National Guideline Centre

Final

Multimorbidity: clinical assessment and management

Multimorbidity: assessment, prioritisation and management of care for people with commonly occurring multimorbidity

NICE guideline NG56

Methods, evidence and recommendations

September 2016

Final .

Commissioned by the National Institute for Health and Care Excellence

Royal College of Physicians

The Royal College of Surgeons of England

RC GP Royal College of General Practitioners

Royal College of Nursing

NICE accredited

1

Funding
National Institute for Health and Care Excellence

Contents

1

Guideline Development Group members

Name	Role
Nina Barnett	Consultant Pharmacist, Care of Older People
Sam Barnett-Cormack	Lay Member
Julia Botsford	Senior Admiral Nurse - Research and Practice Development (resigned from group on 26 03 2015)
Carolyn Chew-Graham	Professor of General Practice Research, Research Institute – Primary Care and Health Sciences, Keele University
Andrew Clegg	Senior Lecturer & Honorary Consultant Geriatrician, University of Leeds & Bradford Teaching Hospitals NHS Foundation Trust
Bruce Guthrie	Guideline Chair, Professor of Primary Care Medicine, University of Dundee
John Hindle	Consultant Geriatrician, Special Interest in Movement Disorders, Betsi Cadwaladr University Health Board
Jonathan Inglesfield	General Practitioner & Medical Director, NHS Guildford and Waverley CCG
David Kernick	General Practitioner
Emily Lam	Lay Member
Rupert Payne	Consultant Senior Lecturer in Primary Health Care, Centre for Academic Primary Care, University of Bristol
Alaster Rutherford	Primary Care Pharmacist Consultant, NHS Bath & North East Somerset CCG
Cate Seton-Jones	Medical Director and Consultant in palliative medicine, Phyllis Tuckwell Hospice

NCGC technical team members

Name	Role
Joanna Ashe	Senior Information Scientist
Kate Ashmore	Document Editor/Process Assistant (from April 2015)
Neil Askew	Health Economist
Katie Broomfield	Document Editor/Process Assistant (until April 2015)
Tamara Diaz	Project Manager
Caroline Farmer	Senior Research Fellow
Elisabetta Fenu	Health Economist Lead
James Gilbert	Research Fellow
Hannah K	Research Fellow
Sophia Kemmis-Betty	Senior Health Economist
Emma Madden	Research Fellow
Norma O'Flynn	Guideline Lead, Operations Director

Acknowledgements

The development of this guideline was greatly assisted by the following people:

Name	Role
Jill Cobb	Information Scientist/Document Delivery Lead
Bethany King	Document Editor/Process Assistant
Giulia Zuodar	Document Editor/Process Assistant
Josh South	Project Administrator
Madeline Logan	Project Administrator
Jessica Flynn	Office Manager

1 Guideline summary

1.1 Full list of recommendations

General principles

1. Be aware that multimorbidity refers to the presence of 2 or more long-term health conditions, which can include:
    - defined physical and mental health conditions such as diabetes or schizophrenia
    - ongoing conditions such as learning disability
    - symptom complexes such as frailty or chronic pain
    - sensory impairment such as sight or hearing loss
    - alcohol and substance misuse.
2. Be aware that the management of risk factors for future disease can be a major treatment burden for people with multimorbidity and should be carefully considered when optimising care.
3. Be aware that the evidence for recommendations in NICE guidance on single health conditions is regularly drawn from people without multimorbidity and taking fewer prescribed regular medicines.
4. Think carefully about the risks and benefits, for people with multimorbidity, of individual treatments recommended in guidance for single health conditions. Discuss this with the patient alongside their preferences for care and treatment.

Taking account of multimorbidity in tailoring the approach to care

5. Consider an approach to care that takes account of multimorbidity if the person requests it or if any of the following apply:
    - they find it difficult to manage their treatments or day-to-day activities
    - they receive care and support from multiple services and need additional services
    - they have both long-term physical and mental health conditions
    - they have frailty (see Chapter 8) or falls
    - they frequently seek unplanned or emergency care (see also recommendation 9)
    - they are prescribed multiple regular medicines (see Chapter 6).
6. When offering an approach to care that takes account of multimorbidity, focus on:
    - how the person's health conditions and their treatments interact and how this affects quality of life
    - the person's individual needs, preferences for treatments, health priorities, lifestyle and goals
    - the benefits and risks of following recommendations from guidance on single health conditions

- improving quality of life by reducing treatment burden, adverse events, and unplanned care
- improving coordination of care across services.

7. Follow these steps when delivering an approach to care that takes account of multimorbidity:
    - Discuss the purpose of an approach to care that takes account of multimorbidity (see recommendation 19).
    - Establish disease and treatment burden (see recommendations 20 to 22).
    - Establish patient goals, values and priorities (see recommendations 23 to 25).
    - Review medicines and other treatments taking into account evidence of likely benefits and harms for the individual patient and outcomes important to the person (see recommendations26 to 33).
    - Agree an individualised management plan with the person (see recommendation 34), including:

    - goals and plans for future care (including advance care planning)
    - who is responsible for coordination of care
    - how the individualised management plan and the responsibility for coordination of care is communicated to all professionals and services involved
    - timing of follow-up and how to access urgent care.

How to identify people who may benefit from an approach to care that takes account of multimorbidity

8. Identify adults with multimorbidity who may benefit from an approach to care that takes account of multimorbidity (as outlined in Chapter 6):
    - opportunistically during routine care
    - proactively using electronic health records.

    Use the criteria in recommendation 5 to guide this.

9. Consider using a validated tool such as eFI, PEONY or QAdmissions, if available in primary care electronic health records, to identify adults with multimorbidity who are at risk of adverse events such as unplanned hospital admission or admission to care homes.

10. Consider using primary care electronic health records to identify markers of increased treatment burden such as number of regular medicines a person is prescribed.

11. Use an approach to care that takes account of multimorbidity for adults of any age who are prescribed 15 or more regular medicines, because they are likely to be at higher risk of adverse events and drug interactions.

12. Consider an approach to care that takes account of multimorbidity for adults of any age who:
    - are prescribed 10 to 14 regular medicines
    - are prescribed fewer than 10 regular medicines but are at particular risk of adverse events.

How to assess frailty

13. Consider assessing frailty in people with multimorbidity.

14. Be cautious about assessing frailty in a person who is acutely unwell.

15. Do not use a physical performance tool to assess frailty in a person who is acutely unwell.

Primary care and community care settings

16. When assessing frailty in primary and community care settings, consider using 1 of the following:

- an informal assessment of gait speed (for example, time taken to answer the door, time taken to walk from the waiting room)
- self-reported health status (that is, 'how would you rate your health status on a scale from 0 to 10?', with scores of 6 or less indicating frailty)
- a formal assessment of gait speed, with more than 5 seconds to walk 4 metres indicating frailty
- the PRISMA-7 questionnaire, with scores of 3 and above indicating frailty.

Hospital outpatient settings

17. When assessing frailty in hospital outpatient settings, consider using 1 of the following:

- self-reported health status (that is, 'how would you rate your health status on a scale from 0 to 10?', with scores of 6 or less indicating frailty)
- the 'Timed Up and Go' test, with times of more than 12 seconds indicating frailty
- a formal assessment of gait speed, with more than 5 seconds to walk 4 metres indicating frailty
- the PRISMA-7 questionnaire, with scores of 3 and above indicating frailty
- self-reported physical activity, with frailty indicated by scores of 56 or less for men and 59 or less for women using the Physical Activity Scale for the Elderly.

Delivering an approach to care that takes account of multimorbidity

18. Follow the recommendations in the NICE guideline on patient experience in adult NHS services which provides guidance on knowing the patient as an individual, tailoring healthcare services for each patient, continuity of care and relationships, and enabling patients to actively participate in their care.

Discussing the purpose of an approach to care that takes account of multimorbidity

19. Discuss with the person the purpose of the approach to care, that is, to improve quality of life. This might include reducing treatment burden and optimising care and support by identifying:

- ways of maximising benefit from existing treatments
- treatments that could be stopped because of limited benefit
- treatments and follow-up arrangements with a high burden

- medicines with a higher risk of adverse events (for example, falls, gastrointestinal bleeding, acute kidney injury)
- non-pharmacological treatments as possible alternatives to some medicines
- alternative arrangements for follow-up to coordinate or optimise the number of appointments.

Establishing disease and treatment burden

20. Establish disease burden by talking to people about how their health problems affect their day-to-day life. Include a discussion of:
    - mental health
    - how disease burden affects their wellbeing
    - how their health problems interact and how this affects quality of life.
21. Establish treatment burden by talking to people about how treatments for their health problems affect their day-to-day life. Include in the discussion:
    - the number and type of healthcare appointments a person has and where these take place
    - the number and type of medicines a person is taking and how often
    - any harms from medicines
    - non-pharmacological treatments such as diets, exercise programmes and psychological treatments
    - any effects of treatment on their mental health or wellbeing.
22. Be alert to the possibility of:
    - depression and anxiety (consider identifying, assessing and managing these conditions in line with the NICE guideline on common mental health disorders)
    - chronic pain and the need to assess this and the adequacy of pain management.

Establishing patient goals, values and priorities

23. Clarify with the patient whether and how they would like their partner, family members and/or carers to be involved in key decisions about the management of their conditions. Review this regularly. If the patient agrees, share information with their partner, family members and/or carers. [This recommendation is adapted from the NICE guideline on patient experience in adult NHS services.]
24. Encourage people with multimorbidity to clarify what is important to them, including their personal goals, values and priorities. These may include:
    - maintaining their independence
    - undertaking paid or voluntary work, taking part in social activities and playing an active part in family life
    - preventing specific adverse outcomes (for example, stroke)
    - reducing harms from medicines
    - reducing treatment burden
    - lengthening life.

25. Explore the person's attitudes to their treatments and the potential benefits and harms of those treatments. Follow the recommendations on patient involvement in decisions about medicines and understanding the patient's knowledge, beliefs and concerns about medicines in the NICE guideline on medicines adherence.

Reviewing medicines and other treatments

26. When reviewing medicines and other treatments, use the database of treatment effects to find information on:
    - the effectiveness of treatments
    - the duration of treatment trials
    - the populations included in treatment trials.
27. Consider using a screening tool (for example, the STOPP/START tool in older people) to identify medicine-related safety concerns and medicines the person might benefit from but is not currently taking. [This recommendation is adapted from the NICE guideline on medicines optimisation.]
28. When optimising treatment, think about any medicines or non-pharmacological treatments that might be started as well as those that might be stopped.
29. Ask the person if treatments intended to relieve symptoms are providing benefits or causing harms. If the person is unsure of benefit or is experiencing harms from a treatment:
    - discuss reducing or stopping the treatment
    - plan a review to monitor effects of any changes made and decide whether any further changes to treatments are needed (including restarting a treatment).
30. Take into account the possibility of lower overall benefit of continuing treatments that aim to offer prognostic benefit, particularly in people with limited life expectancy or frailty.
31. Discuss with people who have multimorbidity and limited life expectancy or frailty whether they wish to continue treatments recommended in guidance on single health conditions which may offer them limited overall benefit.
32. Discuss any changes to treatments that aim to offer prognostic benefit with the person, taking into account:
    - their views on the likely benefits and harms from individual treatments
    - what is important to them in terms of personal goals, values and priorities (see recommendation 24).
33. Tell a person who has been taking bisphosphonate for osteoporosis for at least 3 years that there is no consistent evidence of:
    - further benefit from continuing bisphosphonate for another3 years
    - harms from stopping bisphosphonate after 3 years of treatment.

    Discuss stopping bisphosphonate after 3 years and include patient choice, fracture risk and life expectancy in the discussion.

Agreeing the individualised management plan

34. After a discussion of disease and treatment burden and the person's personal goals, values and priorities, develop and agree an individualised management

plan with the person. Agree what will be recorded and what actions will be taken. These could include:

- starting, stopping or changing medicines and non-pharmacological treatments
- prioritising healthcare appointments
- anticipating possible changes to health and wellbeing
- assigning responsibility for coordination of care and ensuring this is communicated to other healthcare professionals and services
- other areas the person considers important to them
- arranging a follow-up and review of decisions made.

Share copies of the management plan in an accessible format with the person and (with the person's permission) other people involved in care (including healthcare professionals, a partner, family members and/or carers).

Comprehensive assessment in hospital

35. Start a comprehensive assessment of older people with complex needs at the point of admission and preferably in a specialist unit for older people. [This recommendation is from the NICE guideline on transition between inpatient hospital settings and community or care home settings for adults with social care needs.]

1.2 Key research recommendations

1. Is it possible to analyse primary care data to identify characteristics that affect life expectancy and to develop algorithms and prediction tools for patients and healthcare providers to predict reduced life expectancy?
2. What is the clinical and cost effectiveness of stopping preventive medicines in people with multimorbidity who may not benefit from continuing them?
3. What is the clinical and cost effectiveness of a community holistic assessment and intervention for people living with high levels of multimorbidity?
4. What is the clinical and cost effectiveness of alternative approaches to organising primary care compared with usual care for people with multimorbidity?

2 Introduction

Multimorbidity is usually defined as when an individual has two or more long-term conditions. Measuring the prevalence of multimorbidity is not straightforward since this will vary depending on which conditions are counted, but all recent studies show that multimorbidity is common, becomes more common as people age, and is more common in people from less affluent areas.[14,202] A recent large UK based study found that 42% of the population had at least one of the 40 conditions counted, and 23% had multimorbidity. Two-thirds of people aged 65 years or over had multimorbidity, and 47% had three or more conditions. People living in the most deprived areas had double the rate of multimorbidity in middle age than those living in the most affluent areas. Put another way, they developed multimorbidity 10-15 years before their more affluent peers. The recognition of multimorbidity associated with socioeconomic depreivation is particularly important as NHS England has a legal duty to have regard to the need to reduce health inequalities. Whereas rates of multimorbidity in older people was largely due to higher rates of physical conditions, in the less affluent multimorbidity was due to combinations of physical and mental health conditions was common.[14]

For many people multimorbidity will present few problems but multimorbidity matters because it is associated with reduced quality of life, higher mortality, polypharmacy and high treatment burden, higher rates of adverse drug events, and much greater health services use including emergency hospital admissions.[80,246] A particular issue for health services and clinicians is that pharmacological and non-pharmacological treatment regimens can become burdensome in people with complex multimorbidity, and care can become uncoordinated and fragmented.[35,114,143] Polypharmacy in people with multimorbidity is often driven by the introduction of multiple drugs intended to prevent future morbidity and mortality, but the case for using such drugs weakens as life expectancy reduces. The absolute difference made by each additional drug may also reduce when people are taking multiple preventative medicines.[101] The implications of multimorbidity for organisation of healthcare are highly variable depending on which conditions an individual has. Groups of conditions which have closely related or concordant treatment, such as diabetes, hypertension and angina pose fewer problems of co-ordination than groups where treatment is discordant, such as people who experience both physical and mental health conditions.

NICE guidelines have been developed for the management of many individual diseases and conditions. The aim of this guideline is to inform patient and clinical decision-making and models of care for people with multimorbidity who would benefit from an individual approach because of high impact on their quality of life or functioning due to their conditions or their treatments. Although this is a particular concern of generalists such as general practitioners or geriatricians, the guideline is also relevant to specialists since many of the patients they care for will have other significant conditions.

3 Development of the guideline

3.1 What is a NICE clinical guideline?

NICE clinical guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. We base our clinical guidelines on the best available research evidence, with the aim of improving the quality of healthcare. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE clinical guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional.

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills.

We produce our guidelines using the following steps:

- Guideline topic is referred to NICE from NHS England.
- Stakeholders register an interest in the guideline and are consulted throughout the development process.
- The scope is prepared by the National Clinical Guideline Centre (NCGC).
- The NCGC establishes a Guideline Development Group.
- A draft guideline is produced after the group assesses the available evidence and makes recommendations.
- There is a consultation on the draft guideline.
- The final guideline is produced.

The NCGC and NICE produce a number of versions of this guideline:

- the ‘full guideline’ contains all the recommendations, plus details of the methods used and the underpinning evidence
- the ‘NICE guideline’ lists the recommendations
- ‘information for the public’ is written using suitable language for people without specialist medical knowledge
- NICE Pathways brings together all connected NICE guidance.

This version is the full version. The other versions can be downloaded from NICE at www.nice.org.uk.

3.2 Remit

NICE received the remit for this guideline from NHS England. NICE commissioned the NCGC to produce the guideline.

The remit for this guideline is:

Multimorbidity: Assessment, prioritisation and management of care for people with commonly occurring multimorbidity.

3.3 Who developed this guideline?

A multidisciplinary Guideline Development Group (GDG) comprising health professionals and researchers as well as lay members developed this guideline (see the list of Guideline Development Group members and the acknowledgements).

The National Institute for Health and Care Excellence (NICE) funds the National Clinical Guideline Centre (NCGC) and thus supported the development of this guideline. The GDG was convened by the NCGC and chaired by Professor Bruce Guthrie in accordance with guidance from NICE.

The group met approximately every 5 – 6 weeks during the development of the guideline. At the start of the guideline development process all GDG members declared interests including consultancies, fee-paid work, shareholdings, fellowships and support from the healthcare industry. At all subsequent GDG meetings, members declared arising conflicts of interest.

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in Appendix B.

Staff from the NCGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers (research fellows), health economists and information scientists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate and drafted the guideline in collaboration with the GDG.

3.3.1 What this guideline covers

The groups that will be covered by this guideline includes adults (18 years and over) with multimorbidity. For further details please refer to the scope in Appendix A and the review questions in Section 4.1.

3.3.2 What this guideline does not cover

The groups that will not be covered by this guideline include:

- Children and young people under 18 years.
- People who only have multiple mental health problems and no physical health problems.
- People with a single long-term condition.

3.3.3 Relationships between the guideline and other NICE guidance

Related NICE guidelines:

- Care of the dying adult. NICE guidance NG32 (2015).
- Depression in adults with a chronic physical health problem. NICE clinical guideline 91 (2009).
- Medicines Adherence: involving patients in decisions about prescribed medicines and supporting adherence. NICE clinical guideline 76 (2009).
- Medicines optimisation. NICE clinical guideline NG5 (2015).
- Older people: independence and mental wellbeing. NICE guidance (2015).
-

- Older people with social care needs and multiple long-term conditions. NICE social care guidance. NICE guidance NG22 (2015).
- Psychosis with co-existing substance misuse. NICE clinical guideline 120 (2011).
- Transition between inpatient hospital settings and community or care home settings for adults with social care needs. NICE guidance NG 27 (2015).

Related NICE guidance currently in development:

- Dual diagnosis: meeting people's wider health and social care needs when they have a severe mental illness and misuse substances. NICE public health guidance. Publication expected September 2016
- Multimorbidities: system integration to meet population needs. NICE public health guidance. Publication date to be confirmed.

4 Methods

This chapter sets out in detail the methods used to review the evidence and to develop the recommendations that are presented in subsequent chapters of this guideline. This guidance was developed in accordance with the methods outlined in the NICE guidelines manual, 2012 and 2014 versions.[169,173]

Sections 4.1 to 4.2 describe the process used to identify and review clinical evidence (summarised in Figure 1), Sections 4.1 and 4.3.6 describe the process used to identify and review the health economic evidence, and Section 4.5 describes the process used to develop recommendations.

Figure 1: Step-by-step process of review of evidence in the guideline

4.1 Developing the review questions and outcomes

Review questions were developed using a PICO framework (patient, intervention, comparison and outcome) for intervention reviews; using a framework of population, index tests, reference standard and target condition for reviews of diagnostic test accuracy; and using population, presence or absence of factors under investigation (for example prognostic factors) and outcomes for prognostic reviews.

This use of a framework guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the GDG. The review questions were drafted by the NCGC technical team and refined and validated by the GDG. The questions were based on the key clinical areas identified in the scope (Appendix A).

A total of 18 review questions were identified.

Full literature searches, critical appraisals and evidence reviews were completed for all the specified review questions.

Table 1: Review questions

Chapter	Type of review	Review questions	Outcomes
	Qualitative	What principles are important for assessing, prioritising and managing care for people with multimorbidity?	Themes as identified by the evidence
	Qualitative	What are barriers to healthcare professionals optimising care for people with multimorbidity?	Themes as identified by the evidence
	Prognostic risk factor	What risk tool best identifies people with multimorbidity who are at risk of unplanned hospital admissions?	Unplanned hospital admissions (max time point=3 years) Statistical outputs may include: • Area under the ROC curve (c-index, c-statistic) • Sensitivity, specificity, predictive values • Predicted risk versus observed risk (calibration) • Other Statistical measures: for example, D statistic, R^2 statistic and Brier score Reclassification.
	Prognostic risk tool	What risk tool best identifies people with multimorbidity who are at risk of reduced health-related quality of life?	Reductions in health related quality of life (max time point=3 years) Statistical outputs may include: • Area under the ROC curve (c-index, c-statistic) • Sensitivity, specificity, predictive values • Predicted risk versus observed risk (calibration) • Other Statistical measures: for example, D statistic, R2 statistic and Brier score • Reclassification.
	Prognostic risk tool	What risk tool best identifies people with multimorbidity who are at risk of admission to care facility?	Admission to care facility (max time point = 3 years) Statistical outputs may include: • Area under the ROC curve (c-index, c-statistic) • Sensitivity, specificity, predictive values • Predicted risk versus observed risk (calibration)

Chapter	Type of review	Review questions	Outcomes
			• Other Statistical measures: for example, D statistic, R2 statistic and Brier score • Reclassification.
	Prognostic risk tool	What risk tool best identifies people with multimorbidity who are at risk of reduced life expectancy?	Mortality Statistical outputs may include: • Area under the ROC curve (c-index, c-statistic) • Sensitivity, specificity, predictive values • Predicted risk versus observed risk (calibration) • Other Statistical measures: for example, D statistic, R^2 statistic and Brier score • Reclassification.
	Prognostic risk factor	Is polypharmacy associated with a greater risk of unplanned hospital admissions amongst people with multimorbidity?	Unplanned hospital admissions at $\geq$ 1 year Statistical outputs may include: Sensitivity, specificity, C-statistic, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
	Prognostic risk factor	Is polypharmacy associated with a greater risk of reductions in health-related quality of life amongst people with multimorbidity?	Health-related quality of life at $\geq$ 1 year Statistical outputs may include: Sensitivity, specificity, C-statistic, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
	Prognostic risk factor	Is polypharmacy associated with a greater risk of admission to care facility amongst people with multimorbidity?	Admission to care facility at $\geq$ 1 year. Statistical outputs may include: Sensitivity, specificity, C-statistic, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
	Prognostic risk factor	Is polypharmacy associated with a greater risk of mortality amongst people with multimorbidity?	Mortality at $\geq$ 1 year. Statistical outputs may include: Sensitivity, specificity, C-statistic, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
	Diagnostic test accuracy	What is the most accurate tool for assessing frailty?	Sensitivity, specificity, C-statistic
	Questionnaire performance	How can treatment burden be assessed?	Reliability Validity

Chapter	Type of review	Review questions	Outcomes
			Reproducibility Responsiveness Interpretability Time to complete User friendliness
	Bespoke review	How might data from condition-specific guidance best be used and presented to inform a ranking of treatments based on absolute risk and benefit and time to achieve benefits?	• -
	Intervention	What is the clinical and cost-effectiveness of stopping antihypertensive treatment?	Critical: • All-cause mortality • Cardiovascular mortality • Non-fatal myocardial infarction • Stroke • Quality of life • Hospitalisation • Admission to care facility Important: • Blood pressure • Falls
	Intervention	What is the clinical and cost effectiveness of stopping drugs for osteoporosis?	Critical: • Health related quality of life • Functional outcomes (e.g. mobility, activities of daily living, FIM, or Barthel index, performance status) • Fracture • Falls • Pain • Hospitalisation • Admission to care facility Important: • GI bleed • Atypical fracture • Oseonecrosis jaw • Discontinuation of medication due to side effects.
	Intervention	What is the clinical and cost effectiveness of stopping statin treatment?	Critical: • Quality of life (continuous) • Hospitalisation (dichotomous) • All-cause mortality(time to event) • Cardiovascular mortality (time to event) • Stroke (dichotomous)

Chapter	Type of review	Review questions	Outcomes
			• Non-fatal myocardial infarction (dichotomous) • Admission to care home (dichotomous) Important: Myalgia (dichotomous)
	Intervention	What models of care improve outcomes in patients with multimorbidity?	Critical: • Health-related quality of life • Mortality • Functional outcomes (for example mobility, activities of daily living) • Patient and carer satisfaction • Length of hospital stay • Unscheduled care • Admission to care facility Important: • Continuity of care • Patient/carer burden
	Intervention	What is the clinical and cost-effectiveness of self-management and expert patient programmes for people with multimorbidity?	Critical: Health-related quality of life (continuous) Mortality (time to event/dichotomous) Functional outcomes (mobility, activities of daily living) (continuous) Patient and carer satisfaction (continuous) Unplanned hospital admissions (dichotomous) Length of hospital stay (continuous) Important Continuity metrics (continuous) Patient/carer treatment burden (continuous) Patient self-efficacy (continuous)
	Intervention	What format of encounters with healthcare professionals improves outcomes for people with multimorbidity?	Critical • Quality of life (continuous) • Mortality (dichotomous) • Functional outcomes (continuous) • Patient/carer satisfaction (continuous) • Length of hospital stay

Chapter	Type of review	Review questions	Outcomes
			(continuous) • Unscheduled care (dichotomous) Important • Continuity of care (dichotomous) • Patient/carer treatment burden (dichotomous) • Admission to care facility (dichotomous)

4.2 Searching for evidence

4.2.1 Clinical literature search

Systematic literature searches were undertaken to identify the published clinical evidence relevant to the review questions. Searches were undertaken according to the parameters stipulated within the NICE guidelines manual.[169] Databases were searched using relevant medical subject headings, free-text terms and study-type filters where appropriate. Where possible, searches were restricted to articles published in English. Studies published in languages other than English were not reviewed. All searches were conducted in Medline, Embase, and The Cochrane Library. Additional subject specific databases were used for some questions: AMED for models of care; CINAHL for barriers, models or care and burden of treatment; PsycINFO for barriers and burden of treatment. All searches were updated on 4 January 2016. One additional paper [44] published after this date was included following stakeholder consultation and this is discussed in section 7.4.

Search strategies were quality assured by cross-checking reference lists of highly relevant papers, analysing search strategies in other systematic reviews, and asking GDG members to highlight any additional studies. Searches were quality assured by a second information scientist before being run. The questions, the study types applied, the databases searched and the years covered can be found in Appendix G.

The titles and abstracts of records retrieved by the searches were sifted for relevance, with potentially significant publications obtained in full text. These were assessed against the inclusion criteria. Reference lists for papers that met the inclusion criteria were checked for further potentially relevant papers. These papers were obtained in full text and assessed against the inclusion criteria.

During the scoping stage, a search was conducted for guidelines and reports on the websites listed below from organisations relevant to the topic.

- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)
- National Institutes of Health Consensus Development Program (consensus.nih.gov)
- NHS Evidence Search (www.evidence.nhs.uk).

All references sent by stakeholders were considered. Searching for unpublished literature was not undertaken. The NCGC and NICE do not have access to drug manufacturers' unpublished clinical trial results, so the clinical evidence considered by the GDG for pharmaceutical interventions may be different from that considered by the MHRA and European Medicines Agency for the purposes of licensing and safety regulation.

4.2.2 Health economic literature search

Systematic literature searches were also undertaken to identify health economic evidence within published literature relevant to the review questions. The evidence was identified by conducting a broad search relating to multimorbidity in the: NHS Economic Evaluation Database (NHS EED), the Health Technology Assessment database (HTA) and the Health Economic Evaluations Database (HEED) with no date restrictions (NHS EED ceased to be updated after March 2015; HEED was used for searches up to December 2014 but subsequently ceased to be available). Additionally, the search was run on Medline and Embase using a health economic filter, from 2013, to ensure recent publications that had not yet been indexed by the economic databases were identified. This was supplemented by additional searches that looked for economic papers specifically relating to models of care, holistic assessment, burden of treatment and stopping treatments on Medline, Embase, NHS EED, HTA and HEED as it became apparent that some papers in this area had not been identified by the first search. Where possible, searches were restricted to articles published in English. Studies published in languages other than English were not reviewed.

4.3 Identifying and analysing evidence of effectiveness

Research fellows conducted the tasks listed below, which are described in further detail in the rest of this section:

- Identified potentially relevant studies for each review question from the relevant search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against pre-specified inclusion and exclusion criteria to identify studies that addressed the review question in the appropriate population, and reported on outcomes of interest (review protocols are included in Appendix C).
- Critically appraised relevant studies using appropriate study design checklist as specified in the NICE guidelines manual[169,173]. Prognostic risk factor reviews were appraised using QUIPS[104,105] prognostic risk tool reviews were appraised using PROBAST, qualitative studies were critically appraised using NCGC checklists, and previously published guidelines were appraised using AGREE II.
- Extracted key information about interventional study methods and results using 'Evibase', NCGC's purpose-built software. Evibase produces summary evidence tables, including critical appraisal ratings. Key information about non-interventional study methods and results was manually extracted onto standard evidence tables and critically appraised separately (evidence tables are included in Appendix H).
- Generated summaries of the evidence by outcome. Outcome data were combined, analysed and reported according to study design:
  - Randomised data for intervention reviews were meta-analysed where appropriate and reported in GRADE profiles. Where meta-analysis was not appropriate due to heterogeneity across studies, data from individual studies was presented separately.
  - Diagnostic accuracy and prognostic data were meta-analysed where appropriate and reported in adapted GRADE profile tables. Where meta-analysis was not appropriate due to heterogeneity across studies, data from individual studies was presented separately.
  - Qualitative data was summarised across studies where appropriate and reported in themes.
  - Questionnaire performance data was presented as a range of values in adapted GRADE profiles.
- A sample of a minimum of 10% of the abstract lists of the first 3 sifts by new reviewers and those for complex review questions (for example, prognostic reviews) were double-sifted by a senior research fellow and any discrepancies were rectified. All of the evidence reviews were quality assured by a senior research fellow. This included checking:

- papers were included or excluded appropriately
- a sample of the data extractions
- correct methods were used to synthesise data
- a sample of the risk of bias assessments.

4.3.1 Inclusion and exclusion criteria

The inclusion and exclusion of studies was based on the criteria defined in the review protocols, which can be found in Appendix C. Excluded studies by review question (with the reasons for their exclusion) are listed in Appendix L. The GDG was consulted about any uncertainty regarding inclusion or exclusion.

The key population inclusion criterion, relevant across the majority of the reviews in the guideline, was adults with multimorbidity. Multimorbidity was defined as the presence of two or more chronic conditions where these included at least one physical health condition. The key population exclusion criterion was people without multimorbidity, or people with multimorbidity with two or more mental health conditions without a coexisting physical health condition.

During development, it was noted that the majority of papers identified in literature searches did not specify whether the study population was multimorbid, or reported baseline characteristics that were unclear or unreliable measures of multimorbidity. The GDG agreed a standard for including papers without clear reporting of the multimorbidity of the population in a review, and under what circumstances these would be downgraded for indirectness as part of the quality process. This standard was intended to maximise the likelihood that papers included in the reviews were including people with multimorbidity, while also not excluding the vast majority of evidence that was identified. The standard used across the majority of the reviews is as follows:

Where papers clearly reported the proportion of people in the study sample who were multimorbid:

- a paper was included if >95% of the population were multimorbid
- a paper was included if 80%-95% of the population were multimorbid and was downgraded once for indirectness.
- A paper was excluded if <80% of the population were multimorbid

Where papers did not clearly report the proportion of people in the study sample who were multimorbid:

- A paper was included if the study sample was an older adult population (>65 years) and downgraded for indirectness. This standard is based on evidence that approximately 70% of older adults have two or more comorbidities. Papers were excluded if other baseline characteristics indicated that the population was not multimorbid.
- A paper may be included if the reviewer believed that the population is likely to be multimorbid based on the study characteristics reported in the paper. This included consideration of the population characteristics (for example, proportion of study population identified as frail; place of residence) and the study characteristics (for example, study aims and settings). These decisions were agreed with the GDG.

The GDG discussed reliable metrics of multimorbidity. The GDG agreed that the following metrics were not reliable indices of multimorbidity and papers could not be included based on these measures; (i) disease counts (for example, the Charlson comorbidity index) (ii) the mean number of conditions in the study sample, (iii) the N and % of participants with each single condition. These metrics were identified as being unreliable as they do not account for the propensity for conditions

to 'cluster'; such that individuals with one long-term condition are more likely than the general population to develop further long-term conditions.

In some cases, the standard was adjusted according to the need of the review. For example, studies with older adults where the proportion of the study sample with multimorbidity was unclear were not downgraded for indirectness if the GDG felt that this would not contribute to a difference in the effect size. Any alterations to the standard, and the rationale for this, is explained in the introduction for each of the reviews. Further information on the way papers were assessed for indirectness is explained later in this chapter (section 4.3.4).

Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded.

4.3.2 Type of studies

Randomised trials, observational studies (including diagnostic, prognostic, and questionnaire performance studies), qualitative studies, and previously published guidelines were included in the evidence reviews as appropriate.

For all intervention reviews in this guideline, parallel randomised controlled trials (RCTs) were prioritised for inclusion because they are considered the most robust type of study design that can produce an unbiased estimate of the intervention effects. For each intervention review, the GDG considered whether non-randomised trials were appropriate for inclusion. In all instances the GDG felt that RCTs would provide a better standard of evidence and therefore decided to only include non-randomised trials if no RCTs were included. No non-randomised trials were included in the guideline.

For diagnostic review questions, prospective and retrospective cohort studies in which the index test(s) and the reference standard test are applied to the same patients in a cross-sectional design were included. For prognostic review questions, prospective and retrospective cohort studies were included. Case–control studies were not included.

Two types of qualitative review were used in this guideline.

(1) One of these reviews sought the perspectives of individuals with multimorbidity, their carers, and healthcare professionals who provide care for people with multimorbidity. This review included interview and focus group studies.
(2) A separate review sought to identify principles for the care of people with multimorbidity that are recommended by experts in the care of multimorbidity, including people with multimorbidity, their carers, and healthcare professionals who care for people with multimorbidity. This review examined included reported advice and recommendations from already published guidelines relevant to the care of people with multimorbidity, including NICE guidelines, guidelines published by other recognised professional health groups, and other publications where the primary aim was to report recommendations for clinical practice.

In this guideline one questionnaire performance review was conducted to evaluate the performance of questionnaires where there was no established reference standard (gold standard) with which to derive diagnostic accuracy data. Cross-sectional, retrospective and prospective cohort studies were included.

Please refer to the review protocols in Appendix C for full details on the study design of studies selected for each review question.

4.3.3 Methods of combining clinical studies

4.3.3.1 Data synthesis for intervention reviews

Where possible, meta-analyses were conducted using Cochrane Review Manager (RevMan5)[1] software to combine the data given in all studies for each of the outcomes of interest for the review question.

For some questions, the GDG specified that data should be stratified, meaning that studies that varied on a particular factor were not combined and analysed together. Where stratification was used, this is documented in the individual question protocols (see Appendix C). If additional strata were used this led to sub-strata (for example, 2 stratification criteria would lead to 4 sub-strata categories, or 3 stratification criteria would lead to 9 sub-strata categories) which would be analysed separately.

4.3.3.1.1 Analysis of different types of data

Dichotomous outcomes

Fixed-effects (Mantel-Haenszel) techniques (using an inverse variance method for pooling) were used to calculate risk ratios (relative risk, RR) for the binary outcomes, which included:

- mortality
- adverse events
- resource use.

The absolute risk difference was also calculated using GRADEpro[98] software, using the median event rate in the control arm of the pooled results.

For binary variables where there were zero events in either arm or a less than 1% event rate, Peto odds ratios, rather than risk ratios, were calculated. Peto odds ratios are more appropriate for data with a low number of events.

Where sufficient information was provided, hazard ratios were calculated in preference for outcomes such as mortality where the time to the event occurring was important for decision-making. Where incomplete data was reported in a paper to extract Hazard Ratios, these were calculated according to established methods.[228] Hazard ratio data was pooled using the generic inverse variance method in Cochran Review Manager (RevMan5[1] software).

Continuous outcomes

Continuous outcomes were analysed using an inverse variance method for pooling weighted mean differences. These outcomes included:

- heath-related quality of life (HRQoL)
- length of stay in hospital
- symptom scales (such as visual analogue scale)
- function and activities of daily living.

Where the studies within a single meta-analysis had different scales of measurement, standardised mean differences were used (providing all studies reported either change from baseline or final values rather than a mixture of both); each different measure in each study was 'normalised' to the standard deviation value pooled between the intervention and comparator groups in that same study.

The means and standard deviations of continuous outcomes are required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p values or 95% confidence intervals (95% CI) were reported, and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5[1] software). Where p values were reported as 'less than', a conservative approach was undertaken. For example, if a p value was reported as 'p≤0.001', the calculations for standard deviations were based on a p value of 0.001. If these statistical measures were not available then the methods described in Section 16.1.3 of the Cochrane Handbook (version 5.1.0, updated March 2011) were applied.

4.3.3.1.2 Generic inverse variance

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5.[1] If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro.[98] If multivariate analysis was used to derive the summary statistic but no adjusted control event rate was reported no absolute risk difference was calculated.

4.3.3.1.3 Heterogeneity

Statistical heterogeneity was assessed for each meta-analysis estimate by considering the chi-squared test for significance at $p<0.1$ or an I-squared (I^2) inconsistency statistic (with an I-squared value of more than 50% indicating significant heterogeneity) as well as the distribution of effects. Where significant heterogeneity was present, predefined subgrouping of studies was carried out according to subgroup categories specified a priori on the protocol by the GDG (see Appendix C).

If the subgroup analysis resolved heterogeneity within all of the derived subgroups, then each of the derived subgroups were adopted as separate outcomes (providing at least 1 study remained in each subgroup. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups. Any subgroup differences were interpreted with caution as separating the groups breaks the study randomisation and as such is subject to uncontrolled confounding.

If all predefined strategies of subgrouping were unable to explain statistical heterogeneity within each derived subgroup, then a random effects (DerSimonian and Laird) model was employed to the entire group of studies in the meta-analysis. A random-effects model assumes a distribution of populations, rather than a single population. This leads to a widening of the confidence interval around the overall estimate, thus providing a more realistic interpretation of the true distribution of effects across more than 1 population. If, however, the GDG considered the heterogeneity was so large that meta-analysis was inappropriate, then the results were described narratively.

4.3.3.2 Data synthesis for diagnostic test accuracy reviews

For diagnostic test accuracy studies, a positive result on the index test was found if the patient had values of the measured quantity above or below a threshold value, and different thresholds could be used. The thresholds were pre-specified by the GDG including whether or not data could be pooled across a range of thresholds. Diagnostic test accuracy measures used in the analysis were: area under the receiver operating characteristics (ROC) curve (AUC or C-statistic), and, for different thresholds (if appropriate), sensitivity and specificity. The threshold of a diagnostic test is defined as the value at which the test can best differentiate between those with and without the target condition. In practice this varies amongst studies. If a test has a high sensitivity then very few people with the condition will be missed (few false negatives). For example, a test with a sensitivity of 97% will only miss 3% of people with the condition. Conversely, if a test has a high specificity then few people without the condition would be incorrectly diagnosed (few false positives). For example, a test with a specificity of 97% will only incorrectly diagnose 3% of people who do not have the condition as positive. For each review, the GDG discussed the relative importance of sensitivity versus specificity,

taking into consideration the clinical context of the review. Coupled forest plots of sensitivity and specificity with their 95% CIs across studies (at various thresholds) were produced for each test, using RevMan5.[1] In order to do this, 2×2 tables (the number of true positives, false positives, true negatives and false negatives) were directly taken from the study if given, or else were derived from raw data or calculated from the set of test accuracy statistics.

Diagnostic meta-analysis was considered but was not conducted due to insufficient data. Evidence was presented individually, or as the median sensitivity and specificity where more than one study reported evidence for the same tool. If an even number of studies were reported the results of the study with the lower specificity value of the 2 middle studies was reported, alongside the full range of CIs from all studies.

Heterogeneity or inconsistency amongst studies was visually inspected in the forest plots.

Area under the ROC curve (AUC) data for each study were also plotted on a graph, for each diagnostic test. The AUC describes the overall diagnostic accuracy across the full range of thresholds. The following criteria were used for evaluating AUCs:

- ≤0.50: worse than chance
- 0.50–0.60: very poor
- 0.61–0.70: poor
- 0.71–0.80: moderate
- 0.81–0.92: good
- 0.91–1.00: excellent or perfect test.

Heterogeneity or inconsistency amongst studies was visually inspected.

4.3.3.3 Data synthesis for prognostic factor reviews

Evidence on the risk prediction of risk factors (discrimination data) were prioritised for inclusion, as these data can indicate the impact of using a risk factor in clinical practice to identify people who may be at risk of the outcome (that is, the sensitivity and specificity of the tool, as explained above (section 4.3.3.2). In addition, odds ratios (ORs), risk ratios (RRs) or hazard ratios (HRs), with their 95% confidence intervals (95% CIs) for the effect of the pre-specified prognostic factors were extracted from the studies. These data indicate the strength of the association between the risk factor and the outcome (for example, people with a threshold of x and above of a risk factors have twice the risk of the outcome than people under the x threshold of the risk factor). This data only provides an indication of the overall trend in relationship between the risk factor and outcome, and does not account for the fact that this relationship can vary between individuals and across populations and settings. Studies were only pooled if the GDG believed that the population, setting, and outcome were sufficiently similar across studies. Studies of lower risk of bias were preferred, taking into account the analysis and the study design. In particular, prospective cohort studies with a pre-specified threshold of the risk factor were preferred.

4.3.3.1 Data synthesis for risk prediction tools

For evidence reviews on risk prediction tools, results were presented separately for discrimination and calibration. The discrimination data were analysed according to the principles outlined under the section on data synthesis for diagnostic accuracy studies. As explained above (data synthesis for prognostic factor reviews), discrimination data can indicate the clinical impact of using a risk prediction tool in clinical practice, and therefore these data were prioritised for inclusion and decision-making. Calibration data for example, R^2, if reported was presented separately to the

discrimination data. Meta-analysis was considered but not performed due to insufficient data reported for each of the risk prediction tools. The results were presented for each study separately along with the quality rating for the study. Inconsistency and imprecision were assessed consistent with methods used for diagnostic accuracy reviews.

4.3.3.2 Data synthesis for qualitative study reviews

For each included paper subthemes were identified and linked to a generic theme. An example of a subtheme identified in one review is 'viewing the patient individualistically and holistically' and this was linked to a broader generic theme of 'Relationship between patients and healthcare professionals'. In some cases, subthemes related to more than 1 generic theme. A summary evidence table of generic themes and underpinning subthemes was produced, along with a narrative description of the evidence, and a summary of the quality of the evidence.

4.3.3.3 Data synthesis for questionnaire performance reviews

Results for questionnaires included in the questionnaire performance review were presented individually. These reviews are useful to evaluate the performance of questionnaires or other tools where there is no available reference (gold) standard for evaluating the principle outcome. Without diagnostic test accuracy data, it is necessary to evaluate the performance of questionnaires across a number of performance metrics; including reliability, validity, and metrics related to the utility and interpretation of the questionnaire in clinical practice. Guidance from the literature was used to inform the interpretation of performance data, which is summarised below:

Table 2: Interpretation of performance data

Performance metric	Threshold for good performance and/or guidance for interpretation
Internal reliability	Cronbach's alpha for the scale is between 0.70 and 0.95
Construct validity	The authors make clear, a priori hypotheses (including direction) between the scale and more than one related measure; appropriate measures are assessed appropriately and acceptable analysis used; at least 75% of the results are consistent with these hypotheses
Reproducibility	A clear time period to assess test-retest reliability is used; the intraclass coefficient (ICC), weighted kappa or Pearson's correlation coefficient is greater than 0.70; adequate agreement between the repeated tests (as assessed by whether the smallest detectable change or limits of agreement is smaller than the minimally important change
Responsiveness	If the responsiveness ratio is at least 1.96 or the AUC at least 0.70
Interpretability	The authors provide mean scores and standard deviations for relevant subgroups in the sample; the authors provide information on what change in score would be clinically meaningful (MIC)
Time to complete	The time to complete the questionnaire (mean, SD and range) is appropriate to the intended setting of use of the questionnaire
User friendliness	If quantitative data used to assess user friendliness, scores (mean, SD and range) on a validated questionnaire indicate questionnaire is acceptable to an appropriate number of people relevant to the target population (as decided by the GDG). If qualitative data used to assess user friendliness, themes identified demonstrate no significant concerns of using the questionnaire in the intended population (as decided by the GDG),

4.3.4 Appraising the quality of evidence by outcomes

4.3.4.1 Intervention reviews

The evidence for outcomes from the included RCTs and were evaluated and presented using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro[98]) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results.

Each outcome was first examined for each of the quality elements listed and defined in Table 3.

Table 3: Description of quality elements in GRADE for intervention studies

Quality element	Description
Risk of bias	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies in the same meta-analysis.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both clinical benefit AND clinical harm) and thus be imprecise.
Publication bias	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome.
Other issues	Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below. Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

4.3.4.1.1 Risk of bias

The main domains of bias for RCTs are listed in Table 4. Each outcome had its risk of bias assessed within each study first. For each study, if there were no risks of bias in any domain, the risk of bias was given a rating of 0. If there was risk of bias in just 1 domain, the risk of bias was given a 'serious' rating of −1, but if there was risk of bias in 2 or more domains the risk of bias was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome, by taking into account the weighting of studies according to study precision. For

example if the most precise studies tended to each have a score of −1 for that outcome, the overall score for that outcome would tend towards −1.

Table 4: Principle domains of bias in randomised controlled trials

Limitation	Explanation
Selection bias (sequence generation and allocation concealment)	If those enrolling patients are aware of the group to which the next enrolled patient will be allocated, either because of a non-random sequence that is predictable, or because a truly random sequence was not concealed from the researcher, this may translate into systematic selection bias. This may occur if the researcher chooses not to recruit a participant into that specific group because of: • knowledge of that participant's likely prognostic characteristics, and • a desire for one group to do better than the other.
Performance and detection bias (lack of blinding of patients and healthcare professionals)	Patients, caregivers, those adjudicating or recording outcomes, and data analysts should not be aware of the arm to which patients are allocated. Knowledge of the group can influence: • the experience of the placebo effect • performance in outcome measures • the level of care and attention received, and • the methods of measurement or analysis all of which can contribute to systematic bias.
Attrition bias	Attrition bias results from an unaccounted for loss of data beyond a certain level (a differential of 10% between groups). Loss of data can occur when participants are compulsorily withdrawn from a group by the researchers (for example, when a per-protocol approach is used) or when participants do not attend assessment sessions. If the missing data are likely to be different from the data of those remaining in the groups, and there is a differential rate of such missing data from groups, systematic attrition bias may result.
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results can also lead to bias, as this may distort the overall impression of efficacy.
Other limitations	For example: • Stopping early for benefit observed in randomised trials, in particular in the absence of adequate stopping rules. • Use of unvalidated patient-reported outcome measures. • Lack of washout periods to avoid carry-over effects in crossover trials. • Recruitment bias in cluster-randomised trials.

4.3.4.1.2 Indirectness

Indirectness refers to the extent to which the populations, interventions, comparisons and outcome measures are dissimilar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention. As for the risk of bias, each outcome had its indirectness assessed within each study first. For each study, if there were no sources of indirectness, indirectness was given a rating of 0. If there was indirectness in just 1 source (for example in terms of population), indirectness was given a 'serious' rating of −1, but if there was indirectness in 2 or more sources (for example, in terms of population and treatment) the indirectness was given a 'very serious' rating of −2. A weighted average score was then calculated across all studies contributing to the outcome by taking into account study precision. For example, if the most precise studies tended to have an indirectness score of −1 each for that outcome, the overall score for that outcome would tend towards −1.

4.3.4.1.3 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. When estimates of the treatment effect across studies differ widely, this suggests true differences in the underlying treatment effect, which may be due to differences in populations, settings or doses. When heterogeneity existed within an outcome (chi-squared $p<0.1$, or $I^2>50\%$), but no plausible explanation could be found, the quality of evidence for that outcome was downgraded. Inconsistency for that outcome was given a 'serious' score of −1 if the I^2 was 50–74% and a 'very serious' score of −2 if the I^2 was 75% or more.

If inconsistency could be explained based on pre-specified subgroup analysis (that is, each subgroup had an $I^2<50\%$), the GDG took this into account and considered whether to make separate recommendations on new outcomes based on the subgroups defined by the assumed explanatory factors. In such a situation the quality of evidence was not downgraded for those emergent outcomes.

Since the inconsistency score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

4.3.4.1.4 Imprecision

The criteria applied for imprecision were based on the 95% CIs for the pooled estimate of effect, and the minimal important differences (MID) for the outcome. The MIDs are the threshold for appreciable benefits and harms, separated by a zone either side of the line of no effect where there is assumed to be no clinically important effect. If either end of the 95% CI of the overall estimate of effect crossed 1 of the MID lines, imprecision was regarded as serious and a 'serious' score of −1 was given. This was because the overall result, as represented by the span of the confidence interval, was consistent with 2 interpretations as defined by the MID (for example, both no clinically important effect and clinical benefit were possible interpretations). If both MID lines were crossed by either or both ends of the 95% CI then imprecision was regarded as very serious and a 'very serious' score of −2 was given. This was because the overall result was consistent with all 3 interpretations defined by the MID (no clinically important effect, clinical benefit and clinical harm). This is illustrated in Figure 2. As for inconsistency, since the imprecision score was based on the meta-analysis results, the score represented the whole outcome and so weighted averaging across studies was not necessary.

The position of the MID lines is ideally determined by values reported in the literature. 'Anchor-based' methods aim to establish clinically meaningful changes in a continuous outcome variable by relating or 'anchoring' them to patient-centred measures of clinical effectiveness that could be regarded as gold standards with a high level of face validity. For example, a MID for an outcome could be defined by the minimum amount of change in that outcome necessary to make patients feel their quality of life had 'significantly improved'. MIDs in the literature may also be based on expert clinician or consensus opinion concerning the minimum amount of change in a variable deemed to affect quality of life or health. For binary variables, any MIDs reported in the literature will inevitably be based on expert consensus, as such MIDs relate to all-or-nothing population effects rather than measurable effects on an individual, and so are not amenable to patient-centred 'anchor' methods.

In this guideline, MIDs found in the literature were used to assess imprecision for the EQ-5D and SF-36 measures of health-related quality of life. These values are displayed below:

Table 5: MIDs used to assess imprecision for the EQ-5D and SF-36 measures

MIDs for assessing between group differences Outcome	MID for imprecision	MID for clinical importance	Source
SF-36^	Physical component summary: 2 Mental component summary: 3		User's manual for the SF-36v2 Health Survey,

MIDs for assessing between group differences Outcome	MID for imprecision	MID for clinical importance	Source
	Physical functioning: 3 Role-physical: 3 Bodily pain: 3 General health: 2 Vitality: 2 Social functioning: 3 Role-emotional: 4 Mental health: 3		Third Edition
EQ5D*	GRADE defaults	0.03	NICE agreed for use in Low Back Pain & Low back Pain GDG opinion

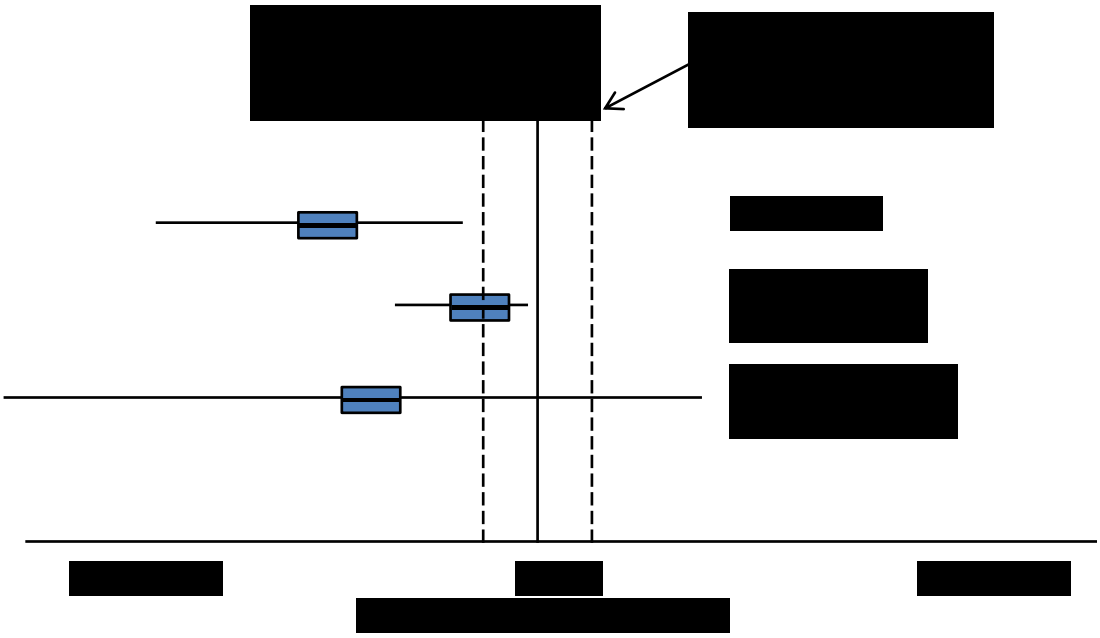
^Note: the SF-12 manual does not specify MIDs. It does however signpost to the SF-36 manual for guidance on interpretation, therefore in this guideline we used the same MIDs for the SF-12.
** Note: this is not based on the literature and was a pragmatic decision for this guideline based on the SF-36 MIDs.*

In the absence of values identified in the literature, the alternative approach to deciding on MID levels is the 'default' method, as follows:

- For categorical outcomes the MIDs were taken to be RRs of 0.75 and 1.25. For 'positive' outcomes such as 'patient satisfaction', the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit. For 'negative' outcomes such as 'stroke', the opposite occurs, so the RR of 0.75 is taken as the line denoting the boundary between no clinically important effect and a clinically significant benefit, whilst the RR of 1.25 is taken as the line denoting the boundary between no clinically important effect and a clinically significant harm.
- For mortality and admission to care home any change was considered to be clinically important and the imprecision was assessed on the basis of whether the confidence intervals crossed the line of no effect, that is, whether the result was consistent with both benefit and harm.
- For continuous outcome variables the MID was taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. Hence the MID denoting the minimum clinically significant benefit was positive for a 'positive' outcome (for example, a quality of life measure where a higher score denotes better health), and negative for a 'negative' outcome (for example, a visual analogue scale [VAS] pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.
- If standardised mean differences have been used, then the MID will be set at the absolute value of +0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the 2 groups, and are thus effectively expressed in units of 'numbers of standard deviations'. The 0.5 MID value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

The default MID value was subject to amendment after discussion with the GDG. If the GDG decided that the MID level should be altered, after consideration of absolute as well as relative effects, this was allowed, provided that any such decision was not influenced by any bias towards making stronger or weaker recommendations for specific outcomes.

Figure 2: Illustration of precise and imprecise outcomes based on the 95% CI of dichotomous outcomes in a forest plot (Note that all 3 results would be pooled estimates, and would not, in practice, be placed on the same forest plot)



4.3.4.1.5 Overall grading of the quality of clinical evidence

Once an outcome had been appraised for the main quality elements, as above, an overall quality grade was calculated for that outcome. The scores (0, −1 or −2) from each of the main quality elements were summed to give a score that could be anything from 0 (the best possible) to −8 (the worst possible). However scores were capped at −3. This final score was then applied to the starting grade that had originally been applied to the outcome by default, based on study design. All RCTs started as High and the overall quality became Moderate, Low or Very Low if the overall score was −1, −2 or −3 points respectively. The significance of these overall ratings is explained in Table 6. The reasons for downgrading in each case were specified in the footnotes of the GRADE tables.

Observational interventional studies started at Low, and so a score of −1 would be enough to take the grade to the lowest level of Very Low. Observational studies could, however, be upgraded if there were all of: a large magnitude of effect, a dose-response gradient, and if all plausible confounding would reduce the demonstrated effect.

Table 6: Overall quality of outcome evidence in GRADE

Level	Description
High	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	Any estimate of effect is very uncertain

4.3.4.2 Diagnostic studies

Risk of bias and indirectness of evidence for diagnostic data were evaluated by study using the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) checklists (see Appendix H in the NICE guidelines manual 2014[169]). Risk of bias and applicability in primary diagnostic accuracy studies in QUADAS-2 consists of 4 domains (see Table 7):

- patient selection
- index test
- reference standard
- flow and timing.

Table 7: Summary of QUADAS-2 with list of signalling, risk of bias and applicability questions.

Domain	Patient selection	Index test	Reference standard	Flow and timing
Description	Describe methods of patient selection. Describe included patients (prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2×2 table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard
Signalling questions (yes/no/ unclear)	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case–control design avoided?	If a threshold was used, was it pre-specified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
Risk of bias; (high/low/ unclear)	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability (high/low/ unclear)	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

4.3.4.2.1 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency was assessed by inspection of the specificity value (based on the primary measure) using the point estimates and 95% CIs of the individual studies on the forest plots. Particular attention was placed on values above or below 50% (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which it would be acceptable to recommend a test). For example, the GDG might have set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example, 0–50%, 50–90% and 90–100%).

4.3.4.2.2 Imprecision

Diagnostic meta-analysis was not conducted in this guideline, and imprecision was assessed according to the range of point estimates or, if only one study contributed to the evidence, the 95% CI around the single study. As a general rule (after discussion with the GDG) a variation of 0–20% was considered precise, 20–40% serious imprecision, and >40% very serious imprecision. Imprecision was assessed on the primary outcome measure for decision-making.

4.3.4.2.3 Overall grading

Quality rating started at High for both prospective and retrospective studies, and each major limitation (risk of bias, indirectness, inconsistency and imprecision) brought the rating down by 1 increment to a minimum grade of Very Low, as explained for intervention reviews.

4.3.4.3 Prognostic risk factor studies

In this guideline, the quality of evidence for prognostic risk factor studies was evaluated according to an amended QUIPS checklist[104,105] which is reported in Table 8. The QUIPS was amended to remove the section on the adequate control of confounding. This is because for the polypharmacy reviews in this guideline, the unadjusted data was preferred and control of plausible confounding was not necessary. If data were meta-analysed the quality for pooled studies was presented. If the data was not pooled then a quality rating was presented for each study.

Table 8: Description of quality elements for prognostic risk factor studies

Quality element	Description of cases where the quality measure would be downgraded
Study design	If case control rather than prospective or retrospective cohort
Participant selection	If potential for selection bias
Prognostic factor measurement	If non-validated and/or unreliable, inappropriate thresholds are chosen (for example, data-driven), and missing data with inappropriate method of imputation
Outcome measurement	If non-validated and/or unreliable
Statistical analysis and reporting	If analysis is not appropriate for the design, insufficient information about the analysis and results
Attrition	If attrition is too high and attrition is related to key characteristics of the study population

Indirectness

Indirectness was assessed as for intervention studies.

Inconsistency

Inconsistency was assessed as for intervention studies.

Imprecision

The criteria applied for imprecision were based on the confidence intervals for the pooled estimate of effect. If either of the 95% confidence intervals of the overall estimate of effect crossed the null line then imprecision was regarded as serious and a 'serious' score of -1 was given. This was because the overall result, as represented by the span of the confidence intervals, was consistent with two conflicting interpretations as defined by the line of no effect (for example, predictive of either low or high risk of the outcome).

Overall grading

Because prognostic reviews were not usually based on multiple outcomes per study, quality rating was assigned by study. However if there was more than one outcome involved in a study, then the quality rating of the evidence statements for each outcome was adjusted accordingly. For example, if one outcome was based on an invalidated measurement method, but another outcome in the same study wasn't, the latter outcome would be graded one grade higher than the other.

Quality rating started at high for prospective and retrospective studies, and each major limitation brought the rating down by one increment to a minimum grade of VERY LOW, as explained for interventional studies. For prognostic studies, prediction tool studies for prognosis are regarded as the gold standard because RCTs are usually inappropriate for these types of review for ethical or pragmatic reasons. Furthermore if the study is looking at more than one risk factor of interest then randomisation would be inappropriate as it can only be applied to one of the risk factors.

4.3.4.4 Prognostic risk tool studies

Risk of bias and indirectness (applicability) of evidence for prognostic risk tool data was evaluated using the Prediction Study Risk of Bias Assessment tool (PROBAST)[2] checklist, which is summarised in Table 9. PROBAST is still under development and the version used in this guideline was acquired from the study author and adapted. One item concerning whether all predictors were available at the time the risk tool would be used in practice was excluded from the risk of bias assessment, and instead was incorporated into an assessment of indirectness. Where the information required to complete PROBAST domains was not reported in publications, this was taken into account for the risk of bias assessment. If the majority of information was available but one domain had limited information there was no obligate downgrade for risk of bias. If more than one domain had limited or no information to inform it's assessment, the study was downgraded once for risk of bias. If very limited or no information was provided for the majority of domains for the study, it was downgraded twice for risk of bias. Ratings were derived for the validation of risk tools; no ratings were provided for the original development phase of the tools.

Table 9: Summary of PROBAST

Quality element	Description of cases where the quality measure would be downgraded
Participant selection	If case control rather than cohort, RCT or nested case-control, or if potential for selection bias
Predictors	If predictors were not defined or assessed in a similar way for all participants, if assessors were not blinded to outcome data
Outcome	If outcome was not defined or assessed in a similar way for all

Quality element	Description of cases where the quality measure would be downgraded
	participants, if assessors were not blinded to predictor information, if predictors were included in the outcome definition
Sample size and participant flow	If there was a low event rate relative to the number of predictors, if there was an inappropriate time interval between predictor assessment and outcome, if risk of selection bias
Analysis	If analysis is not appropriate for the design, if relevant outcome measures were not reported
Applicability	If concerns that the study participants, predictors or outcome are dissimilar to those specified in the review protocol

Inconsistency

Inconsistency refers to an unexplained heterogeneity of results for an outcome across different studies. Inconsistency was assessed by inspection of the specificity value (based on the primary measure) using the point estimates and confidence intervals of the individual studies on the forest plots. Particular attention was placed on values above or below 50% (prognostic accuracy based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 90% as an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 50–90% and 90–100%) and by 2 increments if the individual studies varied across 3 areas (for example, 0–50%, 50–90% and 90–100%).

Imprecision

The judgement of precision was based on visual inspection of the confidence region around the summary sensitivity and specificity point from the diagnostic meta-analysis, if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates or, if only one study contributed to the evidence, the confidence interval around the single study. As a rule of thumb (after discussion with the GDG) a variation of 0–20% was considered precise, 20–40% serious imprecision, and >40% very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Overall grading

Because prognostic reviews were not usually based on multiple tools or outcomes per study, quality rating was assigned by study. However if there was more than one tool or outcome involved in a study, then the quality rating of the evidence statements for each tool and for each outcome was adjusted accordingly. For example, if one outcome was based on an invalidated measurement method, but another outcome in the same study wasn't, the latter outcome would be graded one grade higher than the other.

Quality rating started at HIGH for prospective and retrospective cohort studies, and each major limitation (risk of bias, indirectness, inconsistency and imprecision) brought the rating down by one increment to a minimum grade of VERY LOW, as explained for interventional studies.

4.3.4.5 Qualitative reviews

As explained above in Section 4.3.2, two types of qualitative reviews were included in this guideline. For the review where interviews and focus groups studies were included, the checklist summarised in Table 10 below was used to appraise the quality for each sub-theme. The overall quality rating for each theme is reported in a summary table in the evidence report.

Table 10: Summary of factors used to assess quality in qualitative studies

Quality element	Signalling questions
Limitations of evidence	• Were qualitative studies or surveys an appropriate approach? • Were the studies approved by an ethics committee? • Were the studies clear in what they seek to do? • Is the context clearly described? • Is the role of the researcher clearly described? • How rigorous was the research design and research methods? • Was the data collection rigorous? • Was the data analysis rigorous? • Are the data rich (for qualitative study and open ended survey questions)? • Are the findings relevant to the aims of the study? • Are the findings and conclusions convincing?
Coherence of findings	• Do the subthemes identified complement, reinforce or contradict each other?
Applicability of evidence	• Are the findings of the study applicable to the evidence review? (For example, are the population and setting relevant?)

In the other qualitative review we included recommendations from already published guidelines and other publications where the primary aim was to report recommendations for clinical practice. The quality of this evidence was assessed using AGREE II criteria.[37] The AGREE II tool is used to appraise the quality of guidelines, and is comprised of 6 individual domains and an overall quality rating, summarised in Table 11 below. Consistent with the AGREE II approach, a quality rating for each domain was reported for every guideline. No summary quality rating was produced for the themes identified in the analysis.

Table 11: Summary of domains used to assess quality of published guidelines

Quality element	Signalling questions
Scope and purpose	• Are the overall objectives of the guideline specifically described? • Is the health question(s) covered by the guideline specifically described? • Is the population (patients and public) to whom the guideline is meant to apply is specifically described?
Stakeholder involvement	• Does the guideline committee include individuals from all of the relevant professional groups? • Have the views and preferences of the target population (patients, public) been sought? • Are the target users of the guideline clearly defined?
Rigour of development	• Are systematic methods used to search for evidence? • Are the criteria for selecting evidence clearly described?

Quality element	Signalling questions
	• Are the strengths and limitations of the body of evidence clearly described? • Are the methods for formulating the recommendations clearly described? • Have the health benefits, side effects, and risks been considered in formulating the recommendations? • Is there an explicit link between the recommendations and the supporting evidence? • Has the guideline been externally reviewed by experts prior to its publication? • Is there a procedure for updating the guideline?
Clarity of presentation	• Are the recommendations specific and unambiguous? • Are the different options for the management of the condition clearly presented? • Are key recommendations easily identifiable?
Applicability	• Does the guideline describe facilitators and barriers to its application? • Does the guideline provide advice and/or tools on how the recommendations can be put into practice? • Have the potential resource implications of applying the recommendations been considered? • Does the guideline provide monitoring and/or auditing criteria?
Editorial independence	• The views of the funding body have not influenced the content of the guideline • Have competing interests of GDG members been recorded and addressed?
Overall quality	• Rate the overall quality of the guideline on a scale of 1-7, with higher scores indicating better overall quality

4.3.4.6 Questionnaire performance reviews

Risk of bias of evidence for questionnaire performance data was evaluated using the Questionnaire Bias Assessment Tool (Q-BAST)[78], which is summarised in Table 12 below. Q-BAST consists of six domains, with risk of bias for each domain rated as high, low or unclear. An unclear rating was only given if there was insufficient information provided in the report to make a judgement. An overall rating for each questionnaire was derived, which represented the overall risk of bias across all six domains:

Table 12: Summary of Q-BAST with list of signalling questions.

Quality element	Signalling questions
Research question and study design	• Is there a clear definition of the measurement aim of the questionnaire? • Are there existing measures that the researchers could have used? If so, is the reason for a new measure justified?
Methodological rigour	• Was an appropriate method used to derive questionnaire items? • Was more than one individual involved in choosing the questionnaire items?

Quality element	Signalling questions
	• Are the study methods rigorous? (for example, was the questionnaire adequately piloted? Was the sampling frame sufficiently large? Was more than one individual involved in choosing the questionnaire items?)
Analysis	• Is the analysis sufficiently rigorous? (for example, correct statistical tests for quantitative items, appropriate qualitative analysis). • Were analyses all hypothesis driven? • Was there sufficient justification for any alterations to the measure? • Has an appropriate factor analysis been conducted and used to identify factors within the questionnaire, in the development of individual scales?
Outcome reporting	• Were all relevant data reported?
Missing data	• Was there a high rate of missing data from responders? • Was there a high rate of attrition?
Other	• Is there evidence of floor or ceiling effects? • Were there any other sources of bias?

Indirectness

Indirectness was assessed as for intervention studies.

4.3.5 Assessing clinical importance

The GDG assessed the evidence by outcome in order to determine if there was, or potentially was, a clinically important benefit, a clinically important harm or no clinically important difference between interventions. To facilitate this, binary outcomes were converted into absolute risk differences (ARDs) using GRADEpro[98] software: the median control group risk across studies was used to calculate the ARD and its 95% CI from the pooled risk ratio.

To interpret the clinical evidence for EQ-5D and SF-36 health related quality of life outcomes, the default MIDs (as described in 4.3.4.1.4) were used to identify if the difference between the intervention and comparison indicated a clinical benefit or harm. For other outcomes where MIDs from the literature were not available, the GDG discussed and agreed on whether the point estimate of absolute effect indicated a clinical benefit, harm, or no benefit or harm for each critical outcome. For the critical outcomes of mortality and admission to care home, the GDG agreed that any change would be clinically important; that is, any reduction represented a clinical benefit and any increase represented a clinical harm.

An evidence summary table was produced to compile the GDG's assessments of clinical importance per outcome, alongside the evidence quality and the uncertainty in the effect estimate (imprecision).

4.3.6 Clinical evidence statements

Clinical evidence statements are summary statements that are included in each review chapter, and which summarise the key features of the clinical effectiveness evidence presented. For reviews in this guideline with a limited amount of clinical effectiveness evidence, the evidence statements are presented by outcome and encompass the following key features of the evidence:

- The number of studies and the number of participants for a particular outcome.

- An indication of the direction of clinical importance (if one treatment is beneficial or harmful compared to the other or whether there is no difference between the 2 tested treatments).
- A description of the overall quality of the evidence (GRADE overall quality).

Some of the reviews in this guideline contained a large amount of clinical effectiveness evidence (for example, where a large number of different risk tools were evaluated). For these reviews, a summary of the clinical effectiveness evidence was provided, which encompassed the following key features of the evidence:

- The overall direction of the evidence (for example, the GDG's impression of the clinical effectiveness of the interventions identified and whether any interventions emerged as being strongly clinically beneficial or harmful across critical outcomes)
- Any variation in the direction or quality of the evidence (for example, if the evidence for an intervention was weaker or stronger in a particular strata or subgroup
- More detailed description of key evidence, such as that which was integral to the GDG's discussion and formulation of a recommendation (for example, interventions that emerged as strongly beneficial for people with multimorbidity); including the number of studies and participants for a particular outcome, and a description of the overall quality of the evidence (GRADE overall quality).

4.4 Identifying and analysing evidence of cost-effectiveness

The GDG is required to make decisions based on the best available evidence of both clinical effectiveness and cost-effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected health benefits (that is, their 'cost-effectiveness') in addition to the total implementation cost.[169]

Health economic evidence was sought relating to the key clinical issues being addressed in the guideline. Health economists:

- Undertook a systematic review of the published economic literature.
- Undertook new cost-effectiveness analysis in priority areas.

4.4.1 Literature review

The health economists:

- Identified potentially relevant studies for each review question from the health economic search results by reviewing titles and abstracts. Full papers were then obtained.
- Reviewed full papers against pre-specified inclusion and exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using economic evaluations checklists as specified in the NICE guidelines manual.[169,173]
- Extracted key information about the studies' methods and results into economic evidence tables (included in Appendix I).
- Generated summaries of the evidence in NICE economic evidence profile tables (included in the relevant chapter for each review question) – see below for details.

4.4.1.1 Inclusion and exclusion criteria

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost-utility, cost-effectiveness, cost-benefit and cost-consequences analyses) and

comparative costing studies that addressed the review question in the relevant population were considered potentially includable as economic evidence.

Studies that only reported cost per hospital (not per patient), or only reported average cost-effectiveness without disaggregated costs and effects were excluded. Literature reviews, abstracts, posters, letters, editorials, comment articles, unpublished studies and studies not in English were excluded. Studies published before 1999 and studies from non-OECD countries or the USA were also excluded, on the basis that the applicability of such studies to the present UK NHS context is likely to be too low for them to be helpful for decision-making.

Remaining health economic studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available, then other less relevant studies may not have been included. Where exclusions occurred on this basis, this is noted in the relevant section.

For more details about the assessment of applicability and methodological quality see Table 13 below and the economic evaluation checklist (Appendix G of the 2012 NICE guidelines manual[173]) and the health economics review protocol in Appendix D.

4.4.1.2 NICE economic evidence profiles

NICE economic evidence profile tables were used to summarise cost and cost-effectiveness estimates for the included health economic studies in each review chapter. The economic evidence profile shows an assessment of applicability and methodological quality for each economic study, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from the NICE guidelines manual.[173] It also shows the incremental costs, incremental effects (for example, quality-adjusted life years [QALYs]) and incremental cost-effectiveness ratio (ICER) for the base case analysis in the study, as well as information about the assessment of uncertainty in the analysis. See Table 13 for more details.

When a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity.[184]

Table 13: Content of NICE economic evidence profile

Item	Description
Study	Surname of first author, date of study publication and country perspective with a reference to full information on the study.
Applicability	An assessment of applicability of the study to this guideline, the current NHS situation and NICE decision-making:[(a)] • Directly applicable – the study meets all applicability criteria, or fails to meet 1 or more applicability criteria but this is unlikely to change the conclusions about cost-effectiveness. • Partially applicable – the study fails to meet 1 or more applicability criteria, and this could change the conclusions about cost-effectiveness. • Not applicable – the study fails to meet 1 or more of the applicability criteria, and this is likely to change the conclusions about cost-effectiveness. Such studies would usually be excluded from the review.
Limitations	An assessment of methodological quality of the study:[(a)] • Minor limitations – the study meets all quality criteria, or fails to meet 1 or more quality criteria, but this is unlikely to change the conclusions about cost-effectiveness. • Potentially serious limitations – the study fails to meet 1 or more quality criteria, and this could change the conclusions about cost-effectiveness. • Very serious limitations – the study fails to meet 1 or more quality criteria, and

Item	Description
	this is highly likely to change the conclusions about cost-effectiveness. Such studies would usually be excluded from the review.
Other comments	Information about the design of the study and particular issues that should be considered when interpreting it.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
Cost-effectiveness	Incremental cost-effectiveness ratio (ICER): the incremental cost divided by the incremental effects (usually in £ per QALY gained).
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

(a) Applicability and limitations were assessed using the economic evaluation checklist in Appendix G of the 2012 NICE guidelines manual[173]

4.4.2 Undertaking new health economic analysis

As well as reviewing the published health economic literature for each review question, as described above, new health economic analysis was undertaken by the health economist in selected areas. Priority areas for new analysis were agreed by the GDG after formation of the review questions and consideration of the existing health economic evidence.

The GDG identified outpatient holistic assessment as the highest priority area for original health economic modelling. This area was prioritised as there was uncertainty around the cost effectiveness of holistic assessment as it increases costs but the evidence showed some benefits. More details on the original analysis are reported in Chapter 11 and Appendix N

The following general principles were adhered to in developing the cost-effectiveness analysis:

- Methods were consistent with the NICE reference case for interventions with health outcomes in NHS settings.[169,174]
- The GDG was involved in the design of the model, selection of inputs and interpretation of the results.
- Model inputs were based on the systematic review of the clinical literature supplemented with other published data sources where possible.
- When published data were not available GDG expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.
- The model was peer-reviewed by another health economist at the NCGC.

Full methods for the cost-effectiveness analysis for holistic assessment are described in Appendix N.

4.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that GDGs should consider when judging whether an intervention offers good value for money.[171] In general, an intervention was considered to be cost-effective (given that the estimate was considered plausible) if either of the following criteria applied:

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy.

If the GDG recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in the 'Recommendations and link to evidence' section of the relevant chapter, with reference to issues regarding the plausibility of the estimate or to the factors set out in 'Social value judgements: principles for the development of NICE guidance'.[171]

When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

4.4.4 In the absence of economic evidence

When no relevant published health economic studies were found, and a new analysis was not prioritised, the GDG made a qualitative judgement about cost-effectiveness by considering expected differences in resource use between options and relevant UK NHS unit costs, alongside the results of the review of clinical effectiveness evidence.

The UK NHS costs reported in the guideline are those that were presented to the GDG and were correct at the time recommendations were drafted. They may have changed subsequently before the time of publication. However, we have no reason to believe they have changed substantially.

4.5 Developing recommendations

Over the course of the guideline development process, the GDG was presented with:

- Evidence tables of the clinical and economic evidence reviewed from the literature. All evidence tables are in Appendices H and I.
- Summaries of clinical and economic evidence and quality (as presented in Chapters 5 - 12).
- Forest plots (Appendix K).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (Appendix N).

Recommendations were drafted on the basis of the GDG's interpretation of the available evidence, taking into account the balance of benefits, harms and costs between different courses of action. This was either done formally in an economic model, or informally. Firstly, the net clinical benefit over harm (clinical effectiveness) was considered, focusing on the critical outcomes. When this was done informally, the GDG took into account the clinical benefits and harms when one intervention was compared with another. The assessment of net clinical benefit was moderated by the importance placed on the outcomes (the GDG's values and preferences), and the confidence the GDG had in the evidence (evidence quality). Secondly, the GDG assessed whether the net clinical benefit justified any differences in costs between the alternative interventions.

When clinical and economic evidence was of poor quality, conflicting or absent, the GDG drafted recommendations based on its expert opinion. The considerations for making consensus-based recommendations include the balance between potential harms and benefits, the economic costs compared to the economic benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were agreed through discussions in the GDG. The GDG also considered whether the uncertainty was sufficient to

justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation (see Section 4.5.1 below).

The GDG considered the appropriate 'strength' of each recommendation. This takes into account the quality of the evidence but is conceptually different. Some recommendations are 'strong' in that the GDG believes that the vast majority of healthcare and other professionals and patients would choose a particular intervention if they considered the evidence in the same way that the GDG has. This is generally the case if the benefits clearly outweigh the harms for most people and the intervention is likely to be cost-effective. However, there is often a closer balance between benefits and harms, and some patients would not choose an intervention whereas others would. This may happen, for example, if some patients are particularly averse to some side effect and others are not. In these circumstances the recommendation is generally weaker, although it may be possible to make stronger recommendations about specific groups of patients.

The GDG focused on the following factors in agreeing the wording of the recommendations:

- The actions health professionals need to take.
- The information readers need to know.
- The strength of the recommendation (for example the word 'offer' was used for strong recommendations and 'consider' for weaker recommendations).
- The involvement of patients (and their carers if needed) in decisions on treatment and care.
- Consistency with NICE's standard advice on recommendations about drugs, waiting times and ineffective interventions (see Section 9.2 in the 2014 NICE guidelines manual[169]).

The main considerations specific to each recommendation are outlined in the 'Recommendations and link to evidence' sections within each chapter.

4.5.1 Research recommendations

When areas were identified for which good evidence was lacking, the GDG considered making recommendations for future research. Decisions about the inclusion of a research recommendation were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

4.5.2 Validation process

This guidance is subject to a 6-week public consultation and feedback as part of the quality assurance and peer review of the document. All comments received from registered stakeholders are responded to in turn and posted on the NICE website.

4.5.3 Updating the guideline

Following publication, and in accordance with the NICE guidelines manual, NICE will undertake a review of whether the evidence base has progressed significantly to alter the guideline recommendations and warrant an update.

4.5.4 Disclaimer

Healthcare providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may

not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Clinical Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of this guideline and the literature used in support of this guideline.

4.5.5 Funding

The National Clinical Guideline Centre was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline.

5 People who may benefit from an approach to care that takes account of multimorbidity

5.1 Introduction

Multimorbidity is most commonly defined simply as having 2 or more long term conditions. While this type of definition may be helpful for research purposes it is not necessarily helpful when providing clinical care. The guideline development group considered that it was important to provide more detail about those patients for whom the recommendations in the guideline would be most helpful. The guideline committee therefore worked as a group to consider how they wished to define and describe the population for the guideline. This chapter provides a summary of their discussions and the information they used.

5.2 Defining multimorbidity

Members of the GDG were aware of the existence of a number of different definitions of multimorbidity. The guideline scope includes two definitions (1) the co-existence of 2 or more long term conditions and (2) the combination of 1 chronic disease with at least 1 other disease or bio psychosocial factor or somatic risk factor. The latter definition was considered too broad to be useful, and the GDG recognised that defining multimorbidity by simple counts of any kind was also unlikely to be helpful. This is because very large numbers of people have multimorbidity defined as two or more long term conditions (16-58% of adults in the UK, depending on how many conditions are included in the count), but for many such people their multimorbidity will present them few problems in their life (for example, someone with well-controlled asthma and modest hyperlipidaemia) or in the organisation of their healthcare (for example, someone with type 2 diabetes, hypertension and hay fever). The GDG considered that rather than adjudicate between different definitions of multimorbidity they needed to focus on a pragmatic definition of the broad target population for the guideline. The guideline is targeted towards people with multiple conditions where these present significant problems to everyday functioning or where the management of their care has become burdensome to the patient and/or involves a number of services working in an uncoordinated way. From this perspective, the problems faced by patients may be due to the severity or nature of their conditions, but commonly relates to the organisation of the healthcare system and their interaction with it.

5.3 Recommendations and link to evidence

Recommendations	**General principles** **1. Be aware that multimorbidity refers to the presence of 2 or more long-term health conditions, which can include:** • **defined physical and mental health conditions such as diabetes or schizophrenia** • **ongoing conditions such as learning disability** • **symptom complexes such as frailty or chronic pain** • **sensory impairment such as sight or hearing loss**

	• **alcohol and substance misuse.** **2. Be aware that the management of risk factors for future disease can be a major treatment burden for people with multimorbidity and should be carefully considered when optimising care.** **3. Be aware that the evidence for recommendations in NICE guidance on single health conditions is regularly drawn from people without multimorbidity and taking fewer prescribed regular medicines.** **4. Think carefully about the risks and benefits, for people with multimorbidity, of individual treatments recommended in guidance for single health conditions. Discuss this with the patient alongside their preferences for care and treatment.** **<u>Taking account of multimorbidity in tailoring the approach to care</u>** **5. Consider an approach to care that takes account of multimorbidity if the person requests it or if any of the following apply:** • **they find it difficult to manage their treatments or day-to-day activities** • **they receive care and support from multiple services and need additional services** • **they have both long-term physical and mental health conditions** • **they have frailty (see Chapter 8) or falls** • **they frequently seek unplanned or emergency care (see also recommendation 9)** • **they are prescribed multiple regular medicines (see Chapter 6).**
Other considerations	The GDG considered it important that readers of the guideline should be given clear direction as to how the guideline should be used and the population for whom it was intended. They therefore developed a recommendation to indicate that the guideline should be used in conjunction with single disease guidelines. The GDG considered that the information in single disease guidelines can be relevant and appropriate to many people with multimorbidity and that this guideline was not intended to supplant those guidelines but to help healthcare professionals consider how to best to implement those guidelines considering the needs of people with multimorbidity. The populations included in evidence that informs single disease guidelines is likely to come from people who are younger and fitter and are taking fewer medicines than with multimorbidity. The important point the GDG wished to emphasise was the importance of thinking carefully about the needs of the person with multimorbidity particularly when they are already taking multiple medicines, and explicitly deciding whether single disease guideline recommendations were relevant to that individual. The risks and benefits of recommended treatments are likely to be different for people with multimorbidity than people without multimorbidity. The risk and benefits should be part of the discussion alongside people preferences for treatment. The GDG agreed that a basic definition of multimorbidity as 2 or more long term conditions (LTC) was a straightforward place to start. They accepted the English Department of Health definition of a LTC as 'a condition that cannot, at present be cured, but can be controlled by medication and other therapies' as a starting point although this definition has been described as outdated by the House of Commons

Select Committee, with not enough emphasis on the patient and their preferences and priorities. The use of the term 'condition' however is helpful as it allows the definition to include symptom complexes which are not easily or currently classified as diseases, or management of significant risk factors which may be a major part of treatment burden for many people but again not classified as disease. The GDG however agreed to include a recommendation at the start of the list of recommendations to ensure there was clarity about what consititued a long term condition. They considered that when people had defined medical diagnoses it was relatively straightforward to ensure they were considered to be people with multimorbidity but that ongoing conditions such as learning disability, frailty or chronic pain and longterm sensory impairments were also relevant.

In modern medical care people preventative treatments are very common and the GDG wished to remind healthcare pratitioners that these treatments may also cause a significant treatment burden for patients and could be considered a 'condition'.

The GDG considered that there are many people with multimorbidity with excellent quality of life and everyday functioning, whose self-management and professional care is straightforward. This group do not need a bespoke approach to care and are not the target of the guideline. The GDG agreed that ageing is not itself a condition to be included, although some of the common consequences of ageing such as frailty will be. Although multimorbidity becomes more common with age, people of any age can have the need for a bespoke approach to their care.

The GDG considered that both complexity of care and complexity of conditions are likely to influence whether someone may benefit from an approach to care that takes account of multimorbidty.

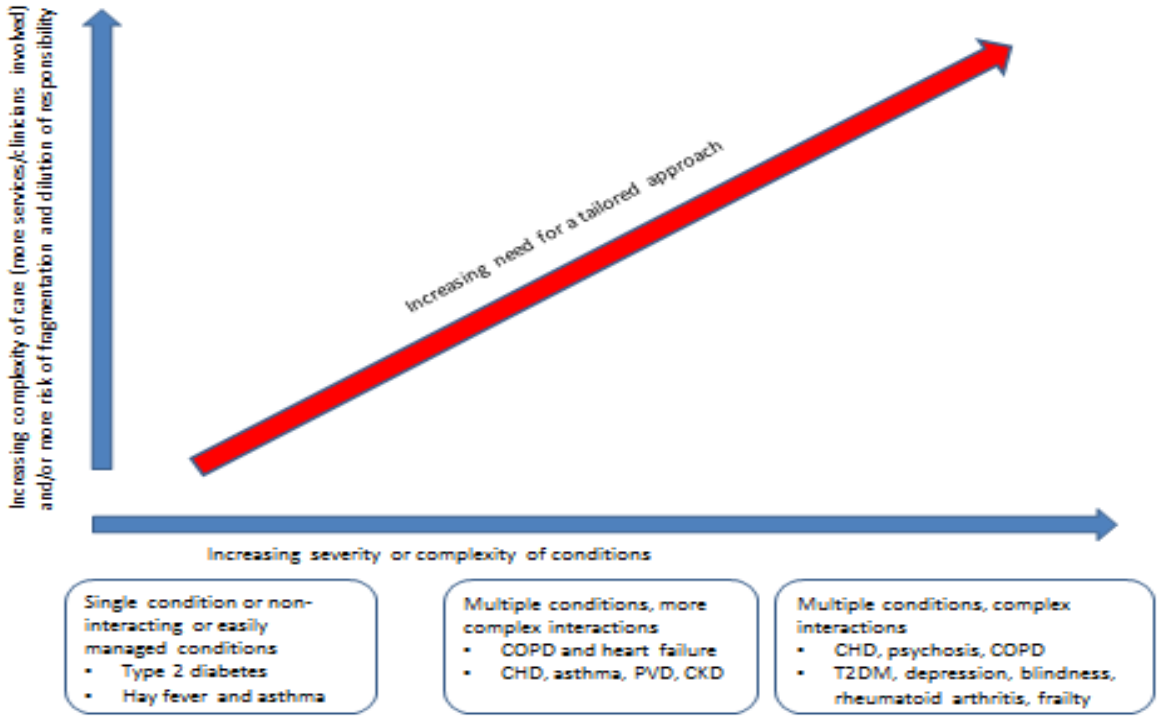
The epidemiology of multimorbidity indicates that as people age they have a larger number of conditions and increased polypharmacy. This commonly occurs because each condition is treated separately from others according to single condition guidelines where recommendations focus on optimising care for the condition focused, with little consideration of the wider context. This focus at least partly reflects that the evidence on which recommendations are based is also largely drawn from relatively younger, fitter, less multimorbid and less co-prescribed people. However, depending on the person, the conditions they have and the recommended treatments, the cumulative impact of individually rationale single condition recommendations may be irrational because of the risk of harms from interactions between treatments in the face of polypharmacy, and between treatments and conditions in the face of multimorbidity, and because of the development of burdensome treatment regimens. The GDG noted that single condition guidelines are explicit that clinicians should not blindly follow recommendations for all patients because treatment decisions should always be made in the context of an individual's circumstances.

The GDG considered that some conditions may of themselves make it likely that multimorbidity may become problematic such as when people have dementia and depression in combination with physical illness. Depression can be common in people with long term conditions and may be present in many younger people, particularly those living in less affluent areas who develop multimorbidity 10-15 years earlier than the more affluent. The onset and progression of dementia can make management of care more difficult and a careful analysis of benefit and harm is required. Personal characteristics such as frailty will be associated with reduced resilience and tolerance of problems with care. Similarly individual social and psychological factors may increase complexity.

Depending on the conditions that a person has, care can also become increasingly fragmented. However, definitive statements about numbers and types of conditions and associations with complex care are difficult – a person with Parkinson's disease for example may have one condition but has a multi-system disease that requires input from and attendance at multiple specialist appointments. Individual discussions and judgement will therefore always be required. The GDG believed however that in general, as shown in Figure 3, an approach to care that takes

	account of multimorbidity was more likely to be of benefit to the person as one or both of the complexity of care and the complexity of conditions increased. The GDG considered that deciding when a multimorbidity approach to care would be of benefit would always be a matter of judgement, made in collaboration by the patient and the professional. The GDG used these considerations to develop a suggested list of people for whom a multimorbidity approach to care may be appropriate. As well as people with physical and mental health problems they included people who express difficulty in managing their conditions and treatments, people who already require input from multiple services particularly if the addition of further services is being considered. There are a number of potential indicators that people may have an increased burden of disease such as being prescribed multiple regular medicines, experiencing unplanned care frequently or there is evidence of reduced resilience such as frailty. These are discussed in more detail in Chapters 7 and 9.

Figure 3: Diagram developed by GDG to indicate need for an approach to care that takes account of multimorbidity



6 Principles of an approach to care that takes account of multimorbidity

6.1 Introduction

There are multiple potential combinations of conditions, treatments and personal circumstances that inform people's experience of multimorbidity and its management. In such circumstances guideline principles for care are required rather than rules which are applicable to all. To inform the development of the guideline and the recommendations two separate but related reviews were carried out. The first sought to consider what principles were important in care for people with multimorbidity; and the second examined barriers to providing good quality care to people with multimorbidity.

6.2 Principles of care

6.2.1 Review question: What principles are important for assessing, prioritising and managing care for people with multimorbidity?

For full details see review protocol in Appendix C.

Table 14: Characteristics of review question

Objective	To identify key principles that healthcare professionals should consider when assessing, prioritising and managing care for people with multimorbidity
Population and setting	Adults (aged 18 years and over) with multimorbidity; healthcare professionals treating adults with multimorbidity
Themes	Themes will be identified from the literature and not specified by the GDG in advance. However, for guidance for the technical team, relevant topics may include: • Communication between healthcare professionals and people with MM • Things clinicians should be aware of when treating people with MM; for example, risk of diagnostic overshadowing, risk that side effects of medications may be misinterpreted (prescribing cascade) • Assessing appropriateness of treatments (for example, awareness of possible interactions between treatments, likelihood that the person will experience benefit) • Providing a holistic or generalist approach • Methods for eliciting a person's preferences or wellbeing • Communicating expected benefits or harms of treatment • Best practice for discontinuation of pharmacological treatments
Review strategy	Databases: Medline, Embase, the Cochrane Library Date: All years Language: Restrict to English only Study designs to be considered: Guidelines and other grey literature that provide guidance for healthcare professionals on the assessment, prioritisation and management of care for people with multimorbidity Evidence quality to be assessed using AGREE II

6.2.2 Clinical evidence

The GDG were aware that many organisations and expert groups had already considered the care of people with multimorbidity specifically, or the area of care likely to be important for people with multimorbidity. They agreed therefore to search at the level of guidelines and policy documents to inform the guideline.

Nine guidelines were included in the review [8,10,97,155,160,167,170,172,178] and these are summarised in Table 15 below. Key findings from these studies are summarised in the clinical evidence summary tables (Tables 17-23) below. See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, and excluded studies list in Appendix L.

We searched for guidelines that sought to generate recommendations specifically for the assessment, prioritisation and management of care for populations that were predominantly multimorbid, or recommendations that were deemed to be pertinent to people with multimorbidity. We prioritised guidelines that were produced by professional organisations and where the guideline committee included clinicians from multiple specialties, and preferably also included patient members. We did not include recommendations reported in editorials or commentaries by independent research groups where these were not supported by a systematic review of the evidence. We did not include systematic reviews where the primary aim of the review was not to derive recommendations for clinical practice.

As guidelines were often wide-ranging and not exclusively about multimorbidity, some selectivity had to be applied when extracting recommendations. An inclusive approach was selected to provide the GDG with the most information possible by extracting all recommendations that the one reviewer felt was pertinent to the clinical care of people with multimorbidity.

Four guidelines included a broad range of recommendations on care in the NHS. Three guidelines were sub-populations of people with multimorbidity; 1 for people with depression and co-existing physical health conditions, 1 for people with cardiovascular disease and co-morbidities, and 1 for the care of older adults with multimorbidity. One guideline was focused specifically on the handling of multimorbidity in primary care. The final guideline was focused on the formulation of guidelines to include those with multimorbidity.

The principles extracted from the included guidelines were divided into themes and sub-themes. Wherever possible the original themes used in the guidelines were maintained. Some principles related to multiple themes and in this situation they were included under the dominant theme.

6.2.2.1 Summary of included guidelines

Table 15: Summary of guidelines included in the review

Guideline (developed by)	Population	Objective	Themes	Comments
AHA/ACC/HHS strategies to enhance application of clinical practice guidelines in people with cardiovascular disease and comorbid conditions (2014) (American Heart	US people with cardiovascular disease and co-morbid conditions	To identify core principles for CPGs (clinical practice guidelines) in the effective management of people with multiple chronic conditions and related actions that might be taken by developers of CPGs	Need for research, Guideline development,	No search for evidence was conducted, panel discussion by physicians only without any other disciplinary input

Association, American College of Cardiology and US Department of Health and Human Services)				
Depression in adults with a chronic physical health problem (2009)(NICE)	UK NHS people with depression and a chronic physical health problem	To make recommendations for the treatment and management of depression in adults with a chronic physical health problem	Principles of assessment, Effective delivery of care for depression, Collaborative care,	NICE methodology with systematic searches
Guiding principles for the care of older adults with MM (2012) (American Geriatric Society)	US older adults with multiple chronic conditions	To present the guiding principles for the clinical management of older adults with MM	A person's preferences, Interpreting the evidence, Prognosis, Clinical feasibility, Optimising therapies,	Panel discussion supported by a non-systematic review of the evidence, funded by American Geriatric Society with potential for bias
IOM and DHHS meeting on making clinical practice guidelines appropriate for people with multiple conditions (2014) (US Department of Health & Human Services, Institute of Medicine)	USA people with multiple chronic conditions	To identify guiding principles for clinical guidelines in the effective management of multiple chronic conditions and identifying actions that should be taken by developers and users of guidelines for people with multiple chronic conditions	Improving stakeholder process, Strengthening substance, Increase focus on patient-centeredness,	Panel discussion lacking patient representation or allied health professionals no search for evidence
Medicines adherence (2009) (NICE)	UK NHS adult patients	To provide recommendations to clinicians and others on how to involve adults and carers in decisions about prescribed medicine	Patient involvement, Supporting adherence, Reviewing medicines, Communication between healthcare professionals	NICE methodology with systematic searches and consensus recommendations when evidence insufficient. Wide ranging guideline with some aspects of particular relevance for people with MM.
Medicines	UK NHS adult	To review the	Identifying	NICE methodology

optimisation (2015) (NICE)	patients	evidence available to support health and social care practitioners, and health and social care organisations, in considering the systems and processes required to ensure safe and effective medicines optimisation	incidents, Medicines-related communication systems for transitions, Medication review, Self-management plans, Patient decision aids,	with systematic searches. Wide ranging guideline with some aspects of particular relevance for people with MM.
Patient experience in adult NHS services (2012) (NICE)	UK NHS adult patients	To provide the NHS with clear guidance on the components of a good patient experience	Knowing the person as an individual, Individualised services, Continuity of care, Patient autonomy, Discussing risks & benefits,	NICE methodology with systematic searches and development of domains of a person's experience. Wide ranging guideline with some aspects of particular relevance for people with MM.
Polypharmacy guidance (2012) (NHS Scotland)	UK NHS people on multiple medications or "frail" in a medical sense	To provide information about patient groups that NHS boards should consider as a priority for polypharmacy review, an outline of medication review process in these people and provide NHS boards with tools to be adapted for local guideline use	Reviewing medicines, High risk medication.	No systematic search for evidence described. Best practice guideline with references highlighting issues.
The Ariadne principles: how to handle multimorbidity in primary care consultations (2014) (International symposium of primary care physicians)	People from North America, Europe, Australia, with multimorbidity in primary care	To develop a set of principles for handling multimorbidity in primary care consultations	Interaction assessments, Prioritisation & patient preferences, Individualised management,	No evidence search conducted but describes a well detailed semiformal consensus approach with many opportunities for feedback from primary care physicians.

6.2.2.2 Evidence

Seven themes were identified across the included guidelines; these are presented in Table 16 below. Where possible, the terminology of themes used in the guidelines was preserved.

Table 16: Themes and sub-themes

Main theme	Sub-themes	Statement of finding
Patient centred care	Patient priorities Patient involvement	Involving people in decisions around their care and taking account of their priorities when dealing with risks and benefits is particularly important for people with multimorbidity.
Interpreting and discussing the evidence	Discussing evidence Supporting decisions	When discussing evidence around treatments with people with multimorbidity, consider using decisions aids if available and be aware of the best approaches when presenting evidence.
Medication management	Medication review Supporting adherence	Adherence is a common and complex issue, medication reviews can include consideration of adherence support as well as the stopping and starting of treatments.
Optimising care plans	No sub-themes	Complex care plans for people with multimorbidity must take into account more than simple single disease treatments and assessments
Communication between healthcare professionals	No sub-themes	Communication between healthcare professionals is particularly important when caring for people with multimorbidity as they are likely to be accessing may different services.
Guideline development	No sub-themes	Single condition guideline specific care may not be appropriate for people with multimorbidity, due to the potential interactions between diseases and drugs as well as the total treatment burden incurred.
Need for research	No sub-themes	People with multimorbidity are often excluded from clinical trials; there is a relative paucity of evidence in this population.

6.2.2.2.1 *Patient centred care*

Several guidelines (n=5) published recommendations on the way clinicians should ensure that their care is patient centred, particularly for older adults with multimorbidity and when dealing with multimorbidity in primary care. This included recommendations on promoting the involvement of people in their care (sub-theme: patient involvement), and how to identify and use a person's priorities in guiding their care (sub-theme: patient priorities).

Table 17: Patient centred care

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
Sub-theme 1: Patient involvement		
3	Medicines adherence; Medicines optimisation; Guiding principles for the care of older adults with MM	Subtheme synopsis People with multimorbidity should be encouraged to be involved with their care at their preferred level. Healthcare professionals should facilitate this involvement of people (and their family or carers as appropriate) by optimising the level and type of information they provide. Healthcare professionals should be aware that the consultation skills required for this facilitation can be improved. Healthcare professionals should explain the medical aims of treatment and clarify what the person hopes their treatment will achieve. Healthcare professionals should be aware that people may have differing views around their care and that encouraging their involvement in decisions may lead to them prioritising the stopping of treatment. Recommendations from included guidelines • Healthcare professionals should adapt their consultation style to the needs of individual people so that all people have the opportunity to be involved in decisions about their medicines at the level they wish • Offer all people the opportunity to be involved in making decisions about prescribed medicines. Establish what level of involvement in decision making the person would like. • Explore a person's preferences about the level and type of information they want. Based on this, give the person (and their family members and carers if appropriate) clear, consistent, evidence-based, tailored information throughout all stages of their care. • Avoid making assumptions about a person's preferences about treatment. Talk to the person to find out their preferences, and note any non-verbal cues that may indicate you need to explore the person's perspective further.

Study design and sample		Descriptors of themes
		• Accept that people may have different views from healthcare professionals about the balance of risks, benefits and side effects of medicines. • Encourage people to ask about their condition and treatment. • Be aware that the consultation skills needed for increasing patient involvement can be improved. • Discuss with the person why they might benefit from the treatment. Clearly explain the disease or condition and how the medicine will influence this. • Explain the medical aims of the treatment to people and openly discuss the pros and cons of proposed medicines. The discussion should be at the level preferred by the person. • Clarify what the person hopes the treatment will achieve. • Healthcare professionals have a duty to help people to make decisions about their treatment based on an understanding of the likely benefits and risks rather than on misconceptions. • Be aware that increasing patient involvement may mean that the person decides not to take or to stop taking a medicine. If in the healthcare professional's view this could have an adverse effect, then the information provided to the person on risks and benefits and the person's decision should be recorded. • Encourage and support the person, families and carers to keep an up to date list of all medicines the person is taking. The list should include the names and dosages of prescription and non-prescription medicines and herbal and nutritional supplements. If the person has any allergic or adverse reactions to medicines, these should be noted. • Be aware that people may wish to minimise how much medicine they take • Be aware that people may wish to discuss what will happen if they do not take the medicine suggested by their healthcare professional, non-pharmacological alternatives to medicines, how to reduce and stop medicines they may have been taking for a long time, particularly those known to be associated with withdrawal symptoms, how to fit taking the medicine into their daily routine and how to make a choice between medicines if they believe they are taking too many medicines. • When discussing medicines with people who have chronic or long-term conditions, consider using an individualised self-management plan to support people who want to be involved in managing their medicines.
Sub-theme 2: Patient priorities		
4	Medicines adherence; Patient experience; Guiding	Subtheme synopsis

Study design and sample		Descriptors of themes
	principles for the care of older adults with MM; The Ariadne principles: how to handle multimorbidity in primary care consultations	Having ensured the person is sufficiently informed, health care professionals should recognise when patient priorities are particularly relevant, support a person to communicate them, not allow their own priorities to influence the person's values and preferences and regularly review these as they are liable to change. In discussing a person's priorities, healthcare professionals should define treatment goals in terms of time to aid the process of regular review. Recommendations from included guidelines Treatment goals should be defined in terms of time, this clarification will support monitoring and re-discussing priorities at appropriate time points. A person's prognosis should always be taken into consideration. Healthcare decisions need to be made on a background of the person's values and preferences, these should be thoroughly elucidated and treatment goals agreed upon as a consequence. People may prioritise desired outcomes or the avoidance of negative outcomes. Family physicians should be aware of their own potentially differing preferences. Recognise when the older adult with multimorbidity) is facing a "preference sensitive" decision. Ensure that older adult with multimorbidity are adequately informed about the expected benefits and harms of different treatment options. Elicit preferences only after the older adult with multimorbidity

6.2.2.2.2 Interpreting and discussing the evidence

Two guidelines included recommendations on how clinicians should discuss the risks and benefits of treatments. This includes thinking about whether the evidence is relevant and how people should be informed about risks and benefits (sub-theme: discussing evidence), and how patient-decision aids might be used to support shared decision-making around the management of care (sub-theme: supporting decisions).

Table 18: Interpreting and discussing the evidence

Study design and sample		Descriptors of themes
No. of guideli	Guideline(s)	

Study design and sample		Descriptors of themes
nes		
Sub-theme 1: Discussing the evidence		
2	Guiding principles for the care of older adults with MM; Patient experience	Subtheme synopsis Healthcare professionals should refer to clinical evidence to inform the care of people with multimorbidity. Healthcare professionals should consider the quality and applicability of a body of clinical evidence before discussing its implications with a person. When healthcare professionals do discuss evidence with people with multimorbidity, they should present it in an accessible manner including using relative risks in conjunction with absolute risks, use consistent natural frequencies and not percentages, present risk over a defined period of time, be aware that different people interpret terms like 'rare' or 'common' differently, use both positive and negative framing, personalise risks and benefits as much as possible, and consider using graphs or icons to support numerical data. Recommendations from included guidelines Question whether a study is applicable to the population in question. Consider the quality of a study (for example, RCT vs NRS) and prefer reviews of multiple studies. Consider whether the outcomes reported are clinically important and important to patients. Consider the balance between any benefits and the harms incurred including the burden required to commit to treatment. Always consider the baseline risk not just a relative risk change, that is, ARR is more useful than RRR. NNT and NNH data should be interpreted in conjunction with time factors, clinicians should look for a time horizon to benefit or harm (that is, the length of time needed to accrue an observable clinically meaningful benefit or harm). Use absolute risk rather than relative risk. Use natural frequency rather than a percentage (for example, 10 in 100 not 10%). Be consistent in the use of data (for example, 1 in 100 vs 10 in 100, not 1 in 100 vs 1 in 10). Present a risk over a defined period of time. Be aware that different people interpret terms such as rare, unusual and common in different ways, and use numerical data if available. Think about using a mixture of numerical and pictorial formats. Include both positive and negative framing. Personalise risks and benefits as far as possible. Offer support to the person when they are considering options. Use the principles of shared decision making, that the person is aware

Study design and sample		Descriptors of themes
		of the options available, understands the risks, benefits and consequence of these, that the person understands the information and encourage the person to clarify what is important to them and check their choice is consistent with this.
Sub-theme 2: Supporting decisions		
1	Medicines optimisation	Subtheme synopsis Healthcare professionals should make use of all available resources to support the decision making of people with multimorbidity. Where available, reliable decision aids should be offered to people by those trained in their use, healthcare professionals should be aware of the limitations of decision aids and the need to adjust discussions according to a person's baseline risk. People should be given plenty of time to make these decisions and the aids should not replace consultations but help them. It may be appropriate for more than one consultation to ensure a person can make an informed decision. Recommendations from included guidelines Offer all people the opportunity to be involved in making decisions about their medicines. Find out what level of involvement in decision-making the person would like and avoid making assumptions about this. Find out about a person's values and preferences by discussing what is important to them about managing their conditions and their medicines. Recognise that the person's values and preferences may be different from those of the health professional and avoid making assumptions about them. Apply the principles of evidence based medicine when discussing the available treatment options with a person in a consultation about medicines. Use the best available evidence carefully when making decisions together with clinical expertise and the person's values and preferences. In a consultation about medicines, offer the person the opportunity to use a patient decision aid (when 1 is available) to help them make a preference-sensitive decision that involves trade-offs between benefits and risks. Ensure the patient aid is appropriate in the context of the consultation as a whole. Do not us a patient decision aid (PDA) to replace discussions with a person in a consultation about medicine. Recognise that it may be appropriate to have more than 1 consultation to ensure that a person can make an informed decision about their medicines. Give people the opportunity to review their decision as appropriate. Ensure that PDAs have followed a robust and transparent development process, in line with International Patient Decision Aid Standards (IPDAS) criteria. Before using a PDA, read and understand its content paying particular attention to its limitations and the need to adjust discussions

Study design and sample		Descriptors of themes
		according to the person's baseline risk. Have the necessary skills and knowledge when using a PDA including clinical knowledge, communication skills, numeracy skills, ability to explain the trade-off between benefits and risks. Consider training and education to support healthcare professionals and patients in developing the skills to use PDAs.

6.2.2.2.3 Medication management

Several guidelines (n=5) published recommendations on the way clinicians should conduct medication management with people, particularly older adults with multimorbidity. This included general recommendations on when and how a review should be conducted (sub-theme: medication review), and how to recognise, assess and support people struggling with adherence (sub-theme: supporting adherence).

Table 19: Medication management

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
Sub-theme 1: Medication review		
4	Medicines optimisation; Medicines adherence; Polypharmacy guidance; Guiding principles for the care of older adults with MM	Subtheme synopsis Medication reviews should be triggered by certain events or situations (for example hospital admission or discharge), they should be conducted by appropriately trained personnel and be personalised. Healthcare professionals should be aware that people may alter their own medication (for example stopping or altering the dose) and ask about this. Healthcare professionals should ask about a person's particular concerns during medication review and consider all over the counter and complementary medicine. Medication that is deemed of little continued benefit should be slowly tapered unless appropriate advice exists to the contrary, in particular healthcare professionals should consider specialist input before suggesting stopping if a medication is preventing rapid symptomatic deterioration or fulfilling an essential replacement function. Stopping of medication should be done one at a time ideally and is generally guided by little evidence. Medication reviews should be regularly repeated as a person's preferences and situation is likely to change over time.

Study design and sample		Descriptors of themes
		Recommendations from included guidelines Review a person's knowledge, understanding and concerns about medicines, and a person's view of their need for medicine at intervals agreed with the person, because these may change over time. Offer repeat information and review to people, especially when treating long term conditions with multiple medicines. Review at regular intervals the decision to prescribe medicines, according to the person's choice and need. Medication should be reviewed regularly. Medication appropriateness should be evaluated at hospital admission, ICU admission and hospital discharge. Be aware that people sometimes evaluate prescribed medicines using their own criteria such as their understanding of their condition or the symptoms most troubling to them. They may, for example, stop and start the medicine or alter the dose and check how this affects their symptoms. Ask the person whether they have done this. Consider using a screening tool (for example, STOPP/START) to identify potential medicines related patient safety incidents in some patient groups, including those with polypharmacy or chronic conditions. Determine locally the most appropriate health professional to carry out a medication review, based on their knowledge and skills, including technical knowledge of medicine managing processes, therapeutic knowledge and effective communication skills. During a medication review, take into account the person's understanding about their medicines, their concerns about their medicines, all over the counter and complementary medicines, how safe & effective their medicines are and any monitoring tests that are needed. Medication should ideally be stopped 1 at a time. Little evidence exists to guide stopping of medications and if there is uncertainty it is sensible to use a tapering regimen when stopping drugs. People with a 40-60% risk of emergency admission within the next 12 months (as per Scottish Patients at Risk of Readmission and Admission (iSPARRA)), on multiple medicines from 10 or more particular BNF sections and high risk medicines, reviews should be started on people >75 years. Question whether each prescription is preventing rapid symptomatic deterioration or fulfilling an essential replacement function as these should be continued or only discontinued with specialist input. For medicines without clear essential indications or contraindications, check their effectiveness in the specific patient group against a reference summary (version included in guideline – based on NNTs in specific situations). High risk combinations should be avoided unless completely necessary, these combinations include: NSAID + ACEi/diuretic, NSAID + tricyclic antidepressant/glitazone, warfarin + antiplatelet drug/macrolide/NSAID/quinolone. PPIs and H2 antagonists should be considered for reduction particularly if antibiotics are required due to the increased risk of

Study design and sample		Descriptors of themes
		C.difficile. When using diuretics for ankle oedema consider alternative ways to manage the oedema particularly if there is medication causes (for example, calcium channel blockers). Consider stopping or reducing dose of digoxin if being used in presence of CKD. Review combinations of antidepressants such as tricyclic antidepressants for analgesia used in combination with other antidepressants for depression. In general SSRIs are better tolerated in people with dementia who also have depression. Consider cumulative GI effects when co-prescribing SSRIs & NSAIDs/aspirin. Use metformin with caution in renal impairment and avoid if eGFG <30 ml/min. When an antidepressant is to be prescribed for a person with depression and a chronic physical health problem, take into account the presence of additional physical health disorders, the side effects of the antidepressants which may impact on the physical health disorders, that there is no evidence supporting the use of specific antidepressants for people with particular physical health problems and interactions with other medicines.
Sub-theme 2: Supporting adherence		
2	Medicines adherence; The Ariadne principles: how to handle multimorbidity in primary care consultations	Subtheme synopsis Healthcare professionals should be aware of risk factors (for example complex medication regimens) and triggers for non-adherence and routinely assess it in a non-judgmental manner. Causes, either intentional (for example a person's beliefs and concerns around a medication) or unintentional (practical problems), should be identified and healthcare professions should discuss possible solutions with patients Healthcare professionals should be aware that no specific intervention can be recommended to address adherence for all people and any intervention should be tailored to the situation. Recommendations from included guidelines Recognise that non-adherence is common and that most people are non-adherent sometimes. Routinely assess adherence in a non-judgemental way whenever you prescribe, dispense and review medicines. Consider assessing non-adherence by asking the person if they have missed any doses of medicine recently. Make it easier for them to report non-adherence by asking the question in a way that does not apportion blame, explaining why you are asking the question, mentioning a specific time period such as 'in the past week', asking about medicine-taking behaviours such as reducing the dose,

Study design and sample		Descriptors of themes
		stopping and starting medicines. Consider using records of prescription re ordering, pharmacy patient medication records and return of unused medicines to identify potential non-adherence and people needing additional support. If a person is not taking their medicines, discuss with them whether this is because of beliefs and concerns or problems about the medicines (intentional non-adherence) or because of practical problems (unintentional non-adherence). Be aware that although adherence can be improved, no specific intervention can be recommended for all people. Tailor any intervention to increase adherence to the specific difficulties with adherence the person is experiencing. Complex medication regimens should trigger awareness of increased risk of reduced adherence.

6.2.2.2.4 Optimising care plans

Two guidelines published recommendations on general ways clinicians should optimise care plans for people with multimorbidity, particularly in primary care and in older adults.

Table 20: Optimising care plans

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
2	The Ariadne principles: how to handle multimorbidity in primary care consultations; Guiding principles for the care of older adults with MM	Subtheme synopsis Complex care plans for people with multimorbidity must take into account more than simple single disease treatments and assessments. Healthcare professions should assess for possible interactions between diseases, between medications, and between diseases and medications. When formulating care plans, healthcare professionals should offer to discuss prognosis with people, but be aware that they may not wish to do, and may consider prioritising treatment based on a person's anticipated life expectancy. Healthcare professionals should take into account a person's social participation, functional autonomy, coping strategies and health seeking behaviour when optimising a care plan in discussion with the person themselves.

Study design and sample		Descriptors of themes
		Recommendations from included guidelines It is helpful to prioritise decisions based on life expectancy so they are categorised as short term (within the next year), midterm (within the next 5 years) or long term (beyond 5 years). Clinicians should offer to discuss prognosis but not all older adults with multimorbidity may wish to do so. In contrast to people with a single disease, interactions rather than single diseases need assessment. These include drug-drug, drug-disease and disease-disease interactions. It is important to keep a list of all individual diagnoses and to assess impact on quality of life and functioning. A list of other physicians and therapists should be kept and updated regularly. Active monitoring for signs and symptoms of psychological disorders, cognitive dysfunction and deleterious social circumstances that may influence care seeking, is vital. A person's social participation, functional autonomy, coping strategies and health seeking behaviour should be elicited and considered. When assessing a person with a chronic physical health problem who may have depression, conduct a comprehensive assessment that does not rely simply on a symptom count. Take into account the degree of functional impairment. When providing interventions for people with a learning disability or acquired cognitive impairment who have a chronic physical health problem and a diagnosis of depression, provide the same interventions as for other people with depression where possible but if necessary adjust the method of delivery or duration of the intervention to take account of the disability or impairment. If a person's chronic health problem restricts their ability to engage with a preferred psychosocial or psychological treatment for depression consider alternatives in discussion with the person, such as antidepressants or delivery of psychosocial or psychological interventions by telephone if mobility or other difficulties prevent face to face contact. The first step is to identify treatments that may be inappropriate in older adult with multimorbidity consensus statements and expert derived criteria exist to identify these potentially inappropriate medications (PIMs) and should be consulted.

6.2.2.2.5 Communication between healthcare professionals

Several guidelines (n=4) published recommendations on ways that communication between healthcare professionals should be optimised to ensure best care for people, particularly those with multimorbidity being seen in primary care.

Table 21: Communication between healthcare professionals

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
4	Medicines adherence; Medicines optimisation; Patient experience; The Ariadne principles: how to handle multimorbidity in primary care consultations	Subtheme synopsis Organisations and healthcare professionals should be aware of the many services a person with multimorbidity may need to use, which may impact on the person. Healthcare professionals should ensure that communication between services is efficient, confidential and pro-active where possible. Effective communication is particularly important regarding prescribing decisions, where changes are made to a person's medication healthcare professionals should inform the original prescriber where possible. The aim of organisations and healthcare professionals should be to ensure effective co-ordination of care such that the impact of utilising multiple services on the person is minimised. Organisations should consider co-ordinating additional support for some people with multimorbidity when they have been discharged from hospital, for example pharmacist counselling, telephone or home follow-up. Healthcare professionals should consider collaborative care for people with moderate to severe depression and a chronic physical health problem. Recommendations from included guidelines For people who use a number of different services ensure effective co-ordination and prioritisation of care to minimise the impact on the person. Ensure clear and timely exchange of patient information between healthcare professionals and between healthcare and social care professionals. Healthcare professionals involved in prescribing, dispensing or reviewing medicines should ensure that there are robust processes for communicating with other healthcare professionals involved in the person's care. Organisations should ensure that information about medicines is shared with the person and their GP; they should identify when local

Study design and sample		Descriptors of themes
		systems are in place for this and take account of HSCIC's guide to confidentiality. Consider collaborative care for people with moderate to severe depression and a chronic physical health problem with associated functional impairment whose depression has not responded to initial high-intensity psychological interventions, pharmacological treatment or a combination of psychological and pharmacological interventions. Collaborative care should normally include supervised case management with support from a senior mental health professional, close collaboration between primary and secondary physical health services and specialist mental health services, a range of interventions consistent with latest guidelines and long term co-ordination of care and follow-up. Healthcare professionals involved in reviewing medicines should inform the prescriber of the review and its outcome. This is particularly important if the review involves discussion of difficulties with adherence and further review is necessary. Organisations should ensure that robust and transparent processes are in place so that when a person is transferred to another setting complete and accurate information about medicines is shared, received, document and acted on. Organisations should consider additional support for some patient groups (including those with polypharmacy or chronic conditions) when they have been discharged from hospital, for example, pharmacist counselling, telephone follow up, GP and or nurse home visits.

6.2.2.2.6 Guideline development

Two guidelines published recommendations on ways that guideline development can be improved to provide better care for people with multimorbidity, particularly those with cardiovascular disease and co-morbidities.

Table 22: Guideline development

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
2	AHA/ACC/HHS strategies to enhance application of clinical practice guidelines in people	Subtheme synopsis Guidelines must proactively integrate care for those with co-morbidities into their content. They should provide information on common co-morbidities to their index condition, and prompt healthcare professionals to consider these, while also openly addressing

Study design and sample		Descriptors of themes
	with cardiovascular disease and comorbid conditions; IOM and DHHS meeting on making clinical practice guidelines appropriate for people with multiple conditions	the lack of knowledge around co-morbidity and its impact on guideline feasibility. Guidelines should explicitly discuss the applicability and quality of recommendations for the most frequent co-morbidities that accompany their index condition. Guideline development panels should include expert representation for conditions other than the index condition. Guidelines should be patient-centred rather than focused solely on the management of single conditions in isolation. Recommendations from included guidelines In light of the paucity of evidence around MM, CPGs need to be nuanced to account for clinical judgement and acknowledge the role of individualised, patient-centred decision making in implementation. Organisations that develop CPGs must now consider comorbidities in the development process. Involving people in the CPG development process is critically important to fully appreciate a person's perspectives; this becomes even more important when dealing with MM. CPGs should explicitly discuss the applicability and quality of recommendations for the most frequent combinations of comorbidities that accompany the named condition. Guideline development should harmonize co-morbidity related content across guidelines created by different groups. Guideline development panels should include appropriate expert representation for conditions other than the index condition. Guidelines should take into account factors associated with adherence as a function of the number and types of comorbid conditions in individual people. Guidelines should prompt clinicians to consider comorbidities in addition to the index condition. Discussion of comorbidities should be integrated into guidelines rather than addressed in supplemental sections. In addition to addressing what is known about relevant comorbidities, condition-specific guidelines should concisely summarise what key information is unknown. Guidelines should call attention to and integrate, preventative measures across certain index conditions which may have implications for other conditions and modifiable risk factors. Guidelines should address care co-ordination across providers and settings. Guidelines should be patient-centred rather than focused solely on the management of specific conditions. Because of the complexity of management plans for persons with multiple chronic conditions, the application of guidelines should take into account the need for and importance of shared decision making.

6.2.2.2.7 Need for research

One guideline published recommendations on areas in which future research can improve care for people with cardiovascular disease and co-morbidities.

Table 23: Need for research

Study design and sample		Descriptors of themes
No. of guidelines	Guideline(s)	
1	AHA/ACC/HHS strategies to enhance application of clinical practice guidelines in people with cardiovascular disease and comorbid conditions	Subtheme synopsis There is a general need for more evidence in multimorbidity; this will be aided by greater inclusion of those with multimorbidity in clinical trials and collection of longitudinal data from clinical registries. Recommendations from included guidelines There is a need for external validation of clinical and drug approval trials to ensure that people with multiple comorbid conditions are not excluded unnecessarily. The use of electronic health records and clinical registries can allow for longitudinal evaluation of the management strategies and clinical outcomes of people with multimorbidity. Comorbidity data for selected CPG conditions outlining the most common associations should be developed to inform further CPG research.

6.2.2.3 Quality of the evidence

The quality of the guidelines was assessed using the AGREE II criteria [37]The GDG agreed that scores in categories “scope and purpose”, “stakeholder involvement”, “rigour of development” and “editorial independence” were particularly relevant to assessing the quality of published guidelines in this review, and placed greater emphasis on these criteria when using the evidence to inform their recommendations. Full AGREE II scores are displayed in **Table 24**.

Table 24: Quality ratings: AGREE II criteria

Guideline	Scope and purpose (%)	Stakeholder involvement (%)	Rigour of development (%)	Clarity of presentation (%)	Applicability (%)	Editorial independence (%)	Overall quality (1-7; higher is better)
AHA/ACC/HHS cardiovascular disease and comorbid conditions	56	33	10	42	8	50	2
The Ariadne principles: how to handle multimorbidity in primary care consultations	67	56	67	75	38	67	5
Depression in adults with a chronic physical health problem	94	100	90	78	75	42	5
Guiding principles for the care of older adults with MM	50	72	42	92	63	58	4

Guideline	Scope and purpose (%)	Stakeholder involvement (%)	Rigour of development (%)	Clarity of presentation (%)	Applicability (%)	Editorial independence (%)	Overall quality (1-7; higher is better)
IOM and DHHS meeting on making clinical practice guidelines appropriate for people with multiple conditions	78	56	10	75	13	17	3
Medicines adherence	89	94	92	92	79	67	6
Medicines optimisation	78	100	96	75	63	83	6
Patient experience	100	100	94	83	67	83	6
Polypharmacy guidance (NHS Scotland)	61	72	23	50	67	17	3

6.2.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

6.2.4 Recommendations and link to evidence

Recommendations	**6. When offering an approach to care that takes account of multimorbidity, focus on:** • **how the person's health conditions and their treatments interact and how this affects quality of life** • **the person's individual needs, preferences for treatments, health priorities, lifestyle and goals** • **the benefits and risks of following recommendations from guidance on single health conditions** • **improving quality of life by reducing treatment burden, adverse events, and unplanned care** • **improving coordination of care across services.**
Key principles for assessing, prioritising and managing care for people with multimorbidity	The review identified a number of overlapping principles relevant to the care of people with multimorbidity. Some guidelines targeted specific aspects of care (such as medicines management) while others were directed more generally at patient care. The GDG considered that the Ariadne principles, which focussed on the management of multimorbidity in primary care, were particularly relevant to this review. The GDG also agreed that the recommendations could be generalised beyond primary care settings. The findings from the review indicated the need to go beyond single disease treatments and assessments as may be found in single condition guidelines. Single condition guideline specific care may not be appropriate for people with multimorbidity, due to the potential interactions between diseases and drugs as well as the total treatment burden incurred. The evidence indicated the importance of remembering that people with multimorbidity are often excluded from clinical trials, and therefore the evidence available to guide decisions about treatments is often lacking. Individual care plans are necessary and these need to take account of people's priorities when dealing with risks and benefits. Communication between healthcare professionals is important as people with multimorbidity are likely to be accessing may different services. Some of the evidence included was more specific to medicines management and included recommendations on principles of supporting medicines adherence and medicines optimisation.
Economic considerations	No identified economic evidence was found. This area was deemed to have no major economic implications as no additional costs are expected to be associated with healthcare professionals considering the recommended key principles when assessing, prioritising and managing care for people with multimorbidity. Currently patients have their medications reviewed every year or more often; this recommendation aims at changing the content of these discussion rather than changing the quantity or intensity of the reviews. The GDG agreed that the majority of these conversations would take place within usual consultation time with no associated increased in cost of GP's time. In addition, an individualised management

	plan would lead to a more efficient care and decision making further down the line, and it aims at reducing treatment burden, adverse events, and unplanned or uncoordinated care, which would create some cost savings to the NHS.
Quality of evidence	The quality of the evidence ranged from 2-6 in terms of overall quality as assessed by the AGREE II tool (range 1-7; 7 is highest quality). The GDG discussed the utility of the lower quality guidelines and noted that for the purposes of this review, the principle limitation on the utility of a guideline was its applicability rather than the guideline methodology. However, the quality of the guidelines were considered when assessing the sources of each identified principle. The GDG discussed whether the documents and guidelines were relevant to people with multimorbidity in England. The majority of the guidelines were developed in the UK or in Europe or Australia, which have similar healthcare systems to the UK. Some guidelines were developed in the USA which does have a substantially different healthcare system to the UK; this was taken into account when assessing which guideline principles might inform recommendations. The majority of the guidelines were focused on a population of people with multimorbidity, although some guidelines were focused on a general UK population but on topics that the GDG judged to have particular significance for people with multimorbidity (for example, medicines adherence and polypharmacy). The GDG noted that the principles identified in this review were not a comprehensive list of all principles that are relevant to assessing, prioritising and managing care for people with multimorbidity; however, they reflected a number of important topics.
Other considerations	The GDG used this review of existing principles and also information from the review of barriers to optimising care, and their experience to develop guiding principles for the care of people with multimorbidity. The GDG considered that the aims to improve quality of life and reduce burden of treatment, were important principles for providing an approach to care that takes into account multimorbidity. This has to include attention to an individual's needs, priorities and preferences which should inform consideration of the person's conditions and their treatments and how these may interact and affect their quality of life. The GDG wished to make a clear reference to the need to consider whether the recommendations in single disease guidelines are appropriate for people with multimorbidity. The risk when people are accessing multiple services is that care becomes fragmented and the GDG considered that an important principle is to improve co-ordination of care. Lack of co-ordination was also a significant finding in the review on barriers to optimising care. The GDG chose not to include specific reference to areas already covered in existing NICE guidance but to refer to those guidelines, for example the Patient Experience guideline and the Medicines Adherence guideline. They also chose not to make any recommendations about guideline development or research into multimorbidity as these topics were not the primary aim of the review. These are important and interesting and will be of interest to guideline development organisations but are beyond the remit of this guideline. They do however emphasise the limitations of single conditions guidelines which the GDG did include in this recommendations. The GDG noted that some recommendations referred to the use of decision aids to support decision making. However the GDG were unaware of the existence of any decision aids specifically for people with multimorbidity. The GDG noted that collaborative care was mentioned in some of the guidelines, however, no clinical evidence for this model was found in the models of care review for its clinical or cost effectiveness in multimorbidity.

6.3 Barriers to optimising care

6.3.1 Review question: What are barriers to healthcare professionals optimising care for people with multimorbidity?

For full details see review protocol in Appendix C.

Table 25: Characteristics of review question

Objective	To identify what patients, carers and healthcare professionals believe are the barriers to optimising care for patients with multimorbidity.
Population and setting	Adults with multimorbidity, their family/carers, and healthcare professionals who treat patients with **multimorbidity.**
Review strategy	Qualitative studies will be included, surveys will be included if no qualitative studies are retrieved. Studies will be added until saturation is reached. Studies will be analysed using thematic analysis. Results to be presented as a narrative, and diagrammatically where appropriate. Study quality will be assessed using CERQUAL and GRADE.

6.3.2 Evidence

6.3.2.1 Methods

Two systematic reviews and 11 qualitative studies were included in the review[6,13,52,55,84,90,118,126,212,217,229,230,244] which are summarised in Table 156 below. The 2 systematic reviews, which comprised 22 original papers, were extracted primarily. Subsequent papers investigating barriers to optimising care for people with multimorbidity that were identified in the search were added only where these identified themes that had not already been identified and had reached saturation in the analysis; that is, where these contributed to the further development of existing themes or led to the development of new themes.

Themes and subthemes are presented in Table 27. Themes identified in the review highlighted a number of important barriers to care. This included structural barriers to care at the patient-level (for example lack of time, limited financial resources) and service-level (for example the length of consultations, and difficulties getting appointments with the same GP). Themes identified also included resource barriers, such as the ability of the patient to engage in care (for example due to self-efficacy and emotional distress), and resource on behalf of the healthcare professional (for example, professionals' knowledge of multimorbidity). Communication between patients and their healthcare professionals, and between healthcare professionals, was also identified as important to optimising care for people with multimorbidity. Although the focus of the evidence search and review was intended to be on barriers to optimisation of care, both barriers and facilitators emerged from the analysis.

Evidence from the studies is summarised in the clinical summary tables 28 – 33. For further discussion of the quality assessment of the evidence summarised in the tables below, see the clinical evidence tables in Appendix H. See also the study selection flow chart in Appendix E and excluded studies list in Appendix L.

6.3.2.2 Summary of included studies

Table 26: Summary of studies included in the review

Study	Design	Population	Research aim
Allen 2015[6]	Interviews and focus	n= 17 (n=6 patients; n=11	To better understand how

Study	Design	Population	Research aim
	groups with thematic qualitative analysis	healthcare professionals) Patients: adults with multimorbidity (ESRD and comorbid condition) Healthcare professionals: medical specialists, nurses, social worker (n=1), dietician (n=1) Canada	people with multimorbidity who receive care in institutions designed for treatment of acute illness experience and engage in health-related decisions
Bardach 2012[13]	Semi-structured interviews with constant comparison analysis	n=12 (healthcare professionals) Healthcare professionals: family practice physicians n=6, internal medicine n=5, specialist in OB/GYN n=1 USA	To explore primary care physicians perspectives on prevention counselling among people with multimorbidity
Coventry 2014[52]	Semi-structured with thematic content analysis	n=40 (n=20 patients, n=20 healthcare professionals) Patients: Adults with multimorbidity (with 2 or more of the following conditions: coronary heart disease, diabetes, osteoarthritis, chronic obstructive pulmonary disease). Healthcare professionals: n=16 GPs, n=4 practice nurses) England	To evaluate patient and practitioner views about barriers to self-management in people with multimorbidity
Cowie 2009[55]	Semi-structured interviews with thematic qualitative analysis	n=33 (patients) Adults (median age 67; range 42-83), 90.9% with multimorbidity Male/female ratio: 17:16 England	To examine patients' experiences of continuity of care in the context of different long-term conditions and models of care, and to explore implications for the future organization care of long-term conditions
Fried 2008[84]	Focus groups with constant comparison analysis	n=66 (patients) Older adults (aged 65 years or older) with multimorbidity (median 5 chronic conditions; range 3-8), living in the community	To examine the ways in which older persons with multiple conditions think about potentially competing outcomes, in order to gain insight into how processes to elicit values regarding these outcomes can be

Study	Design	Population	Research aim
		Male/female ratio: 33:67 USA	grounded in the patient's perspective
Gill 2014[90]	Semi-structured interviews with constant comparison analysis	n=27 (patients, informal caregivers, physicians) Patients: older adults (aged 65 years or older; average 82.3±7.7 years) with multimorbidity (median number of conditions 5±2.43) Male/female ratio: 56: 44 Canada	To explore the care challenges experienced by older people with multimorbidity, their informal caregivers and family physicians
Jowsey 2009[118]	Semi-structured interviews (with patients and carers) and focus groups (with health professionals), with qualitative content analysis	n=129 (52 patients; 12 carers, 63 health care professionals) Patients: adults (aged 45-85) with); 86.5% with multimorbidity Male/female ratio: 54:46 Australia	To identify the common challenges co-morbidity poses to patients and carers in their experiences of self-management; to detail the views and perceptions of health professionals about these challenges; and to discuss policy options to improve health care for people with co-morbid chronic illness
Koch 2015[126]	Systematic review N = 12 studies	n=426 (patients) Adults (aged 18 years or older) with multimorbidity England, Scotland, USA	To conduct a systematic review of the literature on patient's perceptions of barriers and facilitators to managing multiple chronic conditions
Schoenberg 2011[212]	Semi-structured interviews with thematic content analysis	n=20 (patients) Adults (aged 41 years or older; mean age 55) with multimorbidity (average number of conditions 4) USA	To improve understanding of how vulnerable rural residents experience and manage several simultaneously occurring chronic health conditions
Sinnott 2013[217]	Systematic review (n=10 studies)	n= 275 (GPs) Belgium, England, Germany, Ireland, Scotland, The Netherlands, USA	To synthesise the existing published literature on the perceptions of general practitioners (GPs) or their equivalent on the clinical management of multimorbidity
Townsend 2003[229] Townsend	Semi-structured interviews with constant	n=23 (patients) Adults (aged 50 years or over)	To examine attitudes towards drug use among middle aged respondents

Study	Design	Population	Research aim
2008[230]	comparative analysis	with multimorbidity (4 or more chronic conditions) Male/female ratio: 10:13 Scotland	with high levels of chronic morbidity.
Williams 2004A[244]	Semi-structured interviews with constant comparative analysis	n=12 (patients) Adults (aged 18 years or over; range 34-77 years) with multimorbidity (average of 5.75 conditions), discharged from acute care Male/female ratio: 1:1 Australia	To investigate perceptions of quality of care by patients experiencing comorbidities who required an acute hospital stay

6.3.2.3 Qualitative evidence synthesis

Table 27: Review findings

Main theme	Sub-themes	Statement of finding
Nature of multimorbidity	Complexity of multimorbidity	Multimorbidity is complex due to the interactions which can occur between conditions and treatments.
Knowledge of multimorbidity	Patient knowledge of multimorbidity	People were reported to have a poor understanding of the complex interactions between their conditions and their treatments.
	Healthcare professionals' knowledge of multimorbidity	Healthcare professionals were reported to have a poor knowledge of how to tailor guidelines to people with multimorbidity.
Services	Patient-level access to services	People with multimorbidity can be prevented from accessing care and engaging in lifestyle changes due to patient-level structural barriers (for example lack of time, limited financial resources).
	Format and coordination of services	Patients can be prevented from accessing care and engaging in lifestyle changes due to service-level structural barriers (for example length of consultation, difficulties getting appointments with the same GP).
	Communication between healthcare professionals	Communication between healthcare professionals was reported to be poor (for example incomplete/delayed feedback, incomplete patient records).
Emotional and psychological factors	Patient emotion	People with multimorbidity can experience negative emotions when discussing options for their care and making decisions, which can impact on shared decision making.
	Motivation and control	People with multimorbidity can lack motivation to engage in the management of their conditions, this may be due to patient's perceived lack of control.
	Mental health	Depression and anxiety can be a barrier to patients being able to manage their own care.

Main theme	Sub-themes	Statement of finding
	Cognitive impairment	Cognitive impairment can be a barrier to effectively communicating with healthcare providers and adhering to treatment regimes.
Relationship between patients and healthcare professionals	Viewing the patient individualistically and holistically	People with multimorbidity wanted healthcare information and education tailored to them as individuals and for healthcare professionals to address them as a whole person, rather than a single condition.
	Communication between patients and health care professionals	Poor communication between patients and healthcare professionals is a barrier to decision making.
	Patient and healthcare professional relationship continuity	Healthcare professional relationship continuity was viewed as a facilitator to communication with and knowledge of the patient.
Additional support	Support from family and friends	Social support can facilitate both accessing and managing care

6.3.2.3.1 Narrative summary of review findings

Theme 1: Nature of multimorbidity

Sub-theme 1.1: Complexity of multimorbidity

People with multimorbidity have a number of chronic conditions and may be in receipt of multiple treatments for their health conditions. People with multimorbidity and the healthcare professionals who care for them reported that this complexity itself was a barrier to optimal care. This is because people with multimorbidity may experience interactions between their conditions and treatments. Interactions between conditions may lead to the symptoms of a condition affecting symptoms of another (for example breathing difficulties can affect mobility and so may worsen arthritis) (n=1; England/Scotland/USA). Interactions between a person's conditions and their treatments may include a condition affecting the efficacy of treatment for another condition. Conditions may also limit the person from engaging in preventative lifestyle changes (for example difficulty breathing is a barrier to increasing exercise) (n=3; England, England/Scotland/USA, Australia). It may also include treatments for a condition affecting symptoms of another condition (for example medication or changes in diet making other conditions problematic) (n=1; Australia). Interactions between treatments may also limit people's access to optimal care (for example, surgery for 1 condition may prevent exercise which would help another condition) (n=1; Canada). The complexity of multimorbidity and treatments was reported to be a barrier to complying with treatment recommendations (n=2; England/Scotland/USA, Canada). However, healthcare professionals suggested that the complexity of multimorbidity may also encourage people with multimorbidity to engage in preventative health behaviour, so as to prevent the onset of new conditions and thus greater complexity (n=1; USA).

The complexity of multimorbidity may also be a barrier to people with multimorbidity and healthcare professionals monitoring current conditions, because it is difficult to distinguish between symptoms of conditions and the effects of treatment (including side effects). Furthermore it was seen as a barrier to identifying the onset of new conditions, as people found it hard to identify the cause of new symptoms (n=1; Canada). The complexity of multimorbidity may affect people with multimorbidity and their healthcare professionals' knowledge and understanding of conditions and treatment (see Theme 2: Knowledge of multimorbidity).

Theme 2: Knowledge of multimorbidity

Sub-theme 2.1: Patient knowledge of multimorbidity

The knowledge of people with multimorbidity about their health conditions, symptoms, treatments, and treatment effects influences their access to optimal care. Healthcare professionals and people with multimorbidity suggested that people with multimorbidity may have a poor understanding of the complex interactions between their conditions (n=3; England/Scotland/USA, Canada) and their treatments (n=1; England/Scotland/USA).

Poor knowledge can be a barrier to the diagnosis of new conditions and to identifying the effects of treatment. This is because people may find it difficult to recognise the signs and symptoms of their conditions, and to differentiate them from drug interactions and side-effects (n=1; Australia). A person's knowledge of multimorbidity may affect their ability to communicate with healthcare professionals (see subtheme 5.2), for example by limiting their ability to report new symptoms. People with multimorbidity reported that healthcare professionals often relied on them to be knowledgeable about their own conditions and treatments to inform decisions about management. This may lead to higher treatment burden for people, and may be problematic when people are less knowledgeable. It may also be problematic in cases where people may not be able to provide reliable information (for example people with dementia) (see subtheme 4.4).

Poor patient knowledge can also be a barrier to effective treatment, as it may affect a person's ability to use treatment as indicated. For example, patients may take suboptimal doses of medication. Also, poor knowledge of brand names of medications could lead to people taking higher doses of medication than prescribed if people unknowingly take 2 doses of the same medication with different brand names (n=2; England/Scotland/USA , Australia)). Furthermore people with multimorbidity reported that they were unable to find information on the interactions between their conditions and treatments across their conditions (n=1; England/Scotland/USA). People reported that they would like more information on their conditions to inform the management of their conditions (n=1; Australia). Poor patient knowledge of the healthcare system can also be a barrier to seeking care when people do not know who to contact after identifying new symptoms (n=1; Australia).

However, good knowledge of conditions and treatments can facilitate care. Good patient knowledge of their own conditions, current treatments and treatment effects can facilitate communication with healthcare professionals, as patients can communicate changes in their circumstances and preferences more effectively. This in turn can enable the choice of a suitable treatment for the person (n=1; Australia). One study also indicated that people who had greater knowledge of their conditions were more likely to engage in the management of their conditions and undertake health improving behaviour (n=1; England).

People's knowledge of their conditions and treatments may be influenced by the complexity of their multimorbidity (see Theme 1: complexity of multimorbidity).

Sub-theme 2.2: Healthcare professionals' knowledge of multimorbidity

Healthcare professionals' who lack knowledge about treating people with multimorbidity and how to tailor single-condition clinical guidelines to them may be a barrier to care for people with multimorbidity. Healthcare professionals reported concerns that clinical guidelines are generally focused on the care of a single condition and do not consider the specific circumstances of an individual person. This can make it difficult for them to apply recommendations to the care of people with multimorbidity, which requires the consideration of a number of different conditions and treatments (n=2; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA, USA). Healthcare professionals reported that they were not confident in tailoring clinical guidelines to a person's existing conditions and wider social circumstances as there is little guidance on what constitutes good practice in this area (n=1, England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA).

It was also reported that poor healthcare professional knowledge of the brand names of medications could lead to a healthcare professionals prescribing the same medication twice under different brand names, which may lead to the person taking higher doses of their medication than prescribed (n=1, Australia).

Healthcare professionals' knowledge of multimorbidity may affect their ability to communicate with patients (see subtheme 5.2) and with other healthcare professionals (see subtheme 3.3).

Theme 3: Services

Limited access to care, such as healthcare appointments, can be a barrier to care for people with multimorbidity. This could be through difficulties at both the patient- and service-level. Difficulties at the patient-level are driven by difficulties accessing care due to individual characteristics of the person with multimorbidity, whereas those at the service-level are driven by the structure of the healthcare system in itself.

Sub-theme 3.1: Patient-level access to services

People with multimorbidity may be required to attend multiple appointments and be in receipt of multiple treatments for their health conditions. People with multimorbidity reported that they did not always have the time to access care and to engage in lifestyle changes, such as exercise (n=2; England, England/Scotland/USA). People with multimorbidity and healthcare professionals also reported that financial resources may limit access to medication for people with multimorbidity, and may also limit their ability to engage in preventative strategies, for example exercise (n=5; England, England/Scotland/USA, Australia, USA). People with multimorbidity and healthcare professionals also reported that some people may have difficulties in accessing services due not having access to a telephone or to the internet (n=1; England), and through having limited access to transportation, which may affect attendance at appointments (n=5; England, England/Scotland/USA, USA, Canada). Healthcare professionals reported that people who live in economically deprived areas are more likely to have problems accessing care due to financial constraints and limited access to transportation (n=1; England).

Patient-level barriers to care may be exacerbated or may be reduced by having support from family and friends (see Theme 6: Additional support).

Sub-theme 3.2: Format and coordination of services

People with multimorbidity and healthcare professionals reported that the length of normal primary healthcare consultations was too short, as there was not enough time to discuss more than one issue or to discuss issues in the wider context of the person's conditions and treatments (n=4; England, England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA, USA). People with multimorbidity reported incidents of poor coordination of care across healthcare settings and providers. People thought that the coordination of care across organisational boundaries was poor (n=3; England, England/Scotland/USA, England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA, Canada). For example poor coordination of the results of diagnostic tests may result in delayed reporting of the test results back to people, thereby putting people under unnecessary stress (n=1; Canada). Furthermore, poor coordination of appointments across different departments/providers can lead to people having to attend multiple appointments on different days, putting unnecessary burden on them (n=2; England/Scotland/USA, England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA). People reported long waits for appointments, diagnostic tests and results (n=1; Canada) and that it was difficult to get urgent appointments (n=2, England), especially with specialists (n=2; Australia, USA). People also reported that it was difficult to get appointments with the same GP (n=1; England). People expressed desire for a healthcare professional who could coordinate their care (n=1; USA). Carers also discussed the

benefits of having a single person who could manage communication between and coordinate across different specialities for the patient, as a facilitator to care (n=1, Canada). Poor coordination of care could act as a barrier to communication, for example, people who are unsure who is in control of their care and so do not know who to contact (n=2; England, USA).

The format of consultations may affect the relationship between patients and healthcare professionals (see Theme 5: Relationship between patients and healthcare professionals). The coordination of services may also affect the communication between healthcare professionals (subtheme 3.3).

Sub-theme 3.3: Communication between healthcare professionals

People with multimorbidity, their carers and healthcare professionals reported that there was poor communication between healthcare professionals who are involved in a person's care (n=6; England, England/Scotland/USA, Australia, Canada, USA). For example, incomplete patient records and incomplete or delayed feedback from clinicians, particularly specialists (n=4; England, Canada, Australia, USA). People with multimorbidity reported that poor communication between healthcare professionals was reported to be a barrier to effective treatment. For example, people could receive conflicting information from different clinicians (n=2; England/Scotland/USA, Australia), and poor communication could lead to delays in receiving treatment (n=1, England).

Often, poor communication between healthcare professionals was associated with increased treatment burden for people with multimorbidity (see sub-theme 4.1). Poor communication between healthcare professionals was viewed as a barrier to coordination as difficulties sharing information between healthcare professionals meant that people often need to coordinate their own care, for example through knowing information about their own conditions, treatments and healthcare appointments. Carers reported that GPs were often not aware of the person's medical history (n=1; Canada). Healthcare professionals recognised that they often relied on people to communicate feedback from other clinicians (n=2; Canada). People felt that the length and complexity of their patient records make it difficult and time consuming for healthcare professionals to navigate (n=1; Australia) and that they often were responsible for communicating symptom and functional status changes to different providers (n=1; England/Scotland/USA).

Theme 4: Emotional and psychological factors

Sub-theme 4.1: Patient emotion

People with multimorbidity can find making decisions about their care to be a source of emotional distress (n=1; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA). Healthcare professionals reported reluctance to raise some topics, as they were concerned about causing people distress. For example, discussions about prioritising treatment to address a balance between life expectancy and quality of life, and talking about dietary/exercise interventions with people with multimorbidity who are overweight (n=2, England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA, USA).

Patient emotion may affect a person's motivation and control (see subtheme 4.2) and their mental health (see subtheme 4.3).

Sub-theme 4.2: Motivation and control

Healthcare professionals highlighted that some people may lack motivation to engage in the management of their conditions, for example adhering to medication. This may because some people lack self-efficacy to manage their conditions, their treatments, or to change their wider social circumstances. Consistent with this, people described the perceived lack of control over their

conditions and treatments as a barrier to managing their own care (n=1; England/Scotland/USA). However, people reported that the ability to participate in healthcare decisions and to manage their own care can feel empowering (n=1; England/Scotland/USA). Healthcare professionals reported that socioeconomic deprivation may negatively impact people's motivation to engage in their care (n=1; England). Health professionals considered that this could be due to having a perceived lack of control over their lives, which could reduce self-efficacy to participate in treatment and health behaviours (n=1; England). Healthcare professionals also noted that people in deprived areas may prioritise other concerns (for example, paying bills) over engaging in health-related behaviours.

Sub-theme 4.3: Mental health

People with multimorbidity and healthcare professionals reported that depression and anxiety can be a barrier to people with multimorbidity managing their own care, including medication compliance and preventative action, for example diet and exercise regimes (n=4; England, England/Scotland/USA, Australia, Canada).

Sub-theme 4.4: Cognitive impairment

Cognitive impairment was reported as a barrier to effectively communicating with healthcare providers (n=1, England/Scotland/USA) and to adherence to treatment regimens (n=1; Australia). For example people with dementia may not be able to accurately report changes in symptoms or treatment effects to healthcare professionals, and may forget whether they have taken their medication.

Theme 5: Relationship between patients and healthcare professionals

Sub-theme 5.1: Viewing the patient individualistically and holistically

People with multimorbidity reported that they wanted healthcare professionals to consider them as a whole person when making decisions about care, rather than focus on a single condition only (n=1; England). People with multimorbidity and carers reported that healthcare professionals focused on a single condition only and did not consider the full complexity of their multimorbidity. This was reported to be particularly the case amongst healthcare professionals who are specialists in a particular condition (n=3; Australia; Canada). People with multimorbidity reported that they wanted healthcare professionals care that was tailored to their individual needs, taking into consideration their personal preferences, all of their conditions, treatments and wider social circumstances (n=1; England/Scotland/USA).

Sub-theme 5.2: Communication between patients and health care professionals

Poor communication between people with multimorbidity and healthcare professionals may be a barrier to making decisions about care. People with multimorbidity reported that they felt that they had little support from healthcare professionals in making decisions about their care (n=3; Scotland, Canada) and that they wanted healthcare professionals to listen, be sympathetic and take time to explain things to them (n=1; England).

Healthcare professionals also discussed how they may find it difficult to discuss the outcome and of treatment with people, and to involve people in making decisions about treatment, because of concerns about causing them distress. This may be particularly the case when discussing the balance

between life expectancy and quality of life (n=1; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA). Healthcare professionals feel that they need to develop enhanced communication skills to discuss with people the interactions between conditions and to discuss treatment options with them, including stopping medications (n=1; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA).

Patients also expressed experiencing distress in instances where they felt that healthcare professionals did not communicate the results of tests in a timely manner (n=1; Canada) or communicate clearly the reason for appointments (n=1; England).

Communication between patients and healthcare professionals can be affected by both the person and healthcare professionals' knowledge of multimorbidity (see Theme 2: knowledge of multimorbidity), the person's emotions (see subtheme 4.1) and cognitive impairment (see subtheme 4.4).

Sub-theme 5.3: Patient and healthcare professional relationship continuity

Relationship continuity between people with multimorbidity and healthcare professionals can facilitate healthcare professionals' knowledge of the person. Continuity was also viewed as a facilitator to communication between people and healthcare professionals. For example, it may facilitate healthcare professionals' knowledge of the person's personal and clinical history (n=1; England). This may mean that healthcare professionals can offer more suitable treatments, and may reduce the amount of time required to discuss treatment options. Healthcare provider continuity was viewed as important to building a trusting relationship between people and healthcare professionals (n=1; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA). People discussed how greater trust in their clinician may encourage them to agree to changes in their care and may also increase their adherence to treatment (n=1; USA). Continuity may therefore encourage people to disclose new symptoms or concerns about treatment, because the person trusts or feels more comfortable disclosing information to the healthcare professional (n=1; England/Scotland/Ireland/Belgium/Germany/The Netherlands/USA).

Continuity in the relationship between people with multimorbidity and healthcare professionals may affect the ability of healthcare professionals to see the person as a whole individual (see subtheme 5.1) and their ability to communicate with each other (5.2).

Theme 6: Additional support

Sub-theme 6.1: Support from family and friends

Healthcare professionals reported that social isolation or a lack of a network of social support may be a barrier to people with multimorbidity managing their health conditions (n=1; England). People with multimorbidity reported that family and friends provided them with support in managing their conditions (n=2; England, England/Scotland/USA). This included financial support (for example helping them pay for transport to healthcare appointments), emotional support (for example motivating them to exercise), informational support (for example helping them to find and interpret relevant information), and behavioural support (for example reminding them to take their medication). People with multimorbidity also reported that family and friends helped them to access healthcare appointments through providing transportation. Some people also reported that family and friends also helped to facilitate decision-making, such as which treatment to undertake, (n=1; Canada), others reported that they did not (n=1; England/Scotland/USA.

6.3.2.4 Evidence summary

For further discussion of the quality assessment of the evidence summarised in the tables below, see the clinical evidence tables in appendix H.

Table 28: Summary of evidence: Nature of multimorbidity

Study design and sample		Findings	Quality assessment		
No. of studies	Design		Criteria	Rating	Overall assessment of confidence
Sub-theme 1.1: Complexity of multimorbidity					
8	Systematic review (1), interviews (5), interviews and focus groups (1), and focus groups (1)	Multimorbidity is complex due to the interactions which can occur between conditions and treatments.	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	No concerns about adequacy	

Table 29: Summary of evidence: Knowledge of multimorbidity

Study design and sample		Findings	Quality assessment		
No. of studies	Design		Criteria	Rating	Overall assessment of confidence
Sub-theme 2.1: Patient knowledge of multimorbidity					
5	Systematic review (2), interviews (2), interviews and focus groups (2)	People were reported to have a poor understanding of the complex interactions between their conditions and their treatments.	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	

Study design and sample		Findings	Quality assessment		
No. of studies	Design		Criteria	Rating	Overall assessment of confidence
			Adequacy	Minor concerns about adequacy	
Sub-theme 2.2: Healthcare professionals' knowledge of multimorbidity					
3	Systematic review (1), interviews (1), interviews and focus groups (1)	Healthcare professionals were reported to have a poor knowledge of how to tailor guidelines to people with multimorbidity.	Limitations	Minor limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	Partially relevant	
			Adequacy	No concerns about adequacy	

Table 30: Evidence summary: Services

Study design and sample		Findings	Quality assessment		
No. of studies	Design Sample		Criteria	Rating	Overall assessment of confidence
Sub-theme 3.1: Patient-level access to services					
5	Systematic review (1), interviews (3), interviews and focus groups (1)	People with multimorbidity can be prevented from accessing care and engaging in lifestyle changes due to patient-level structural barriers (for example lack of time, limited financial resources).	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	No concerns about adequacy	
Sub-theme 3.2: Format and coordination of services					
6	Systematic review (2), interviews (4)	People with multimorbidity can be prevented from accessing care and engaging in lifestyle changes due to service-level structural barriers	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns	

Study design and sample		Findings	Quality assessment		
No. of studies	**Design Sample**		**Criteria**	**Rating**	**Overall assessment of confidence**
		(for example length of consultation, difficulties getting appointments with the same GP).		about coherence	
			Relevance	No concerns about relevance	
			Adequacy	No concerns about adequacy	
Sub-theme 3.3: Communication between healthcare professionals					
6	Systematic review (1), interviews (4), interviews and focus groups (1)	Communication between healthcare professionals was reported to be poor (for example incomplete/delayed feedback, incomplete patient records).	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	No concerns about adequacy	

Table 31: Evidence summary: Emotional and psychological factors

Study design and sample		Findings	Quality assessment		
No. of studies	**Design Sample**		**Criteria**	**Rating**	**Overall assessment of confidence**
Sub-theme 4.1: Patient emotion					
2	Systematic review (1), interviews (1)	People with multimorbidity can experience negative emotions when discussing options for their care and making decisions, which can impact on shared decision making.	Limitations	Minor limitations	LOW
			Coherence	Minor concerns about coherence	
			Relevance	Partially relevant	
			Adequacy	Substantial concerns about adequacy	

Study design and sample		Findings	Quality assessment		
No. of studies	**Design Sample**		**Criteria**	**Rating**	**Overall assessment of confidence**
Sub-theme 4.2: Motivation and control					
2	Systematic review (1), interviews (1)	People with multimorbidity can lack motivation to engage in the management of their conditions, this low self-efficacy may be due to patient's perceived lack of control.	Limitations	Moderate limitations	LOW
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	Substantial concerns about adequacy	
Sub-theme 4.3: Mental health					
3	Systematic review (1), interviews (1), interviews and focus groups (1)	Depression and anxiety can be a barrier to patients being able to manage their own care.	Limitations	Moderate limitations	LOW
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	Minor concerns about adequacy	
Sub-theme 4.4: Cognitive impairment					
2	Systematic review (1), interviews and focus groups (1)	Cognitive impairment can be a barrier to effectively communicating with healthcare providers and adhering to treatment regimes.	Limitations	Moderate limitations	LOW
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	Substantial concerns about adequacy	

Table 32: Summary of evidence: Relationship between patients and healthcare professionals

Study design and sample			Quality assessment		
No. of studies	**Design Sample**	**Findings**	**Criteria**	**Rating**	**Overall assessment of confidence**
Sub-theme 5.1: Viewing the patient individualistically and holistically					
5	Systematic review (1), interviews (3), interviews and focus groups (1)	People with multimorbidity wanted healthcare information and education tailored to them as individuals and for healthcare professionals to address them as a whole person, rather than a single condition.	Limitations	Moderate limitations	LOW
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	Substantial concerns about adequacy	
Sub-theme 5.2: Communication between patients and health care professionals					
5	Systematic review (1), interviews (3), interviews and focus groups (1)	Poor communication between patients and healthcare professionals is a barrier to decision making.	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	Minor concerns about adequacy	
Sub-theme 5.3: Patient and healthcare professional relationship continuity					
3	Systematic review (1), interviews (2)	Healthcare professional relationship continuity was viewed as a facilitator to communication with and knowledge of the patient.	Limitations	Moderate limitations	MODERATE
			Coherence	Minor concerns about coherence	
			Relevance	No concerns about relevance	
			Adequacy	No concerns about adequacy	

Table 33: Clinical summary: Additional support

Study design and sample		Findings	Quality assessment		
No. of studies	Design Sample		Criteria	Rating	Overall assessment of confidence
Sub-theme 6.1: Support from family and friends					
3	Systematic review (1), interviews (1), interviews and focus groups (1)	Social support can facilitate both accessing and managing care.	Limitations	Severe limitations	LOW

6.3.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

Economic

- No relevant economic evaluations were identified.

6.3.4 Recommendations and link to evidence

Recommendations	**7. Follow these steps when delivering an approach to care that takes account of multimorbidity:** • **Discuss the purpose of an approach to care that takes account of multimorbidity (see recommendation 19).** • **Establish disease and treatment burden (see recommendations 20 to 22).** • **Establish patient goals, values and priorities (see recommendations 23 to 25).** • **Review medicines and other treatments taking into account evidence of likely benefits and harms for the individual patient and outcomes important to the person (see recommendations26 to 33).** • **Agree an individualised management plan with the person (see recommendation 34), including:** **- goals and plans for future care (including advance care planning)** **- who is responsible for coordination of care** **- how the individualised management plan and the responsibility for coordination of care is communicated to all professionals and services involved** **- timing of follow-up and how to access urgent care.**
Barriers and facilitators	The evidence suggests that the complexity of multimorbidity itself is an important barrier to optimising the care of people with multimorbidity. Complexity includes the interactions between conditions, between treatments, and between conditions and treatments. The complexity of multimorbidity is also a barrier to the amount of knowledge required of both patients and healthcare professionals. In addition to the knowledge of their conditions and their treatments, people may also need knowledge about the interactions between conditions and treatments. The GDG noted that the complexity of multimorbidity may place considerable burden on the person, and may make it difficult for healthcare professionals to monitor and treat people with multimorbidity. The GDG noted that healthcare professionals should take into consideration the complexity of multimorbidity when discussing current and new treatments with patients (and their carers). The evidence suggested that mental health issues and cognitive impairments can also be a barrier to people managing their own care and to treatment adherence. For example, people with depression may lack the motivation to adhere with their

	treatments. People with cognitive impairments, such as dementia, may not be able to accurately report changes in symptoms or treatment effects to healthcare professionals, and may forget whether they have taken their medication. The GDG noted that mental health difficulties and cognitive impairment should be taken into consideration when discussing care with the patient (and their carer). The evidence suggested that the length of consultation is a barrier to care as people with multimorbidity and the healthcare professional lack time to discuss more than one issue, or to discuss care in the wider context of the person's conditions and treatments. The GDG discussed the format of consultations, noting that the evidence suggested that patients and healthcare professionals felt that standard consultations did not necessarily provide enough time to discuss important issues. The GDG considered making a recommendation for healthcare professionals to consider routinely using double appointments to consult with people with multimorbidity. The GDG believed that this may optimise communication between healthcare professionals and people with multimorbidity about their conditions and treatments. The GDG suggested that this may not necessarily increase the overall appointment time spent with people with multimorbidity, as otherwise people with multimorbidity may have to make multiple appointments with some repetition. The GDG also considered making a recommendation about the use of alternative formats of consultations, such as telephone support. The GDG agreed that longer consultation times and alternative formats of consultations may be of benefit to people but noted that there was little quantitative evidence in this area (please see section 12). The evidence also suggested that people found it difficult to get urgent appointments and appointments with specialists and with the same GP. The GDG noted that many primary care healthcare appointments are usually -driven by patient need and are episodic, based on the immediate issues a person may have. The GDG discussed the benefits of a regular planned review of people with multimorbidity in order to discuss any issues about their care. Without needing to discuss any immediate health issues, this may allow greater time for discussing the wider management and longer-term care of a person's conditions. The GDG also noted that this may identify 'non-immediate' issues which they may not book an appointment for, but which addressing may lead to benefit for the person. The evidence suggested that healthcare professionals find it difficult to discuss the outcomes and risks of treatment with people, and to involve them in the decision making process particularly when discussing sensitive issues. This is a barrier to optimal care because it may lead healthcare professionals to avoid discussions that may lead to changes in treatment or in stopping certain treatments, which in turn may reduce the person's treatment burden. The GDG noted that it was important that healthcare professionals feel adequately supported and trained to discuss sensitive topics with people with multimorbidity, and to be able to incorporate people's values and preferences in decisions about care. The evidence also suggested that people can find making decisions about their care to be a source of stress and that they often would like support from healthcare professionals, as well as family and friends, when making decisions. The GDG noted that the complexity of multimorbidity can impact on the decision making abilities of both patients and healthcare professionals. The evidence suggested that poor communication between healthcare professionals may be a barrier to the care of people with multimorbidity. The GDG noted that poor communication between healthcare professionals involved in a person's care could increase the burden of treatment on people and lead to distress. The GDG agreed that healthcare professionals in different services should seek to communicate effectively with each other. In particular, the GDG suggested that the healthcare professional who is referring patient person for specialist care should provide clear and precise information on what care is needed and the context of referral, including the person's other conditions and treatments.

Economic considerations	No economic evidence was identified on the barriers and facilitators to optimising care for people with multimorbidity as this was a qualitative review and therefore did not look at the comparative effectiveness or cost effectiveness of interventions. The recommendations made by the GDG are not expected to have any major cost implications as they indicate what elements have to be considered by health care professionals in the discussion of care with the patients. Currently patients have their conditions and medications reviewed every year or more often; this recommendation aims at changing the content of these discussion rather than changing the quantity or intensity of the reviews. The GDG agreed that the majority of these conversations would take place within usual consultation time with no associated increased in cost of GP's time. In addition, an individualised management plan would lead to a more efficient care and decision making further down the line, and it aims at reducing treatment burden, adverse events, and unplanned or uncoordinated care, which would create some cost savings to the NHS.
Quality of evidence	All of the evidence was of low to moderate quality. The majority of the subthemes had moderate limitations. The majority of studies had no or only minor concerns about relevance and adequacy. The studies were all conducted in a population of people with multimorbidity, or in carers or health professionals who worked with people with multimorbidity. The majority of the studies were conducted within the UK, or within Europe, Canada or Australia, which have similar healthcare systems to the UK. A few studies were conducted in the USA which does have a substantially different healthcare system to the UK but this was taken into account when assessing the applicability of the themes around the delivery of services. The GDG were specifically interested in identifying evidence in groups of people at particular risk of multimorbidity (for example low socioeconomic status). However very little information specifically relating to these groups was identified.
Other considerations	The GDG used the review of barriers to optimising care, in combination with the information from the review of principles of care and other reviews in the guideline to develop a plan for how an approach to care that takes account of multimorbidity should be delivered. The GDG considered that the barriers review presented clear evidence of the difficulty felt by people with multimorbidity and also that expressed by their doctors of making sense of their multiple conditions and treatments and how these may interact. The evidence provided clear expression of not only the difficulties of the conditions and their interactions but the added burden of mental health and mood issues, and the difficulties of managing interactions with health services. The GDG noted that there is limited understanding of the interactions that may occur between conditions and treatments in people with multimorbidity. This is partly due to the diversity of this patient group, but also because of the lack of research conducted specifically in people with multimorbidity to explore such interactions. The GDG believed that further research with people with multimorbidity may inform understanding and care. In the interim the GDG suggested that healthcare professionals should take into consideration all of a person's conditions and treatments, and the wider context of their lives, when making decisions about care. The GDG agreed that a healthcare professional needs to be explicit and needs to explain to people the purpose of an approach to care that takes account of multimorbidity. This explanation will have to be individualised to the person with multimorbidity and the GDG considered it important that a multimorbidity approach to care be seen as a positive way of managing a person's care. One of the important stages in delivering this approach is understanding the person's experience of their illness and treatments. This is discussed in more detail in section 9. To be able to offer an individualised approach the healthcare professional needs to understand an individual person's preferences, values and priorities. This is an important concept in the area of shared decision making which is discussed more fully in NICE guideline on patient experience in adult NHS services and here in

	section 6. Once people's preferences and priorities are explored and any burdens of treatment understood, healthcare professional and patient can review medicines and other treatments a person is taking and consider whether they the serve a person's interests. The result of this should be a plan for the person's continuing care which includes the person's goals and plans for future care, The GDG acknowledged that some discussions with people with multimorbidity require skill and sensitivity; for example, discussing the trade-off between quality of life and life expectancy when considering prioritising treatments. An important part of delivering a multimorbidity approach to care is a discussion about how any decisions made will be communicated to other healthcare services and professionals and how future care will be co-ordinated. The person with multimorbidity and healthcare professional should agree appropriate follow up. The GDG noted that use of Summary Care Records (SCR), where available, may be beneficial in improving communication. SCRs should make it easier for the variety of different healthcare professionals involved in the care of people with multimorbidity to document plans for care. The GDG discussed the potential use of enriched Summary Care record but noted that patients need to give explcicit consent for this and that healthcare practitioners have a role in encouraging its use if appropriate. The GDG noted that the majority of evidence came from a population of people with multimorbidity in primary care. They agreed that these individualised approaches to care will be largely carried out in primary care settings but that this approach will also be familiar to specialists in care of the elderly. They considered the principles of the approach and the principles of stopping an approach useful to all healthcare professionals. Additionally, they considered awareness of the concept of treatment burden and the importance of patient priorities also helpful to all healthcare professionals.

7 Identification

Multimorbidity is common and most people with multimorbidity have uncomplicated care. Using systematic methods to find people with problematic multimorbidity would allow better identification of people who might benefit from an individualised approach and be a better use of resources than less systematic methods. Multimorbidity is associated with reduced quality of life, higher mortality, polypharmacy and high treatment burden, and much greater health services use including emergency hospital admissions. This chapter reports on evidence reviews that sought to explore whether these factors could be used to identify those people with multimorbidity that might benefit from tailored care and GDG discussion about their use. Principles of an approach to care that takes account of multimorbidity are discussed in chapter 6 and details of what this approach to care would involve are discussed in detail in chapter 9.

7.1 Approach to identification

Recommendations	**8. Identify adults with multimorbidity who may benefit from an approach to care that takes account of multimorbidity (as outlined in Chapter 6):** • **opportunistically during routine care** • **proactively using electronic health records.** **Use the criteria in recommendation 5 to guide this.**
Relative values of different outcomes	The GDG were interested in ensuring that practitioners were alert to people with multimorbidity and considered that this could best be achieved using multiple methods including clinical judgement when people are seen opportunistically or more proactively taking advantage of electronic health records.
Trade-off between clinical benefits and harms	The GDG considered that the benefits or harms of identifying people depended on the quality of review that they received. Carried out professionally and sensitively they considered that such a review would be of benefit to most people offered it.
Economic considerations	No relevant economic evaluations were identified. The GDG did not think there were any resource costs to identifying people opportunistically as people who might benefit from an approach to care that takes account of multimorbidity. The healthcare professional time associated with using electronic health records is minimal (less than a minute per patient), while there are potential benefits of identifying people who may benefit from this approach to care. These are people who are already in contact with the health care system. The use of electronic health records and resource issues is discussed in sections below where individual tools are reviewed.
Quality of evidence	These recommendations were informed by evidence reviews in this chapter and chapter 8 and quality of evidence contributing to the recommendations is discussed in those sections.
Other considerations	The GDG used the evidence their experience and the information in the reviews of risk tools and of association between polypharmacy and harms to develop these recommendations. They considered that identifying people who may benefit from an approach to care that takes account of multimorbidity needs to be done using clinical judgement and using tools and quantitative measures. The GDG agreed that in many clinical settings people could be identified opportunistically during routine

	care. The criteria the GDG outlined in section 5.3 such as people finding it difficult to manage their treatments or day to day activities, or people who are receiving care from multiple services are likely only to be identified within clinical encounters. The GDG considered it important to remind healthcare professionals that those people who might particularly benefit from a multimorbidity approach to care might not be targeted if electronic tools and purely quantitative ways of identifying people were used. The GDG considered that many healthcare practitioners, particularly those working in primary care could make a qualitative judgement about which of their patients were likely to be appropriate for a multimorbidity approach to care. People with falls or unplanned hospital admissions and people receiving multiple regular medicines may be identified when electronic health records are used for example when prescribing in an outpatient hospital setting. There is also the potential to use electronic health records proactively to identify people and specific ways of doing this are discussed in further sections of this chapter.

7.2 Unplanned hospital admissions

7.2.1 Review question: What risk tool best identifies people with multimorbidity who are at risk of unplanned hospital admission?

For full details see review protocol in Appendix C.

Table 34: PICO characteristics of review question

Question	What risk tool best identifies people with multimorbidity who are at risk of unplanned hospital admission?
Population	Adults (aged >17 years) with multimorbidity
Risk tool	Risk tools identified in the literature validated for predicting unplanned hospital admissions in people with multimorbidity.
Target condition or Reference standard	Unplanned hospital admission (max time point = 3 years)
Outcomes (in terms of predictive test accuracy, calibration)	Area under the curve (c-statistic) Sensitivity, specificity, predictive values Predicted risk versus observed risk (calibration) Other outcomes for example, Somers' D statistic, R^2 statistic and Brier score Reclassification
Study types	Prospective and retrospective cohort studies (external validation, internal validation (split half validation))

The GDG wished to identify studies that evaluated the accuracy of risk tools for identifying individuals with multimorbidity who are at risk of an unplanned hospital admission. The GDG considered that both specificity and sensitivity were important but agreed that they would prioritise specificity data for decision-making (that is, the ability of the tool to correctly identify people who were not at risk of unplanned hospital admission). This is because they wanted to ensure that people who were not at risk of an unplanned hospital admission were not identified as requiring additional assessment and support, which may be associated with significant resource implications. However, the GDG believed that the sensitivity of the tool (that is, the ability of the tool to correctly identify people who are at risk of unplanned hospital admission) was also important, so as to ensure that individuals at a high risk of an unplanned hospital admission are identified and can be considered for a multimorbidity approach to care. As a consequence, the GDG prioritised higher specificity, but expected the tool to have high sensitivity to recommend its use in practice.

Nineteen studies evaluating 37 risk tools were included in the review,[3,25,43,44,47,59,63,65,108,116,144,211,223,239,247,248] these are summarised in Table 35 below. One paper (Wallace 2014[10]) is a systematic review that provided pooled discrimination data from 3 studies, which evaluated the prognostic accuracy of the Probability of Repeated Admission (Pra) tool, at a threshold of 0.5 or greater, for identifying people who will have unplanned hospital admissions. We have retained this pooled data but undertaken independent quality assessment of the 3 studies.

Full details of the tools included in this review are provided in the clinical evidence tables in Appendix H. A broad number of factors were included in these tools, including: demographic variables; disease type; presence of comorbidities; medication use; function; quality of life; and laboratory or clinical tests. There was variation across the included studies in the definition of the outcome. We have included studies that evaluated the prognostic accuracy of tools for first hospitalisation, re-hospitalisation, and admission to emergency department where these were defined in the studies as unplanned. We have also included 2 studies where the outcome was defined as 2 or more admissions to hospital or ED (Coleman 1998[47] Susser 2008[223]), and reported this data separately. The number of unplanned admissions in each study ranged from 7.45% to 76.8%. Follow up of the studies ranged from 30 days to 10 years.

The studies included populations in a variety of settings: living in the community (n=15) and previously hospitalised and discharged (n=4). Only 1 study evaluated the prognostic accuracy of a tool to identify people who will have an unplanned hospital admission amongst a multimorbid population (Zeng 2014[15]). Sixteen studies were included that were conducted with an older adult population, which did not report the number of people with multimorbidity. Following discussion with the GDG, a further 2 studies were included (Hippisly-Cox 2013[108] and Donnan 2008[65]) that were conducted with adults from the general population (aged 18-100 years old and aged over 40 years old, respectively). The GDG agreed to include these studies as they report data on tools currently used in practice in the UK (PEONY in Scotland, and QAdmissions in England). Evidence from these studies is summarised in the clinical evidence summary below. See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, and excluded studies list in Appendix L.

The paper describing validation of the eFI [44] was published after the cut off date for literature review and was highlighted in stakeholder comments at consultation. This was included following GDG discussion because of its particular relevance to the guideline population and that it predicted both admission to care home facility and hospitalisation.

Summary of included studies

Table 35: Summary of included studies in the review

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
Abbatecola 2011[3]	Hospitalised Older Patient Examination (HOPE) Index	n=1510 Older adults (aged 70 or older; mean age 81±6 years), currently hospitalised Multimorbidity: number of people with multimorbidity not reported	Unplanned readmission to an acute geriatric ward (2 years) C-statistic Sensitivity Specificity	76.8%	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
		Italy			
Boeckxstans 2015[25]	Cumulative Illness Rating Scale (CIRS) Unweighted disease count	n=560 Older adults (aged 80-101 years; mean age 84.7±3.7), living in the community Multimorbidity: 37.6% reported 5 or more comorbid diseases (range 1-16 diseases) Belgium	Time to first hospitalisation (3 years) C-statistic Sensitivity Specificity	50.9%	Prospective cohort
Clegg 2016[43,44]	Electronic Frailty Index (eFI) Unweighted deficit count (36 items)	n = 723727 (internal validation cohort = 207720, external validation cohort = 516007) Older adults (aged 65 to 95) registered at relevant GPs Multimorbidity: number of people with multimorbidity not reported UK	Hospitalisation (1 and 3 years) C-statistics Sensitivity Specificity	101998 by 3 years across both cohorts (14%)	Retrospective cohort
Coleman 1998[47]	Probability of Repeated Admission (Pra): calculated using administrative data Probability of Repeated Admission (Pra): calculated using self-report data	n=2174 Older adults (aged 65 years or over), living in the community Multimorbidity: number of people with multimorbidity not reported USA	≥2 admissions (4 years) C-statistic	Not reported	Prospective cohort
Daniels 2012[59]	Groningen Frailty Indicator Dutch Tilburg Frailty Indicator Sherbook Postal	n=430 Older adults (aged 70 or older; mean 77.2±5.5), living in the community Multimorbidity:	Hospital admission (1 year) Sensitivity Specificity C-statistic	n=75 (17%)	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	Questionnaire	number of people with multimorbidity not reported The Netherlands	PPV NPV		
Donate-Martinez 2013[63]	Probability of Repeated Admission (Pra): self-report The Community Assessment Risk Screen (CARS)	n=500 Older adults (aged 65 or over) Multimorbidity: number of people with multimorbidity not reported Spain	Hospital admission (1 year) C-statistic Sensitivity Specificity	15% hospitalised once or more	Retrospective cohort
Donnan 2008[64,65]	Predicting Emergency Admissions Over the Next Year (PEONY)	n = 90, 879 Adults (aged 40 years or over; deviation cohort mean age 61.5 years) Multimorbidity: number of people with multimorbidity not reported UK	Emergency admission in (1 year) C-statistic Sensitivity Specificity PPV	Not reported for validation cohort, n=6793 (7.45%) in derivation cohort	Retrospective cohort
Hippisley-Cox 2013[107,108]	QAdmissions: GP data alone – QResearch cohort QAdmissions: HES-GP linked-data – CPRD cohort QAdmissions: GP data alone – QResearch cohort QAdmissions: HES-GP linked-data – CPRD cohort	n=3,815,982 (1,340,622 QResearch, 2,475,360 CPRD) Adults (aged 18-100; mean age 47.8±18.6), living in the community Multimorbidity: number of people with multimorbidity not reported UK	Emergency admission to hospital (1 and 2 years) C-statistic Pseudo R^2 Sensitivity Specificity	n=132,723 (9.9%, QResearch cohort)	Prospective cohort
Jensen 2001[116]	Probability of Repeated Admission (Pra)	n=386	Admissions (1 year)	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
		Older adults (aged 65 or over) Multimorbidity: number of people with multimorbidity not reported USA	Sensitivity Specificity		
Mazzaglia 2007[144]	Unnamed; 7-item questionnaire including questions about: age; sex; hospitalisations in past 6 months; polypharmacy (≥5 prescriptions)	n=2926 Older adults (aged 65 and older; mean age 75.1±7.2), living in the community Multimorbidity: number of people with multimorbidity not reported Italy	Hospitalisation (15 months) C-statistic	17.2%	Prospective cohort
Ritt 2015[195]	Clinical Frailty Scale Frailty Phenotype	n=307 Older adults (aged 65 years or over), inpatients admitted to a geriatric ward Multimorbidity: number of people with multimorbidity not reported Germany	Unplanned hospital admission (6 months) C-statistic	Not reported	Prospective cohort
Schneeweiss 2001[211]	Chronic Disease Score (CDS-1) Chronic Disease Score (CDS-2) Deyo Charlson Comorbidity Index (CCI) D'Hoore Charlson Comorbidity Index (CCI) Romano Charlson Comorbidity Index	n=141,161 Older adults (aged 65 years or older; mean age 75.4±6.7), living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Emergency hospitalisation (1 year) C-statistic	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	(CCI) Ghali Charlson Comorbidity Index (CCI)				
Soong 2015[221]	Charlson Comorbidity Index (CCI) Patients At Risk of Readmission 30-Day (PARR30) Risk Index for Geriatric Acute Medical Admission (RIGAMA) Cardiovascular Health Study (CHS) model Study of Osteoporotic Fractures (SOF) model Avila-Funes Frailty Index (36-item) Identifying Seniors at Risk (ISAR)	n=2099252 Older adults (aged 65 years or over), discharged after acute emergency admission Multimorbidity: number of people with multimorbidity not reported UK	ED readmission (30-90 days) C-statistic	Not reported	Retrospective cohort
Susser 2008[223]	Charlson Comorbidity Index (CCI): self-report and administrative versions	n=520 Older adults (aged 65 years or over), ready to be discharged from emergency department Multimorbidity: number of people with multimorbidity not reported Canada	Health services utilisation (ED visits, 2+ visits/5 months) C-statistic	Not reported	Retrospective cohort
Wallace 2013 (3	Probability of Repeated	n=8843	Hospital admission (Boult	n=2117 (25.1%)	Systematic review

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
studies included in analysis: Boult 1995[28], Mosley 2009[154], Wagner 2006[238])	Admission (Pra)	Older adults (aged 65 years or older), living in the community Multimorbidity: number of people with multimorbidity not reported Boult 1995: USA Mosley 2009: USA Wagner 2006: UK, Germany, Switzerland	1995, 4 years; Mosley 2009, 1 year; Wagner 2006, 1 year) C-statistic Sensitivity Specificity		
Wallis 2015[240]	CCI CHSA Clinical Frailty Scale	n=5764 Older adult (aged 75 years or over; mean age 84.3±5.9 years), living in the community with previous ED admission Multimorbidity: number of people with multimorbidity not reported UK	ED readmission (30 days) C-statistic	n=759 (13.17%)	Retrospective cohort
Widagdo 2015[243]	Frailty phenotype Simplified frailty phenotype Frailty Index (39-item) Prognostic Frailty Score	n=2087 Older adults (aged 70 years or over; mean age 77±6), majority living in the community (3.3% living in care facility) Multimorbidity: number of people with multimorbidity not reported Australia	Hospitalisation (3 years) Sensitivity Specificity C-statistic	Frailty phenotype n=404 (30.1%) Simplified frailty phenotype n=292 (28.4%) Frailty Index n=513 (30.6%) Prognostic Frailty Score n=379 (29.8%)	Retrospective cohort
Zekry 2012B[247]	Charlson Comorbidity Index	n=444	Rehospitalisation (1 year)	Hospitalised once: 82 (18.5%)	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	(CCI) Cumulative Illness Rating Scale, geriatric adaption (CIRS-G) Index of Coexistent Disease (ICED) Kaplan scale Geriatric Index of Comorbidity (GIC) Chronic Disease Score (CDS-1)	Older adults (aged 75 or older), hospitalised and discharged Multimorbidity: number of people with multimorbidity not reported Switzerland	Pseudo R^2		
Zeng 2014[248]	Quan Charlson Comorbidity Index Quan cumulative Charlson Comorbidity Index Quan baseline Charlson Comorbidity Index Quan Charlson Comorbidity Index trajectory: linear model Quan Charlson Comorbidity Index trajectory: quadratic model	n=13163 Older adults (aged 65 or older), living in the community Multimorbidity: 100% (3 or more chronic conditions) USA	Inpatient admission (10 years) C-statistic	Not reported	Retrospective cohort

7.2.2 Discrimination

Table 36: Clinical evidence profile: risk tools for identifying unplanned hospital admissions in people with multimorbidity

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Avila-Funes	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.55	LOW
Cardiovascular Health Study (CHS) model	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.52	LOW
Community Assessment Risk Screen (CARS) (≥4)	1	500	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.64	0.64	0.69	LOW
Charlson Comorbidity Index (CCI)	2	2105016	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.59 0.54 (0.52-0.56)	LOW
CCI (D'Hoore, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.58	LOW
CCI (D'Hoore, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.60	LOW
CCI (Deyo, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.58	LOW

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
CCI (Deyo, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.60	LOW
CCI (Ghali, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.56	LOW
CCI (Ghali, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.58	LOW
CCI (Quan, baseline)	1	13163	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.65	LOW
CCI (Quan, cumulative)	1	13163	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.65	LOW
CCI (Quan, ICD-10 codes)	1	13163	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.65	LOW
CCI (Quan, trajectory: linear model)	1	13163	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.65	LOW
CCI (Quan, trajectory: quadratic model)	1	13163	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.65	LOW
CCI (Romano, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.58	LOW
CCI (Romano, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.60	LOW

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Chronic Disease Score (CDS-1)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.59	LOW
Chronic Disease Score (CDS-2)	1	141161	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.61	LOW
CHSA Clinical Frailty Scale	2	6071	HIGH[a]	No serious inconsistency[b]	Serious indirectnessc	No serious imprecison[e]	-	-	0.54 (0.50-0.64)	LOW
Cumulative Illness Rating Scale (CIRS) (>3)	1	560	HIGH[a]	Not applicable[b]	Serious indirectnessc	No serious imprecision[d]	0.614	0.593	0.61 (0.57-0.66)	LOW
Dutch Tilburg Frailty Indicator (≥4)	1	430	HIGH[a]	Not applicable[b]	Serious indirectnessc	No serious imprecision[d]	0.53 (0.41-0.64)	0.65 (0.60-0.70)	0.60 (0.52-0.67)	LOW
eFI (Moderate and above, internal cohort, 3yrs)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.26	0.88	0.64	MODERATE
eFI (Moderate and above, external cohort, 3yrs)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.36	0.82	0.69	MODERATE
eFI (Severe, internal cohort, 3 years)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.06	0.98	0.64	MODERATE
eFI (Severe, external cohort, 3 years)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.10	0.97	0.69	MODERATE
eFI (Moderate and above, internal cohort,	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.31	0.86	0.66	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
1yr)										
eFI (Moderate and above, external cohort, 1 year)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.42	0.81	0.71	MODERATE
eFI (Severe, internal cohort, 1yr)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.08	0.97	0.66	MODERATE
eFI (Severe, external cohort, 1yr)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.13	0.97	0.71	MODERATE
Frailty Index (36-item)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectnessc	Not estimable[e]	-	-	0.57	LOW
Frailty Index (39-item) (≥0.25)	1	2087	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.238	0.881	0.56	LOW
Frailty Phenotype (≥3)	2	1973	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable[d]	0.99	0.938	0.5 (0.432-0.568)	LOW
Simplified Frailty Phenotype (≥2)	1	1173	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.034	0.989	0.51	LOW
Groningen Frailty Indicator (≥5)	1	430	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.52 (0.40-0.64)	0.54 (0.50-0.58)	0.54 (0.46-0.61)	LOW
Hospitalised Older Patient Examination (HOPE) Index (≥4)	1	3043	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.882	0.167	0.60 (0.56-0.63)	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
HOPE Index (≥8)	1	3043	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.656	0.55	0.60 (0.56-0.63)	MODERATE
Patients At Risk of Readmission 30-Day (PARR30)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.7	LOW
Predicting Emergency Admissions Over the Next Year (PEONY) (>20)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.761	0.695	0.79	LOW
PEONY (>23)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.689	0.774	0.79	LOW
PEONY(>32)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.42	0.926	0.79	LOW
PEONY (>37)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.271	0.998	0.79	LOW
PEONY (>46)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.079	0.996	0.79	LOW
PEONY (>50)	1	90552	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.042	0.998	0.79	LOW
Probability of Repeated Admission (Pra) (≥0.3)	5	3869729	HIGH[a]	No serious inconsistency[b]	Very serious indirectness[c]	Not estimable[d]	0.52	0.713	0.67 (0.642 – 0.752)	VERY LOW

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Pra (≥0.5)	2	9343	HIGH[a]	No serious inconsistency[b]	Very serious indirectness[c]	No serious imprecision[d]	0.12 (0.105 - 0.136)[f]	0.96 (0.958-0.967)[f]	0.67 (0.642 – 0.752)	VERY LOW
Prognostic Frailty Score (≥3)	1	1485	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.586	0.583	0.58	LOW
QAdmissions – GP data alone (1yr, >7% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.58	0.82	-	MODERATE
QAdmissions – GP data alone (1yr, >12% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.40	0.92	-	MODERATE
QAdmissions – GP data alone (1yr, >18% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.26	0.96	-	MODERATE
QAdmissions – GP data alone, (1yr, >28% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.13	0.99	-	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
QAdmissions – GP data alone, (1yr, >36% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.07	0.99	-	MODERATE
QAdmissions – GP data alone (2yr, >13% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.555	0.84	-	MODERATE
QAdmissions – GP data alone (2yr, >21% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.37	0.93	-	MODERATE
QAdmissions – GP data alone (2yr, >31% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.23	0.97	-	MODERATE
QAdmissions – GP data alone, (2yr, >46% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.11	0.99	-	MODERATE
QAdmissions – GP data alone, (2yr, >57% risk,	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.06	0.99	-	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
QRes cohort)										
QAdmissions – HES-GP linked data (2yr, >13% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.569		-	MODERATE
QAdmissions – HES-GP linked data (2yr, >23% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.393		-	MODERATE
QAdmissions – HES-GP linked data (2yr, >36% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.246		-	MODERATE
QAdmissions – HES-GP linked data, (2yr, >69% risk, QRes cohort)	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.066		-	MODERATE
QAdmissions – HES-GP linked data (2yr, >13% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.568		-	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
QAdmissions – HES-GP linked data (2yr, >22% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.394		-	MODERATE
QAdmissions – HES-GP linked data (2yr, >35% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.249		-	MODERATE
QAdmissions – HES-GP linked data, (2yr, >68% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.067		-	MODERATE
QAdmissions – GP data alone (2yr, >14% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.551		-	MODERATE
QAdmissions – GP data alone (2yr, >24% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.374		-	MODERATE
QAdmissions – GP data alone (2yr, >36% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.232		-	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
QAdmissions – GP data alone, (2yr, >66% risk, CPRD cohort)	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.061		-	MODERATE
QAdmissions – GP data alone, men, QRes cohort	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.769 (0.767-0.771)	MODERATE
QAdmissions – GP data alone, women, QRes cohort	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.764 (0.762-0.766)	MODERATE
QAdmissions - HES-GP linked data, men, QRes cohort	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.776 (0.774-0.778)	MODERATE
QAdmissions – HES-GP linked data, women, QRes cohort	1	1340622	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.773 (0.771-0.774)	MODERATE
QAdmissions – GP data alone, men, CPRD cohort	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.767 (0.765-0.768)	MODERATE
QAdmissions – GP data	1	246536	LOW[a]	Not applicable[b]	Serious	No serious	-	-	0.764	MODERA

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
alone, women, CPRD cohort		0			indirectness[c]	imprecision[e]			(0.763-0.766)	TE
QAdmissions - HES-GP linked data, men, CPRD cohort	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.772 (0.771-0.774)	MODERATE
QAdmissions – HES-GP linked data, women, CPRD cohort	1	2465360	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.771 (0.770-0.773)	MODERATE
Risk Index for Geriatric Acute Medical Admission (RIGAMA)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.55	LOW
Rothman	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.53	LOW
Sherbook Postal Questionnaire (≥2)	1	430	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.76 (0.65-0.85)	0.44 (0.39-0.49)	0.60 (0.53-0.67)	MODERATE
Study of Osteoporotic Fractures (SOF) model	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.53	LOW
Unnamed (7-item questionnaire)	1	2926	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.67 (0.65-0.70)	LOW
Unweighted disease count (>3)	1	560	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.667	0.535	0.63 (0.58-0.67)	LOW

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality

(a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Downgraded because the majority of the evidence included an indirect population (downgraded by 1 increment) or a very indirect population (downgrade by 2 increments)

(d) Sensitivity and specificity data for individual studies is not reported; therefore data cannot be meta-analysed.

Narrative findings

One study (Wagner 2006) reported incomplete discrimination data for the Pra tool. The authors reported the number of participants who had an unplanned admission and how many were predicted to do so by the Pra tool at a threshold of 0.5 or more, however, they did not report the actual number of true or false positives and negatives. From this data, we estimated that the sensitivity of the tool was 11.3% and the specificity of the tool was 100%; however, this method of estimation is subject to a high risk of bias and should be interpreted with caution.

Table 37: Clinical evidence profile: risk tools for identifying 2 or more unplanned admissions in people with multimorbidity

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic Pooled/Median (range)	Quality
CCI, self-report	1	520	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	-	-	0.64 (0.58-0.69)	LOW
CCI, administrative	1	520	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Serious imprecision[d]	-	-	0.65 (0.59-0.70)	LOW

Risk tool	No of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (%)	Specificity (%)	C-statistic Pooled/Median (range)	Quality
Pra (≥0.3)	1	2174	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[d]	-	-	0.696	LOW
Pra(≥0.5)	1	2174	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[d]	-	-	0.696	LOW

a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)

d) Imprecision was assessed based on inspection of the confidence region of the specificity values or, where specificity was not reported, sensitivity or C-statistic values (in order of preference). As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of the confidence interval around sensitivity was considered not imprecise, 0.2-0.4 serious imprecision, and >0.4 very serious imprecision. Imprecision was not estimable where studies did not report confidence intervals.

7.2.3 Calibration

Table 38: Clinical evidence profile: risk tools for identifying adverse outcomes (unplanned hospital admissions) in people with multimorbidity

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	Pseudo R^2 (95%CI)	Brier score (95%CI)	D statistic	RR (p/o)	Quality
QAdmissions – HES-GP linked data, women	1	4190003	LOW[a]	Serious indirectness[b]	Not estimable	40.6%	-	-	-	MODERATE

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	Pseudo R^2 (95%CI)	Brier score (95%CI)	D statistic	RR (p/o)	Quality
QAdmissions – HES-GP linked data, men	1	4190003	LOW[a]	Serious indirectness[b]	Not estimable	42.65%	-	-	-	MODERATE
QAdmissions – GP data alone, women	1	4190003	LOW[a]	Serious indirectness[b]	Not estimable	37.3%	-	-	-	MODERATE
QAdmissions – GP data alone, men	1	4190003	LOW[a]	Serious indirectness[b]	Not estimable	39.5%	-	-	-	MODERATE
CCI	1	444	HIGH[a]	Serious indirectness[b]	Not estimable	3.1%	-	-	-	LOW
CIRS-G	1	444	HIGH[a]	Serious indirectness[b]	Not estimable	5.6%	-	-	-	LOW
eFI (Internal cohort, 3yrs)	1	207720	LOW	Serious indirectness[c]	Not estimable[e]	0.05	-	-	-	MODERATE
eFI (External cohort, 3yrs)	1	516007	LOW	Serious indirectness[c]	Not estimable[e]	0.02	-	-	-	MODERATE
eFI (Internal cohort, 1yr)	1	207720	LOW	Serious indirectness[c]	Not estimable[e]	0.03	-	-	-	MODERATE
eFI (External cohort, 1yr)	1	516007	LOW	Serious indirectness[c]	Not estimable[e]	0.02	-	-	-	MODERATE
ICED	1	444	HIGH[a]	Serious indirectness[b]	Not estimable	0.4%	-	-	-	LOW
Kaplan scale	1	444	HIGH[a]	Serious indirectness[b]	Not estimable	0.5%	-	-	-	LOW
GIC	1	444	HIGH[a]	Serious	Not estimable	14.0%	-	-	-	LOW

Risk tool	No of studies	n	Risk of bias	Indirectness	Imprecision	Pseudo R^2 (95%CI)	Brier score (95%CI)	D statistic	RR (p/o)	Quality
				indirectness[b]						
Chronic Disease Score (CDS-1)	1	444	HIGH[a]	Serious indirectness[b]	Not estimable	1.7%	-	-	-	LOW

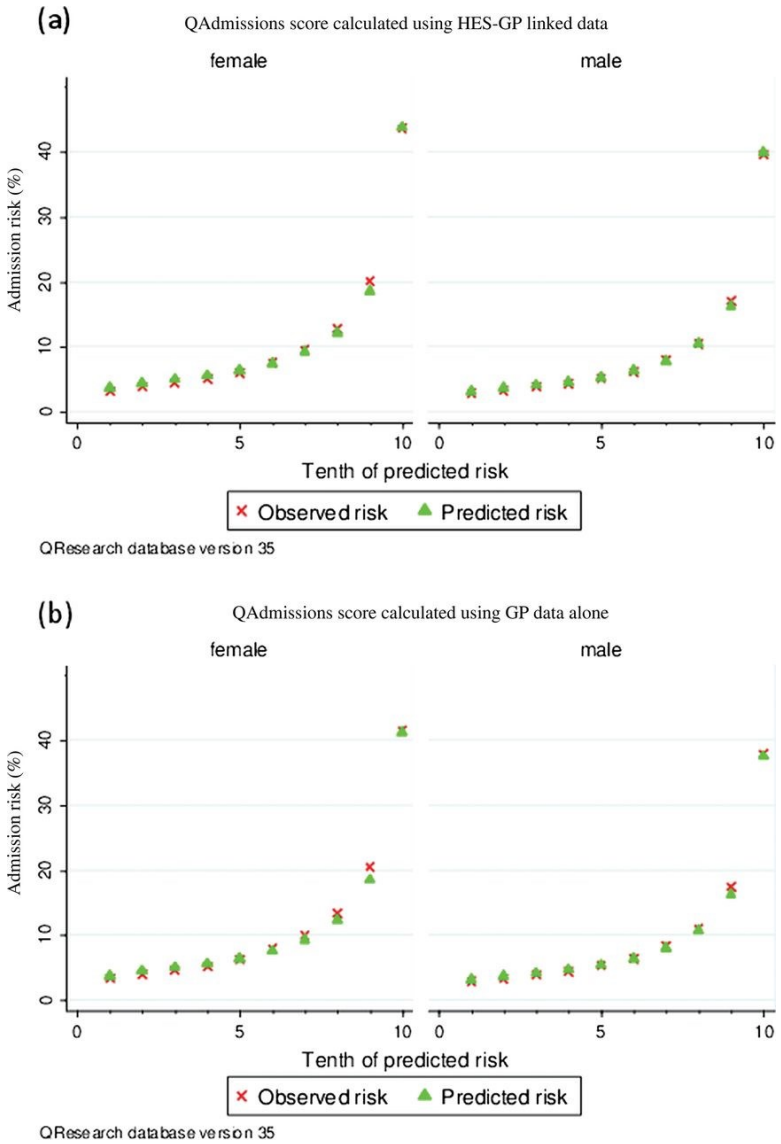
a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

b) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)

Narrative findings

One study (Hippisley-Cox 2013) provided plots demonstrating the calibration performance of the QAdmissions score using HES-GP linked data and GP data alone. These plots indicate that both tools reliably predict emergency hospital admissions. This evidence is at very high risk of bias.

Figure 4: Calibration plot (reproduced from Hippisley-Cox 2013, with permission)



7.2.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.2.5 Evidence statements

Clinical

Eighteen studies that evaluated 36 risk tools were included in the review. Of these, the GDG noted that the majority of the tools demonstrated poor discrimination and calibration for predicting unplanned hospital admissions. The majority of evidence was of low quality. The GDG identified 4 tools that demonstrated moderate accuracy in predicting unplanned hospital admission (prioritising, specificity and calibration data):

- Predicting Emergency Admissions Over the Next Year (PEONY): Low quality from 1 study with 90,552 adults (aged 40 years or over) demonstrated that PEONY had a specificity of 0.695-0.998 and sensitivity of 0.042-0.761 at thresholds ranging between less than 23 and less than 50 and a moderate C-statistic value (0.79); calibration data was not reported for this tool.
- QAdmissions (using GP data): Moderate quality evidence from 1 study with 3,815,982 adults (aged 18 years or over) demonstrated that QAdmissions (using GP data) had a specificity of 0.82-0.99 and sensitivity of 0.06-0.58 at thresholds ranging between less than 7 and less than 57, and moderate C-statistic (0.76) and calibration values (pseudo R^2 37.3-39.5%).
- Patients At Risk of Readmission 30-Day (PARR30): Low quality from 1 study with 2,099,255 older adults demonstrated that PARR30 had a moderate C-statistic value (0.7); but no sensitivity, specificity or calibration data were reported.
- Electronic frailty index (eFI): Moderate quality from 1 study with 732,727 older adults demonstrated that eFI had a moderate C-statistic value (ranging from 0.64 to 0.71 depending on cohort and time point) and sensitivities and specificities ranging from 0.06 to 0.42 and from 0.81 to 0.98 respectively (depending on cohort, threshold and time point).

Economic

- No relevant economic evaluations were identified.

7.2.6 Recommendations and link to evidence

Recommendations	**9. Consider using a validated tool such as eFI, PEONY or QAdmissions, if available in primary care electronic health records, to identify adults with multimorbidity who are at risk of adverse events such as unplanned hospital admission or admission to care homes.** **10. Consider using primary care electronic health records to identify markers of increased treatment burden such as number of regular medicines a person is prescribed.**
Relative values of	The GDG was interested in the prognostic accuracy of risk tools to identify people

different outcomes	with multimorbidity who have a higher risk of unplanned hospital admissions. The GDG wanted to identify a tool which would be able to identify people with multimorbidity who may benefit from additional support, and to inform decisions between people with multimorbidity and clinicians about the optimisation and prioritisation of treatment. The GDG agreed that the relative value of sensitivity and specificity was dependent on how the tool was intended to inform the care of an individual patient. The GDG were clear that while the recommendations in the guideline might have an effect on the rate of unplanned admissions, due to a change in clinical management, the tools were not being used in order to influence hospital admission per se. The main function of the tool in this context was to identify people with multimorbidity who might benefit form a review of their care with a view to optimising their care. This might involve the discontinuation of treatment and as such the specificity of the tool was considered important so as to reduce the risk of people at low risk being referred having treatment withdrawn unnecessarily. However, the GDG also felt that a high sensitivity was important so that people who are at higher risk of unplanned admissions are not missed (fewer false negatives), therefore ensuring that these individuals gain access to treatment that may reduce their risk of adverse events and improve their quality of life. Many of the studies included in the review reported the C-statistic. The GDG felt that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information to make a recommendation.
Trade-off between clinical benefits and harms	Evidence for a large number of tools, each evaluated in very few studies, was identified. Of these tools, only 11 tools reported sensitivity and specificity data: CARS, Dutch Tilburg Frailty Indicator, Groningen Frailty Indicator, HOPE, PEONY, Pra, QAdmissions (GP data linked), Sherbook Postal Questionnaire, eFI and an unweighted disease count. Specificity data was not available for the QAdmissions CPRD cohort nor for the HES-GP linked data for the QAdmissions QResearch cohort. The GDG agreed that the majority of the risk tools demonstrated poor accuracy in identifying people who are at risk of an unplanned hospital admission. The GDG agreed that the PEONY tool and QAdmissions (GP linked data) emerged as the most accurate risk tools identified in the review. PEONY demonstrated moderate specificity and sensitivity at lower thresholds, with excellent specificity and poor sensitivity at higher thresholds. The GDG noted that this tool had a moderate C-statistic value. No reclassification data was reported for this tool. QAdmissions (GP data linked) demonstrated excellent specificity but poor sensitivity. Sensitivity and specificity for QAdmissions (GP data linked) was reported at a number of thresholds and as the threshold increased, the specificity values increased and the sensitivity decreased The QAdmissions tool (GP data linked) explained 37.3-39.5% of the variance in unplanned hospital admissions (women/men, pseudo R2). The GDG noted that the eFI tool appeared to perform slightly worse than PEONY or QAdmissions but was equivalent or better than all other tools tested. The eFI tool also had evidence supporting its use in predicting other adverse outcomes that QAdmissions and PEONY were not used for (see the care home admission and life expectancy reviews). As the eFI tool is currently widely available the GDG decided it would be appropriate to recommend it alongside PEONY and QAdmissions, particularly as all three tools were being used more as a proxy to identify those with multimorbidity at risk of adverse outcomes rather than to specifically predict admissions or any other single outcome.
Economic considerations	No relevant economic evaluations were identified. The GDG considered the cost associated with using health electronic records or risk tools to identify people who are at risk of adverse events and noted that these are generally not associated with any licencing cost although some may require a specific software installation with its associated costs depending on the systems used in GP practices. There are no significant costs associated with using electronic health records as the healthcare professional's time associated is minimal (less than a minute) and there are

	potential benefits of identifying people who may benefit from further care triggered by the risk assessment. Furthermore, the GDG considered the complexity of the tool when making recommendations and the availability of the tools in current practice. The GDG advised that QAdmissions is the most commonly used tool on electronic GP systems and it is accessible online free of charge if it is not already integrated in the GP computer system. For these reasons, QAdmissions was recommended over other risk tools. PEONY showed better accuracy in the clinical review but may have limited implementation in clinical practice outside Scotland where it has been developed.
Quality of evidence	The GDG expressed concern about the limited availability of sensitivity and specificity data for assessing the accuracy of risk tools to predict unplanned hospital admissions in people with multimorbidity. Some studies did provide sensitivity and specificity data, which is a metric of the ability of the tool to predict those who will have and will not have the event. However, the majority of studies only reported a C-statistic value, which provides an overall estimate of the accuracy of the tool, but does not give an indication of the number of false positive and false negative diagnoses that will be made if the tool was used in practice. The GDG noted that the majority of data included in the review was of low quality. The majority of studies were at a high risk of bias, due to risk of bias in sampling and poor outcome reporting. The GDG noted that only 3 risk tools were validated within the UK (PEONY, QAdmissions and eFI). The GDG discussed how tools are developed and validated abroad may have limited applicability to UK practice. The GDG discussed the studies conducted in Europe, Australia and Canada, and agreed that these countries has similar health systems to the UK and so did not downgrade these studies for indirectness. Studies that were conducted in the USA were downgraded for indirectness due to differences between the health system in the USA and UK. The majority of studies did not report whether the people had multimorbidity. The GDG decided to downgrade studies with an unclear number of people with multimorbidity for indirectness because of uncertainty that the same tools would be accurate at predicting unplanned hospital admission in a general population of people with multimorbidity. The GDG were interested in the accuracy of risk tools to predict unplanned hospital admissions within 3 years, as they felt that people with multimorbidity who experience unplanned hospital admissions within that timeframe may most benefit from an approach to care that takes into account multimorbidity. However, due to the scarcity of evidence identified for the review the GDG decided to consider evidence for tools that predicted unplanned hospital admissions for greater than 3 years. The GDG considered that tools that were identified as being accurate at predicting unplanned hospital admissions during a longer timeframe would also be accurate at identifying people with unplanned hospital admissions within 3 years, and therefore decided not to downgrade for indirectness.
Other considerations	The GDG discussed whether QAdmissions, eFI, PARR 30 and PEONY were sufficiently accurate to be used to identify people with multimorbidity who are at risk of unplanned hospital admissions. The GDG noted that no sensitivity or specificity data were reported for the PARR 30 tool, and that PARR 30 was also not validated within the UK. As a consequence, the GDG did not believe that there was sufficient evidence to recommend the use of PARR 30. Some of the GDG members were concerned about the low sensitivity of QAdmissions, eFI and PEONY when thresholds of the tools are used which favour specificity. The majority of the GDG agreed to recommend that healthcare professionals may consider using either tool to identify people with multimorbidity who may benefit from a multimorbidity approach. This decision was informed by awareness that QAdmissions, eFI and PEONY are currently used in clinical practice, and so a recommendation would be relatively easy to implement for many practices. QAdmissions is the most commonly used tool on electronic GP systems. The GDG noted that the GP linked data only version of the

	QAdmissions tool can be automatically populated using data solely from GP computer systems. A stand-alone version is available and can be used to assess individual patients; or it can be integrated into GP clinical computer systems by the system suppliers, similar to other risk prediction tools such as QRISK2. PEONY is currently used in practice mainly in Scotland. The GDG had some concerns that recommending an unplanned hospital admissions tool would be misunderstood and that it could be misinterpreted that they believed that unplanned admission rates for people with multimorbidity could be improved by more holistic care. They acknowledged however current policy initiatives to review care for people at high risk of unplanned admission and that linking holistic care for people with multimorbidity to such policy initiatives could provide traction for implementation of better care for this group. The GDG discussed whether they could recommend a specific threshold on both tools to identify people who may be at risk of unplanned hospital admissions and therefore may benefit from an approach to care that takes into account multimorbidity. The GDG agreed that healthcare professionals may wish to alter the threshold at which they identify risk, according to the how the tool would be used and resources available. The decision to recommend eFI as a potential tool to identify people who may benefit from a multimorbidity approach was made on the basis of its prediction of admission to a care facility which is discussed in section 7.4.5.

7.3 Health-related quality of life

7.3.1 Review question: What risk tool best identifies people with multimorbidity who are at risk of reduced health-related quality of life

For full details see review protocol in Appendix C.

Table 39: PICO characteristics of review question

Question	What risk tool best identifies people with multimorbidity who are at risk of reduced health-related quality of life?
Population	Adults (aged >17 years) with multimorbidity
Risk tool	Risk tools identified in the literature validated for predicting reduced health-related quality of life in people with multimorbidity.
Target condition or Reference standard	Reductions in health related quality of life (max time point = 3 years)
Outcomes (in terms of predictive test accuracy, calibration)	Area under the curve (c-statistic) Sensitivity, specificity, predictive values Predicted risk versus observed risk (calibration) Other outcomes for example, Somers' D statistic, R^2 statistic and Brier score Reclassification
Study types	Prospective and retrospective cohort studies (external validation, internal validation (split half validation))

The GDG discussed the objectives of this review and agreed that where discrimination data for a tool was found, they would prioritise specificity data for decision-making (that is, the ability of the tool to correctly identify people who were not at risk of reduced quality of life). This is because they wanted to ensure that people who were not at risk of reduced quality of life were not identified as requiring additional assessment and support, which may be associated with significant resource implications. However, the GDG believed that sensitivity of the tool (that is, the ability of the tool to correctly

identify people who are at risk of unplanned hospital admission) was also important, so as to ensure that individuals at a high risk of adverse events are assessed for any additional support they may require. As a consequence, the GDG prioritised higher specificity but expected the tool to have adequate sensitivity to recommend its use in practice.

We sought studies that evaluated the accuracy of prognostic risk tools in predicting quality of life, in order to identify people at risk of declining quality of life. Two studies evaluating 4 risk tools were included in the review[79,100] these are summarised in Table 40 below. It was not possible to pool studies and as a consequence, results are presented individually.

One study evaluated a risk tool with population of people with multimorbidity[79]; 1 study was conducted with an older adult population.[100] The number of reductions in health-related quality of life was not reported. Follow up of the studies was not reported.

Evidence from these studies is summarised in the clinical evidence summary below (Table 2). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, and excluded studies list in Appendix L. Full details of the tools included in this review are provided in the clinical evidence tables in Appendix H.

Summary of included studies

Table 40: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
Fortin 2005a[1]	Cumulative Illness Rating Scale (CIRS) Functional Comorbidity Index (FCI) Charlson comorbidity Index (CCI)	n=238 Adults (aged 18 years or over; mean age 59±14.3 years), living in the community Multimorbidity: 100% (mean number of conditions 5.3±2.8) Canada	SF-36 (6 months) R[2]	Not reported	Retrospective cohort
Grimmer 2014[2]	Hospital Admission Risk Profile (HARP)	n=148 Older adults (aged 65 years or over; mean age males 77.8 years, females 74.9 years) Multimorbidity: number of people with multimorbidity not reported Australia	SF-12 low physical component score (PCS) (1 and 3 months) SF-12 low mental component score (MCS) (1 and 3 months) SF-12 low or declining health related quality of life (combined PCS and MCS) (2 months)	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
			NB. definition of 'low' and 'declining' not reported Sensitivity Specificity C-statistic		

7.3.2 Discrimination

Table 41: Clinical evidence profile: risk tools for identifying people with multimorbidity who are at risk of reduced health-related quality of life

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Hospital Admission Risk Profile (HARP) (≥1)	1	148	HIGH [a]	Not applicable[b]	Serious indirectness [c]	Serious imprecision[d]	Low MCS at 3 months: 0.448 (0.326- 0.574) Low PCS at 3 months: 0.572 (0.443- 0.677) Low or declining PCS and MCS over 2 months: 0.538 (0.493- 0.673)	Low MCS at 3 months: 0.573 (0.459- 0.682) Low PCS at 3 months: 0.661 (0.548- 0.745) Low or declining PCS and MCS over 2 months: 0.585 (0.493- 0.673)	Low MCS at 3 months: 0.51 (0.43-0.59) Low PCS at 3 months: 0.62 (0.51-0.68) Low or declining PCS and MCS over 2 months: 0.56 (0.48-0.64)	VERY LOW

(a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Downgraded because the majority of the evidence included an indirect population (downgraded by 1 increment) or a very indirect population (downgrade by 2 increments)

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

7.3.3 Calibration

Table 42: Clinical evidence profile: risk tools for predicting health-related quality of life in inpatient or discharged populations

Risk tool	No. of studies	n	Risk of bias	Indirectness	Imprecision	(Partial) R^2 (95%CI)a	Brier score (95%CI)	D statistic	Quality
Cumulative Illness Rating Scale (CIRS)	1	238	HIGH[b]	No serious indirectness[c]	Not estimable[d]	PCS: 17.75% MCS: 0.75%	-	-	MODERATE
Functional Comorbidity Index (FCI)	1	238	HIGH[b]	No serious indirectness[c]	Not estimable[d]	PCS: 11.81% MCS: 0.02%	-	-	MODERATE
Charlson Comorbidity Index	1	238	HIGH[b]	No serious indirectness[c]	Not estimable[d]	PCS: 5.46% MCS: 2.80%	-	-	MODERATE

(a) The partial R^2 represents the proportion of variance in the outcome explained by the risk tool over and above that explained by age, gender, self-perceived social support and self-perceived economic status.

(b) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1increment) or a very indirect population (downgrade by 2 increments)

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

7.3.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.3.5 Evidence statements

Clinical

Two studies that evaluated 4 risk tools were included in the review. Of these, the GDG noted that all of the tools demonstrated poor discrimination and calibration for predicting reduced health-related quality of life. The evidence was of low to very low quality. Evidence was identified for the following tools:

- Hospital Admission Risk Profile (HARP): low quality evidence from 1 study with 148 older adults showed that HARP (score ≥1) has a specificity of 0.573 and a sensitivity of 0.448, and a C-statistic of 0.51 for predicting a low SF-12 mental component score in people with multimorbidity; a specificity of 0.661 and a sensitivity of 0.572, and a C-statistic of 0.62 for predicting a low SF-12 physical component score in people with multimorbidity and a specificity of 0.585 and a sensitivity of 0.538, and C-statistic of 0.56 for predicting low or declining SF-12 score (mental and physical components) in people with multimorbidity. Calibration data was not reported for this tool.
- Cumulative Illness Rating Scale (CIRS): Moderate quality evidence from 1 study with 238 people with multimorbidity showed that CIRS has an R^2% statistic of 17.5% for predicting health-related quality of life physical component score and 0.75% for predicting mental component score in people with multimorbidity. Discrimination data was not reported for this tool.
- Functional Comorbidity Index (FCI): Moderate quality evidence from 1 study with 238 people with multimorbidity showed that FCI has an R^2% statistic of 11.81% for predicting health-related quality of life physical component score and 0.02% for predicting mental component score in people with multimorbidity. Discrimination data was not reported for this tool.
- Charlson Comorbidity Index: Moderate quality evidence from 1 study with 238 people with multimorbidity showed that Charlson Comorbidity Index has an R^2% statistic of 5.46% for predicting health-related quality of life physical component score and 2.80% for predicting mental component score in people with multimorbidity. Discrimination data was not reported for this tool.

Economic

- No relevant economic evaluations were identified.

7.3.6 Recommendations and link to evidence

Recommendations	No recommendations made.
Relative values of different outcomes	The GDG was interested in the prognostic accuracy of risk tools to identify people with multimorbidity who are at risk of experiencing reductions in health-related quality of life within a 3 year timeframe. The GDG agreed that the relative value of sensitivity and specificity was dependent

	on how the tool was intended to inform the care of an individual patient. The main function of the tool in this context was to identify people with multimorbidity who might benefit form a review of their care with a view to optimising their care. This might involve the discontinuation of treatment and as such the specificity of the tool was considered important as people who are not at risk may be labelled as at risk inappropriately (false positives), and therefore have treatment withdrawn unnecessarily. People identified as at risk of reduced quality of life might also be considered for additional assessment and support which would have resource implications. However, the GDG also felt that a high sensitivity was important so that people who are at higher risk of reduced quality of life are not missed (fewer false negatives), therefore ensuring that these individuals gain access to treatment that may reduce their risk of adverse events and improve their quality of life.
Trade-off between clinical benefits and harms	Discrimination data was only reported for 1 tool (HARP). This tool demonstrated poor sensitivity, specificity, and C-statistic for predicting changes in both physical and mental components of quality of life. Calibration data was reported for 3 tools (CIRS, FCI, and Charlson Comorbidity Index). All of the tools also demonstrated poor calibration, with R^2 values ≤ 17.5%. No studies reported reclassification data. Overall the GDG agreed that no risk tools identified in this review demonstrated adequate accuracy for identifying people with multimorbidity who are at risk of experiencing reductions in health-related quality of life.
Economic considerations	No relevant economic evaluations were identified. The GDG considered the cost associated with using risk tools to identify people who are at risk of reduced HRQoL and noted that these are generally not associated with any licencing cost. The main cost associated with using this risk tool is additional healthcare professional time associated with completing them. If the specificity of a tool is low, there is a risk that a large number of people will be triggered for further care that they do not require (over-treatment), which would make the tool unlikely to be not cost-effective. Conversely, if the tool has low sensitivity then a large number of people will not be identified as being at risk of adverse outcomes, and therefore not receive the additional care they could benefit from. The GDG decided not to make a recommendation as the evidence on the accuracy of risk tools was inconclusive and this increases the uncertainty in the cost effectiveness of tools.
Quality of evidence	The GDG noted that the majority of data included in the review was of moderate or low quality. All of the tools were at a high risk of bias, due to risk of bias in sampling and poor outcome reporting. One of the studies was conducted in a population of people with multimorbidity. The other study was conducted in an older adult population with unclear numbers of people with multimorbidity; this study was downgraded for indirectness because of uncertainty that the same tools would be accurate at predicting mortality in a population of older people with multimorbidity compared to the a general population of people with multimorbidity. This is because some of the items in included in the tools may be more or less prevalent or important in people with multimorbidity Both of the studies were conducted in non-UK populations (Canada and Australia) but the GDG agreed that the studies were applicable to a UK health setting and decided not to downgrade for indirectness.
Other considerations	

7.4 Admission to a care facility

7.4.1 Review question: What risk tool best identifies people with multimorbidity who are at risk of admission to a care facility?

For full details see review protocol in Appendix C.

Table 43: PICO characteristics of review question

Question	What risk tool best identifies people with multimorbidity who are at risk of admission to a care facility?
Population	Adults (aged >17 years) with multimorbidity
Risk tool	Risk tools identified in the literature validated for predicting admission to care facility in people with multimorbidity.
Target condition or Reference standard	Admission to care facility (max time point = 3 years)
Outcomes (in terms of predictive test accuracy, calibration)	Area under the curve (c-statistic) Sensitivity, specificity, predictive values Predicted risk versus observed risk (calibration) Other outcomes for example, Somers' D statistic, R^2 statistic and Brier score Reclassification
Study types	Prospective and retrospective cohort studies (external validation, internal validation (split half validation))

The GDG was interested in identifying studies that evaluated the accuracy of risk tools for identifying individuals with multimorbidity who may be admitted to a care facility. The GDG agreed that they would prioritise specificity data for decision-making (that is, the ability of the tool to correctly identify people who were not at risk of admission to a care facility). This is because they wanted to ensure that people who were not at risk of admission to a care facility were not identified as requiring additional assessment and support, which may be associated with significant resource implications. Admission to a care facility in the UK is also an event that largely takes place in the last year or so of life. A tool might also therefore be used to identify people for whom optimisation of treatments might involve stopping preventative treatments and high specificity is helpful to ensure that these decisions are being made with the right group of patients. However, the GDG believed that the sensitivity of the tool (that is, the ability of the tool to correctly identify people who are at risk of admission to a care facility) was also important, so as to ensure that individuals at a high risk of an admission to a care facility are identified and can be considered for a multimorbidity approach to care. As a consequence, the GDG prioritised higher specificity, but expected the tool to have high sensitivity to recommend its use in practice.

Five studies evaluating 20 risk tools were included in the review[43,44,117,196,221,243]; these are summarised in Table 44 below. All of the studies were conducted with an older adult population who were either living in the community or had been recently discharged from hospital. The proportion of the sample who were admitted to a care facility was only reported in 2 studies[43,44,243] (1.5 – 1.9% and 2.5% of the study samples). Follow up of the studies ranged from 1 to 5 years. Evidence from these studies is summarised in the clinical evidence summary below (Table 45). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, and excluded studies list in Appendix L. Full details of the tools included in this review are provided in the clinical evidence tables in Appendix H.

The paper describing validation of the eFI[44] was published after the cut off date for literature review and was highlighted in stakeholder comments at consultation. This was included following

GDG discussion because of its particular relevance to the guideline population and that it predicted both admission to care home facility and hospitalisation.

Summary of included studies

Table 44: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
Clegg 2016[43,44]	Electronic Frailty Index (eFI) Unweighted deficit count (36 items)	n = 207720 (internal validation cohort) Older adults (aged 65 to 95) registered at relevant GPs Multimorbidity: number of people with multimorbidity not reported UK	Nursing home admission (3 years) C-statistics Sensitivity Specificity	5,239 by 3yrs (2.5%)	Retrospective cohort
Jones 2005[117]	CHSA Frailty Index (70-item) Comprehensive Geriatric Assessment-Frailty Index (FI-CGA)	n=3736 Older adults (aged 65 or older), living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Admission to care facility (5 years) C-statistic	Not reported	Prospective cohort
Rockwood 2005[196]	Cumulative Illness Rating Scale Modified Mini-Mental State Examination Canadian Study of Health and Aging (CSHA) rules-based definition of frailty CSHA Function Scale CSHA Frailty Index (70-item) CSHA Clinical Frailty Scale	n=2305 Older adults (aged 65 years or older) , living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Need for institutional care (5 years) C-statistic	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
Soong 2015[221]	CCI Risk Index for Geriatric Acute Medical Admission (RIGAMA) Cardiovascular Health Study (CHS) model Study of Osteoporotic Fractures (SOF) model Avila-Funes Frailty Index (36-item) Identifying Seniors at Risk (ISAR)	n=2099252 Older adults (aged 65 years or over), discharged after acute emergency admission Multimorbidity: number of people with multimorbidity not reported England	Admission to care facility (1 year) C-statistic	Not reported	Retrospective cohort
Widagdo 2015[243]	Frailty phenotype Simplified frailty phenotype Frailty Index (39-item) Prognostic Frailty Score	n=2087 Older adults (aged 70 years or over; mean age 77±6), majority living in the community (3.3% living in care facility) Multimorbidity: number of people with multimorbidity not reported Australia	Admission to care facility (3 year) Sensitivity Specificity C-statistic	Frailty phenotype n=22 (1.7%) Simplified frailty phenotype n=15 (1.5%) Frailty Index n=31 (1.9%) Prognostic Frailty Score n=21 (1.7%)	Retrospective cohort

7.4.2 Discrimination

Table 45: Clinical evidence profile: risk tools for identifying adverse outcomes (admission to care facility) in people with multimorbidity

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Avila-Funes	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.5	LOW
Cardiovascular Health Study (CHS) model	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.57	LOW
Charlson Comorbidity Index (CCI)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.62	LOW
CIRS	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.62	LOW
CSHA Clinical Frailty Scale	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.75	LOW
CSHA Function Scale	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.80	LOW
CSHA rule-based definition of frailty	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.70	LOW
CHSA Frailty Index Jones 2005	2	6041	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.72	LOW
eFI (Moderate and above, internal cohort)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.38	0.86	0.72	MODERATE

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
eFI (Severe, internal cohort)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.10	0.97	0.72	MODERATE
Frailty Index-Comprehensive Geriatric Assessment (FI-CGA)	1	3736	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.66	LOW
Frailty Index (36-item)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.55	LOW
Frailty Index (39-item)	1	2087	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.355	0.858	0.61	LOW
Frailty phenotype	1	1566	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.182	0.934	0.56	LOW
Simplified frailty phenotype	1	1173	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.067	0.983	0.56	LOW
Identifying Seniors at Risk (ISAR)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.65	LOW
Modified Mini-Mental State Examination	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.69	LOW
Prognostic frailty score	1	1485	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.762	0.548	0.66	LOW
Risk Index for Geriatric Acute Medical Admission (RIGAMA)	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.5	LOW

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Rothman	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.45	LOW
Study of Osteoporotic Fractures (SOF) model	1	2099252	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.44	LOW

(a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) C-statistic Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)

(d) Imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

(e) The judgement of precision was based on the median C-statistic value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

7.4.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.4.4 Evidence statements

Clinical

Five studies that evaluated 20 risk tools were included in the review. Of these, the GDG noted that the majority of the tools demonstrated poor discrimination and calibration for predicting admission to a care facility. The majority of evidence was of low to very low quality. The GDG identified 5 tools that demonstrated moderate accuracy in predicting admission to a care facility (prioritising specificity and calibration data):

- CSHA Clinical Frailty Scale: Low quality from 1 study with 2305 older adults demonstrated that CSHA Clinical Frailty Scale had a moderate C-statistic value (0.75); sensitivity, specificity and calibration data was not reported for this tool.
- CSHA Function Scale: Low quality from 1 study with 2305 older adults demonstrated that CSHA Function Scale had a moderate C-statistic value (0.80); sensitivity, specificity and calibration data was not reported for this tool.
- CSHA rule-based definition of frailty: Low quality from 1 study with 2305 older adults demonstrated that CSHA rule-based definition of frailty had a moderate C-statistic value (0.70); sensitivity, specificity and calibration data was not reported for this tool.
- CSHA Frailty Index: Low quality from 2 studies with 6041 older adults demonstrated that CSHA Frailty Index had a moderate C-statistic value (0.72-0.75); sensitivity, specificity and calibration data was not reported for this tool.
- eFI: Moderate quality from 1 study with 207720 older adults demonstrated that eFI had a moderate C-statistic value (0.72); sensitivity and specificity were 0.38 and 0.86 at the moderate threshold and 0.10 and 0.97 at the severe threshold; the pseudo R^2 value was 0.04.

Economic

- No relevant economic evaluations were identified.

7.4.5 Recommendations and link to evidence

Recommendations	**9. Consider using a validated tool such as eFI, PEONY or QAdmissions, if available in primary care electronic health records, to identify adults with multimorbidity who are at risk of adverse events such as unplanned hospital admission or admission to care homes.**
Relative values of different outcomes	The GDG was interested in the prognostic accuracy of risk tools to identify people with multimorbidity who have a higher risk of admission to a care facility. The GDG felt that avoidance of care facility admission was very important to many people, and that people who are at risk of admission to a care facility may benefit from additional support and a bespoke approach to care. The GDG agreed that the relative value of sensitivity and specificity was dependent on how the tool was intended to inform the care of an individual patient. The GDG considered 1 of the primary functions of the tool would be to identify individuals who may benefit from the discontinuation of treatment. As a consequence, they identified specificity as more critical to decision making than sensitivity, as people who are not at risk may be labelled as at risk inappropriately (false positives), and therefore be referred unnecessarily or have treatment withdrawn unnecessarily. However, the GDG also felt that a high sensitivity was important so that people who are at higher risk of unplanned admissions are not missed (fewer false negatives), therefore ensuring that these individuals gain access to treatment that may reduce their risk of adverse events and improve their quality of life. Many of the studies included in the review reported the C-statistic. The GDG felt that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information to make a recommendation.
Trade-off between clinical benefits and harms	Five tools identified in the review had moderate discrimination as assessed using C-statistic: the eFI, the Cumulative Deficit Model Frailty Index, CSHA function scale, CSHA clinical frailty scale and CSHA rules based definition of frailty. Only the eFI study provided data allowing for calculation of sensitivity and specificity., which is most informative to a recommendation on the use of a risk tool in clinical practice.
Economic considerations	No relevant economic evaluations were identified. The GDG considered the cost associated with using risk tools to identify people who are at risk of adverse events and noted that these are generally not associated with any licencing cost although some of them may require a specific software installation with its associated costs. The main cost associated with using this risk tool is additional healthcare professional time associated with completing them. If the specificity of a tool is low, there is a risk that a large number of people will be triggered for further care that they do not require (over-treatment), which would make the tool unlikely to be not cost-effective. Conversely, if the tool has low sensitivity then a large number of people will not be identified as being at risk of adverse outcomes, and therefore not receive the additional care they could benefit from.
Quality of evidence	The GDG noted the limited availability of prognostic accuracy data for the tools identified. Only one study reported sensitivity and specificity data, the majority only reported a C-statistic value. The C-statistic provides an overall estimate of the accuracy of the tool, but does not give an indication of the number of false positive and false negative diagnoses that will be made if the tool was used in practice. The GDG felt that they could not judge how useful the tools would be in clinical practice as the number of false negatives and positives that would occur when using the tools in practice is unclear. Therefore the GDG decided not to recommend the use of a tool to predict admissions to care facility in people with multimorbidity in clinical practice. The GDG noted that the majority of data included in the review was of low quality.

	The majority of studies evaluating tools were at a high risk of bias, due to risk of bias in sampling and poor outcome reporting. The study assessing eFI provided a greater breadth of outcome data and was moderate quality overall. All of the tools were validated in older adult populations and the studies did not report the number of people with multimorbidity in the population. The evidence was downgraded on this basis because of uncertainty that the same tools would be accurate at predicting admission to care facility in a population of older people with multimorbidity compared to the general population of people with multimorbidity. This is because some of the items in included in the tools may be more or less prevalent or important in people with multimorbidity. The GDG discussed whether the evidence from older adult populations could be generalised to a younger population of people with multimorbidity. The GDG noted that some of the risk tools included items that may be more or less prevalent in an older adult population (for example, certain diseases, frailty indicators). However, the GDG felt that the data was applicable to a younger adult population because they did not expect significant variation in the effect estimate. In particular, they noted that frailty is not an age-specific concept (that is, although fewer young people are frail, younger people with frailty are just as impaired and at risk as older people with frailty). Two of the studies were conducted in Canada in 2005. The GDG discussed the applicability of these studies to UK practice, noting that there have been changes to the thresholds for admission to care facility over the last 10 years. The GDG agreed that the studies were applicable and decided not to downgrade for indirectness. The GDG were interested in the accuracy of risk tools to predict admission to care facility within 3 years, as they felt that people with multimorbidity who experience admission to care facility within that timeframe may be in most need of an approach to care that takes account of multimorbidty. However, due to the scarcity of evidence identified for the review the GDG decided to consider evidence for tools that predicted admission to care facility for greater than 3 years. The GDG considered that tools that were identified as being accurate at predicting admission to care facility during a longer timeframe would also be accurate at identifying people with admission to care facility within 3 years, and therefore decided not to downgrade for indirectness.
Other considerations	When the evidence review for prediction of admission to a care facility was initially conducted, the GDG considered that the evidence was not sufficient to allow a recommendation to be made. Published evidence on validation of eFI became available during consultation. The GDG considered that the eFI tool was potentially useful as it predicted admission to care home as well as unplanned admissions. The GDG considered that admission to care home was an outcome particularly relevant to people with multimorbidity and would more clearly identify a population who might benefit from the type of multimorbidity approach outlined in the guideline. The availability of the tool and the fact that the tool was already known to practitioners was also taken into account by the GDG. The GDG considered that these factors were likely to increase uptake of the recommendation.

7.5 Life expectancy risk tools

7.5.1 Review question: What risk tool best identifies people with multimorbidity who are at risk of reduced life expectancy?

For full details see review protocol in Appendix C.

Table 46: PICO characteristics of review question

Question	What risk tool best identifies people with multimorbidity who are at risk of reduced life expectancy?
Population	Adults (aged >17 years) with multimorbidity Stratum: living in the community, inpatient
Risk tool	Risk tools identified in the literature validated for predicting reduced life expectancy in people with multimorbidity.
Target condition or Reference standard	All-cause mortality at 12 months
Outcomes (in terms of predictive test accuracy, calibration)	Area under the curve (c-statistic) Sensitivity, specificity, predictive values Predicted risk versus observed risk (calibration) Other outcomes for example, Somers' D statistic, R^2 statistic and Brier score Reclassification
Study types	Prospective and retrospective cohort studies (external validation, internal validation (split half validation))

The GDG was interested in identifying studies that evaluated the accuracy of risk tools for identifying individuals with multimorbidity who may have a reduced life expectancy. The GDG believed that a risk tool could be used in primary care to identify individuals who may be unlikely to benefit from some treatments, and so would help to inform decisions on withdrawing treatment in people with multimorbidity and high treatment burden. The GDG agreed that sensitivity and specificity data were of equal importance to decision-making; that is, for use in clinical practice, a risk tool should demonstrate accuracy in identifying those individuals who are at risk of reduced life expectancy as well as those who are not. This is to ensure that individuals who are at a high risk of reduced life expectancy are assessed for any additional support they may require, and to ensure that people who were not at risk of reduced life expectancy are not identified as requiring the withdrawal of treatment.

Twenty four studies evaluating 41 risk tools were included in the review,[3,18,20,25,27,39,40,43,44,59,62,117,142,144,152,177,189,192,196,207,208,211,247,248] these are summarised in Table 47below.

Eighteen risk tools were entirely weighted disease counts[25,39,142,192,196,211,243,247,248]. Twenty one risk tools assessed a variety of factors including: demographics; presence of specific diseases (for example, diabetes coronary heart disease, cancer, depression, dementia); presence of comorbidity; previous hospitalisation; functioning (for example, walking speed, cognitive function); psychological factors (for example, depressed mood, feelings of anxiety); social factors (for example, loneliness); quality of life; laboratory and clinical tests (for example, albumin, creatinine, haemoglobin); and nutritional status.

Three studies evaluated tools within a multimorbid population [20,62,248]. Twenty one studies were conducted in an indirect older adult population with an unclear number of chronic conditions. The studies included populations in the community (n=15), people who are living in a care facility or have been previously hospitalised and discharged (N = 3), and inpatient populations (n=7). For the analysis, the studies were stratified by population into inpatient and community-dwelling populations (including people living in a care facility) as the GDG thought that there would be significant variation between these 2 groups in terms of life expectancy. The number of events ranged from 14.35% to 74% in the inpatient studies, and 2.8% to 43.7% in the community-dwelling studies. The studies assessed mortality at a variety of time points ranging from 1 to 10 years.

Evidence from these studies is summarised in the clinical evidence summary below. See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Summary of included studies

Table 47: Summary of studies included in the review

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
Abbatecola 2011[3]	Hospitalised Older Patient (HOPE) index	n=1510 Older adults (aged 70 or older; mean age 81±6 years), previously hospitalised Multimorbidity: number of people with multimorbidity not reported Italy	Mortality (2 year) Sensitivity Specificity C-statistic	14.2%	Prospective cohort
Boult 1993[27]	Probability of Repeated Admission (Pra; threshold: ≥ 0.5)	n=5876 Older adults (aged 70 years or older), living in the community Multimorbidity: number of people with multimorbidity not reported USA	Mortality (4 years) Sensitivity Specificity	646 (11%)	Longitudinal cohort study
Beland 2012[18]	Geriatric Comorbidity Score (GCfbS)	n=1494 Older adults (aged 65 or over; mean 73.86 years), living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Mortality (1 year) C-statistic	Not reported	Retrospective cohort
Bernabeu-Wittel 2011A[20]	PROFUND index	n=768 Adults (aged 18 or older; mean age 78.8±9.8), inpatient (75%) Multimorbidity: 100% ('polypathological' with 2 or	Mortality (1 year) C-statistic	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
		more chronic conditions) Spain			
Boeckxstans 2015[25]	Cumulative Illness Rating Scale (CIRS)	n=567 Older adults (aged 80-101 years; mean age 84.7±3.7), living in the community Multimorbidity: 37.6% reported 5 or more diseases; range 1-16 diseases Belgium	Mortality (3 year) Sensitivity Specificity C-statistic	23.1%	Prospective cohort
Chan 2012[40]	Unnamed. 12-item scale including assessment of age; Barthel Index; number of hospitalisations in past year	n=535 Older adults (aged 86-90 years; mean age 86.5±7.4), living in care facility Multimorbidity: number of people with multimorbidity not reported Hong Kong, China	Mortality (2 year) C-statistic	31.8%	Prospective cohort
Chan 2014A[39]	Charlson Comorbidity Index (CCI)	n=2050 Older adults (mean age 80.7±7.1 years), living in the community or living in care facility Multimorbidity: number of people with multimorbidity not reported Hong Kong, China	Mortality (1 year) C-statistic	9.4%	Retrospective cohort
Clegg 2016[43,44]	Electronic Frailty Index (eFI) Unweighted deficit count (36 items)	n = 207720 (internal validation cohort), 516007 (external validation cohort) Older adults (aged 65 to 95) registered at relevant GPs Multimorbidity: number of people with multimorbidity not reported UK	Mortality (3 years) C-statistics Sensitivity Specificity	73,578 by 3yrs (10.2%)	Retrospective cohort
Daniels	Groningen Frailty	n=532	Mortality	2.8%	Prospe

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
2012[59]	Indicator Dutch Tilburg Frailty Indicator Sherbook Postal Questionnaire	 Older adults (aged 70 or older; mean 77.2±5.5), living in the community Multimorbidity: number of people with multimorbidity not reported The Netherlands	(1 year) Sensitivity Specificity C-statistic PPV NPV		ctive cohort
Diez-Manglano 2015[62]	PROFUND index	n=465 Adults (mean age 80.9±8.9), inpatient (from care facility 23.5%) Multimorbidity: 100% ('polypathological' with 2 or more chronic conditions) Spain	Morality (1 year) C-statistic	38.5%	Prospective cohort
Jones 2005[117]	CHSA Frailty Index Frailty Index-Comprehensive Geriatric Assessment (FI-CGA)	n=3736 Older adults (aged 65 or older), living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Mortality (5 year) C-statistic	Not reported	Prospective cohort
Martinez-Velilla 2014[142]	CCI CIRS, geriatric adaption (CIRS-G)	n=122 Older adults (75 years or older; mean age 85.4±5.4), inpatient (from care facility 12%) Multimorbidity: number of people with multimorbidity not reported Spain	Mortality (5 year) C-statistic Pseudo R^2	74%	Prospective cohort
Mazzaglia 2007[144]	Unnamed; 7-item questionnaire including items assessing: age; sex; hospitalisations in past 6 months;	n=2926 Older adults (aged 65 and older; mean age 75), living in the community	Mortality (15 months) C-statistic	3.9%	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	polypharmacy (≥5 prescriptions)	Multimorbidity: number of people with multimorbidity not reported Italy			
Min 2009[152]	13-Item Vulnerable Elders Survey (VES-13)	n=508 Older adults (aged 75 years or older; mean age 81.3), living in the community Multimorbidity: number of people with multimorbidity not reported USA	Mortality (4.5 years) Sensitivity Specificity C-statistic	n=222 (43.7%)	Prospective cohort
Ng 2012[177]	VES-13 VES-13, score model	n=97258 No MM reported Older adults (aged 65 or older; mean age 76.1 years), living in the community USA	Mortality (2 year) C-statistic	7.6%	Prospective cohort
Pilotto 2008[189]	Multidimensional Prognostic Index (MPI)	n=857 Older adults (aged 65-100; mean age 78.3±7.1), inpatient Multimorbidity: number of people with multimorbidity not reported Italy	Mortality (1 year) C-statistic	16.7%	Prospective cohort
Radley 2008[192]	Romano CCI	n=43811 Older adults (aged 65-99; 85% aged 75 or older), inpatient with hip fracture Multimorbidity: number of people with multimorbidity not reported USA	Mortality (1 year) C-statistic	27%	Prospective cohort
Rockwood 2005[196]	Cumulative Illness Rating Scale (CIRS)	n=2305 Older adults (aged 65 years or	Mortality (5 year)	Not reported	Prospective cohort

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	CSHA-3 Clinical Frailty Scale CSHA- 3 Frailty Index CSHA Function scale CSHA rules-based definition of frailty Modified Mini-Mental State Examination	older) , living in the community or living in a care facility Multimorbidity: number of people with multimorbidity not reported Canada	C-statistic		
Sancarlo 2011[207]	MPI	n=4412 Older adults (aged 6-100; mean age 78.1±7.1 years), inpatient Multimorbidity: number of people with multimorbidity not reported Italy	Mortality (1 year) C-statistic	19.3%	Prospective cohort
Sancarlo 2012[208]	MPI	n=654 Older adults (aged 66-99; mean age 79.34±6.5), inpatient Multimorbidity: number of people with multimorbidity not reported Italy	Mortality (1 year) C-statistic	14.35%	Prospective cohort
Schneeweiss 2001[211]	CDS-1 CDS-2 Deyo CCI D'Hoore CCI Romano CCI Ghali CCI	n=141161 Older adults (aged 65 years or older; mean age 75.4±6.7), living in the community Multimorbidity: number of people with multimorbidity not reported Canada	Mortality (1 year) C-statistic	5569 (3.95%)	Prospective cohort
Widagdo 2015[243]	Frailty phenotype	n=2087	Mortality (3 year)	Frailty phenotype	Retrospective

Study	Risk tool	Population	Outcomes	No. of events (n)	Study design
	Simplified frailty phenotype Frailty Index, 39-itrem Prognostic Frailty Score	Older adults (aged 70 years or over; mean age 77±6), majority living in the community (3.3% living in care facility) Multimorbidity: number of people with multimorbidity not reported Australia	Sensitivity Specificity C-statistic	n=205 (13.1%) Simplified frailty phenotype n=122 (10.4%) Frailty Index n=346 (16.6%) Prognostic Frailty Score n=188 (12.7%) (note- number of people included in each test was different)	cohort
Zekry 2012B[247]	CCI CIRS-G ICED Kaplan scale GIC CDS-1	n=496 Older adults (aged 75 or older), hospitalised and discharged Multimorbidity: number of people with multimorbidity not reported Switzerland	Mortality (1 year) Pseudo R^2	97 (22%)	Prospective cohort
Zeng 2014[248]	Quan CCI Quan cumulative CCI Quan baseline CCI Quan CCI trajectory: linear model	n=13163 Older adults (aged 65 or older), living in the community Multimorbidity: 100% (3 or more chronic conditions) USA	Mortality (10 year) C-statistic	4.7%	Retrospective cohort

7.5.2 Discrimination

7.5.2.1 Clinical evidence: Prognostic risk tools for identifying people with multimorbidity living in the community who are at risk of reduced life expectancy

Table 48: Clinical evidence profile: risk tools for identifying community dwelling people with multimorbidity who are at risk of reduced life expectancy

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
13-Item VES (≥2)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.92	0.25	0.75 (0.71-0.80)	LOW
13-Item VES (≥3)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.87	0.37	0.75 (0.71-0.80)	LOW
13-Item VES (≥4)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.74	0.56	0.75 (0.71-0.80)	LOW
13-Item VES (≥5)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.66	0.66	0.75 (0.71-0.80)	LOW
13-Item VES (≥6)	2	97766	HIGH[a]	No serious	Serious	No serious	0.58	0.71	0.75 (0.71-	LOW

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
				inconsistency[b]	indirectness[c]	imprecision[d]			0.80)	
13-Item VES (≥7)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.50	0.74	0.75 (0.71-0.80)	LOW
13-Item VES (≥8)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.39	0.85	0.75 (0.71-0.80)	LOW
13-Item VES (≥9)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.21	0.95	0.75 (0.71-0.80)	LOW
13-Item VES (10)	2	97766	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.09	0.98	0.75 (0.71-0.80)	LOW
13-Item VES, score model	1	97258	VERY HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.74	VERY LOW

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Charlson Comorbidity Index (CCI)	1	2050	VERY HIGH[a]	Not applicable[b]	Very serious indirectness[c]	No serious imprecision[e]	-	-	0.68 (0.64-0.72)	VERY LOW
CCI (D'Hoore, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.65	VERY LOW
CCI (D'Hoore, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.68	VERY LOW
CCI (Deyo, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.66	VERY LOW
CCI (Deyo, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.69	VERY LOW
CCI (Ghali, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.62	VERY LOW
CCI (Ghali, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.65	VERY LOW
CCI (Quan, baseline)	1	13163	HIGH[a]	Not applicable[b]	No serious indirectness[c]	Not estimable[e]	-	-	0.77	MODERATE
CCI (Quan, cumulative)	1	13163	HIGH[a]	Not applicable[b]	No serious indirectness[c]	Not estimable[e]	-	-	0.78	MODERATE
CCI (Quan, ICD-10 codes)	1	13163	HIGH[a]	Not applicable[b]	No serious	Not	-	-	0.80	MODERA

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
					indirectness[c]	estimable[e]				TE
CCI (Quan, trajectory: linear model)	1	13163	HIGH[a]	Not applicable[b]	No serious indirectness[c]	Not estimable[e]	-	-	0.77	MODERATE
CCI (Romano, hospital discharge codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.66	VERY LOW
CCI (Romano, ICD-9-CM codes)	1	141161	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.70	VERY LOW
CIRS (>3)	2	2872	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	No serious imprecision[d]	0.672	0.532	0.58 (0.56-0.67)	LOW
CSHA Function scale	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.68	LOW
CSHA rules-based definition of frailty	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.66	LOW
CSHA-3 Clinical Frailty Scale	1	2305	HIGH[a]	Not applicable	Serious indirectness[c]	Not estimable[e]	-	-	0.70	LOW
CSHA-3 Frailty Index	2	6041	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	Not estimable[e]				LOW

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
							-	-	0.69	
Dutch Tilburg Frailty Indicator (≥4)	1	532	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Serious imprecision[d]	0.67 (0.39-0.87)	0.61 (0.56-0.65)	0.64 (0.50-0.78)	LOW
eFI (Moderate and above, internal cohort)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.37	0.88	0.66	MODERATE
eFI (Moderate and above, external cohort)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.43	0.82	0.71	MODERATE
eFI (Severe, internal cohort)	1	207720	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.10	0.98	0.66	MODERATE
eFI (Severe, external cohort)	1	516007	LOW	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	0.13	0.97	0.71	MODERATE
Frailty Index (39-item) ≥0.25)	1	2087	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.344	0.858	0.60	LOW
FI-CGA	1	3736	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.67	LOW
Frailty phenotype (≥3)	1	1566	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.209	0.931	0.57	MODERATE
GCS	1	1494	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	No serious imprecision[e]	-	-	0.67 (0.57-0.70)	VERY LOW
Groningen Frailty Indicator (≥5)	1	532	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Serious imprecision[d]	0.73 (0.44-0.91)	0.54 (0.50-0.58)	0.64 (0.50-0.77)	VERY LOW
Modified Mini-Mental State Examination	1	2305	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.64	LOW

Risk tool (threshold)	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Prognostic frailty score (≥3)	1	1485	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.771	0.547	0.66	MODERATE
Sherbook Postal Questionnaire (≥2)	1	532	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.71 (0.42-0.90)	0.41 (0.37-0.46)	0.56 (0.42-0.71)	LOW
Simplified frailty phenotype (≥2)	1	1173	LOW[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[d]	0.049	0.983	0.52	MODERATE
Unnamed; 12-item questionnaire	1	535	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	No serious imprecision[e]	-	-	0.742 (0.70-0.79)	VERY LOW
Unnamed; 7-item questionnaire	1	2926	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.75 (0.73-0.78)	LOW

(a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by two increments)

(d) The judgement of precision was based on the median C-statistic value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

(e) C-statistic. The judgement of precision was based on the median C-statistic value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

Narrative results

One study[27] with 5876 participants evaluated the prognostic accuracy of the Pra tool to identify people in the community who are at high risk of mortality. This study reported the overall number of deaths during the study duration alongside the number of people predicted to die by the tool. A high probability of repeated admission was indicated by a Pra score of 5 or more. This data were used to calculate a sensitivity value of 60.5 and a specificity value of 100 for the tool. However, as it is not possible to know the true positive or negative rate (for example, how many people predicted by the tool to be admitted were actually amongst those admitted) it is not reported with the other studies in this review. This evidence is at very high risk of bias and is from an indirect older adult population. The overall quality of the study is very low.

7.5.2.2 Clinical evidence: Prognostic risk tools for identifying people with multimorbidity who are in, or recently discharged from, hospital who are at risk of reduced life expectancy

Table 49: Clinical evidence profile: risk tools for identifying people with multimorbidity who are in, or recently discharged from, hospital who are at risk of reduced life expectancy

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Burden of Illness Score for Elderly Persons (BISEP)	1	122	VERY HIGH[a]	Not applicable[b]	Serious indirectness[c]	Serious imprecision[e]	-	-	0.73 (0.63-0.82)	VERY LOW
CCI (Romano, ICD-9-CM codes)	1	43811	HIGH[a]	Not applicable[b]	Very serious indirectness[c]	Not estimable[e]	-	-	0.72	VERY LOW
CDS-1	1	141161	VERY HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.66	VERY LOW
CDS-2	1	141161	VERY HIGH[a]	Not applicable[b]	Serious indirectness[c]	Not estimable[e]	-	-	0.66	VERY LOW
Charlson Comorbidity	1	122	HIGH[a]	Not applicable[b]	Serious	Serious	-	-	0.64	VERY

Risk tool	No. of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	C-statistic Pooled/Median (range)	Quality
Index (CCI)					indirectness[c]	imprecision[e]			(0.53-0.75)	LOW
HOPE ≥4	1	3043	HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.953	0.158	0.67 (0.57-0.70)	LOW
HOPE ≥8	1	3043	LOW[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[d]	0.75	0.487	0.67 (0.57-0.70)	MODERATE
Index of Coexistent Disease (ICED)	1	122	VERY HIGH[a]	Not applicable[b]	Serious indirectness[c]	No serious imprecision[e]	-	-	0.56 (0.45-0.67)	VERY LOW
Multidimensional Prognostic Index (MPI)	3	5923	HIGH[a]	No serious inconsistency[b]	Very serious indirectness[c]	Serious imprecision[e]	-	-	0.75 (0.70-0.81)[f]	VERY LOW
PROFUND index	2	1228	HIGH[a]	No serious inconsistency[b]	Serious indirectness[c]	Serious imprecision[e]	-	-	0.70 (0.55 – 0.73)[f]	VERY LOW
Prognostic Index (PI)	1	122	HIGH[a]	Not applicable[b]	Serious indirectness[c]	Serious imprecision[e]	-	-	0.72 (0.62-0.83)	VERY LOW

a) Risk of bias was assessed using the PROBAST risk of bias checklist (see comments in Clinical Evidence tables for more details).

b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

c) C-statistic. Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by two increments)

d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences

in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

e) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

f) Median (range) AUC values

7.5.3 Calibration

7.5.3.1 Clinical evidence: Prognostic risk tools for predicting mortality in people with multimorbidity who are in hospital/recently discharged from hospital

Table 50: Clinical evidence profile: risk tools for predicting mortality in inpatient or discharged populations

Risk tool	No. of studies	n	Risk of bias	Indirectness	Imprecision	Pseudo R^2	Hosmer-Lemeshow test (p-value)	Brier score (95% CI)	D statistic	Quality
BISEP	1	122	VERY HIGH[a]	Serious indirectness[b]	Not estimable[c]	17.2	-	-	-	VERY LOW
CCI	2	1071	HIGH[a]	Serious indirectness[b]	Not estimable[c]	1.9	-	-	-	LOW
CDS-1	1	496	VERY HIGH[a]	Serious indirectness[b]	Not estimable[c]	2.0	-	-	-	VERY LOW
CIRS-G	2	618	HIGH[a]	Serious indirectness[b]	Not estimable[c]	2.4	-	-	-	LOW
GIC	2	1062	HIGH[a]	Serious indirectness[b]	Not estimable[c]	8.8	-	-	-	LOW

Risk tool	No. of studies	n	Risk of bias	Indirectness	Imprecision	Pseudo R^2	Hosmer-Lemeshow test (p-value)	Brier score (95% CI)	D statistic	Quality
ICED	2	618	VERY HIGH[a]	Serious indirectness[b]	Not estimable[c]	2.0	-	-	-	VERY LOW
Kaplan scale	1	496	HIGH[a]	Serious indirectness[b]	Not estimable[c]	4.1	-	-	-	LOW
PI	1	122	HIGH[a]	Serious indirectness[b]	Not estimable[c]	20.9	-	-	-	LOW
PROFUND index	1	768	HIGH[a]	Serious indirectness[b]	Not estimable[c]	-	0.063	-	-	LOW
Unnamed; 12-item questionnaire	1	535	HIGH[a]	Very serious indirectness[b]	Not estimable[c]	-	0.156	-	-	VERY LOW

(a) Risk of bias was assessed using the PROBAST checklist

(b) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by two increments)

(c) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

7.5.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.5.5 Evidence statements

Clinical

Community-dwelling

Twenty five tools were validated in a community-dwelling population. Of these, the GDG noted that the majority of the tools demonstrated poor discrimination and calibration for predicting reduced life expectancy in community-dwelling populations. The majority of evidence was of low to very low quality. The GDG identified 7 tools that demonstrated moderate accuracy in predicting reduced life expectancy (prioritising specificity and calibration data):

- 13-Item Vulnerable Elders Survey (VES-13): Low quality evidence from 1 study with 508 community-dwelling older adults demonstrated that the VES-13 had a specificity of 0.25-0.98 and a sensitivity of 0.09-0.92 at thresholds ranging between 2 or less to 10 and a moderate C-statistic value (0.75). Low quality evidence from studies with 97,766 community-dwelling older adults demonstrated that the VES-13 had a moderate C-statistic value (0.75-0.77). Calibration data was not reported for this tool.
- Charlson Comorbidity Index – Quan version: Low quality from 1 study with 13,163 community-dwelling older adults demonstrated that the Charlson Comorbidity Index – Quan version had a moderate C-statistic (0.77-0.8); sensitivity, specificity and calibration data was not reported for this tool.
- CHSA-3 Clinical Frailty Scale: Low quality evidence from 1 study with 2305 community or care facility dwelling older adults demonstrated that the CHSA-3 Clinical Frailty Scale had a moderate C-statistic (0.70 ; sensitivity, specificity and calibration data was not reported for this tool.
- CHSA-3 Frailty Index: Low quality evidence from 2 studies with 6041 community/care facility dwelling older adults demonstrated that the CHSA-3 Frailty Index had a moderate C-statistic (0.69-0.70); sensitivity, specificity and calibration data was not reported for this tool.
- Unnamed 12-item questionnaire: Low quality evidence from 1 study with 535 older adults living in a care facility demonstrated that an unnamed 12-item questionnaire had a moderate C-statistic (0.742); sensitivity, specificity and calibration data was not reported for this tool.
- Unnamed 7-item questionnaire: Low quality evidence from 1 study with 2926 community-dwelling older adults demonstrated that an unnamed 7-item questionnaire had a moderate C-statistic (0.75); sensitivity, specificity and calibration data was not reported for this tool.
- eFI: Moderate quality evidence from 1 study with 723727 community-dwelling older adults demonstrated that eFI had a moderate C-statistic (0.66 to 0.71 depending on cohort); sensitivity and specificity ranged from 0.10 to 0.43 and 0.82 to 0.97 depending on cohort and threshold. Pseudo R^2 ranged from 0.02 to 0.06 depending on cohort.

Inpatient

Ten tools were validated in an inpatient population. Of these, the GDG noted that the majority of the tools demonstrated poor discrimination and calibration for predicting reduced life expectancy in inpatients or people who were recently discharged from hospital. The majority of evidence was of very low quality. The GDG identified 5 tools that demonstrated moderate accuracy in predicting reduced life expectancy (prioritising specificity and calibration data):

- Burden of Illness Score for Elderly Persons (BISEP): Very low quality evidence from 1 study with 122 older adult inpatients demonstrated that BISEP had a moderate C-statistic value (0.73) and moderate calibration values (pseudo R^2 17.2%); sensitivity and specificity data was not reported for this tool.
- Charlson Comorbidity Index – Romano version: Low quality evidence from 1 study with 43,811 older adult inpatients demonstrated that Charlson Comorbidity Index – Romano version had a moderate C-statistic value (0.72); sensitivity, specificity and calibration data was not reported for this tool.
- Multidimensional Prognostic Index (MPI): Low quality evidence from 3 studies with 5923 older adult inpatients demonstrated that MPI had a moderate C-statistic (0.70-0.81); sensitivity, specificity and calibration data was not reported for this tool.
- PROFUND Index: Low quality evidence from 2 studies with 1228 older adult inpatients demonstrated that the PROFUND Index had a poor to moderate C-statistic (0.55-0.73). Low quality evidence from 1 study with 768 older adult inpatients demonstrated that PROFUND demonstrated no evidence of poor fit (Hosmer-Lemeshow test, p-value 0.063); sensitivity and specificity data was not reported for this tool.
- Prognostic Index: Low quality evidence from 1 study with 122 older adult inpatients demonstrated that the Prognostic Index had a moderate AUC (0.70-0.81) and moderate calibration values (pseudo R^2 20.9%); sensitivity and specificity data was not reported for this tool.

Economic

- No relevant economic evaluations were identified.

7.5.6 Recommendations and link to evidence

Recommendation	**No recommendations made.**
Research Recommendation	**1. Is it possible to analyse primary care data to identify characteristics that affect life expectancy and to develop algorithms and prediction tools for patients and healthcare providers to predict reduced life expectancy?**
Relative values of different outcomes	The GDG was interested in the prognostic accuracy of risk tools to identify people with multimorbidity who have a higher risk of a reduced life expectancy. The GDG wanted a tool which would be able to identify people with multimorbidity that may benefit from additional support through bespoke treatment programmes and inform decisions between patients and clinicians about the optimisation and prioritisation of treatment. The GDG did not wish to use the tool to specifically estimate the number of years a person may live. The GDG agreed that the relative value of sensitivity and specificity was dependent on how the tool was intended to inform the care of an individual patient. The GDG considered 1 of the primary functions of the tool would be to identify individuals who may benefit from the discontinuation of preventative treatment. As a consequence, they identified specificity as more critical to decision making than sensitivity, as people who are not at risk may be labelled as at risk inappropriately

	(false positives), and have treatment withdrawn unnecessarily. The GDG were aware of the potential sensitivity of introducing such a topic and using a tool to do this. The GDG also felt that a high sensitivity was important so that people who are at higher risk of mortality are not missed (fewer false negatives), therefore ensuring that these individuals gain access to treatment that may be most appropriate to a reduced life expectancy. Many of the studies included in the review reported the area under the curve (AUC). The GDG felt that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information to make a recommendation.
Trade-off between clinical benefits and harms	Evidence was identified for a number of risk prediction tools. The data were stratified between people in hospital and people living in the community (or living in a care facility). Community The GDG considered the following tools to demonstrate moderate discrimination in the community-dwelling population, as assessed using AUC: eFI; VES-13; CCI Quan; unnamed 7-item questionnaire (Mazzaglia 2007); unnamed 12-item questionnaire (Chan 2012); CHSA-3 Clinical Frailty Scale; FI. The GDG noted that sensitivity and specificity data were available for 10 tools in the community-dwelling population (VES-13, Groningen Frailty Indicator, CIRS, Dutch Tilburg Frailty Indicator, Frailty Index (39-item), Frailty Phenotype, Simplified Frailty Phenotype Pra, Prognostic Frailty Score, Sherbook Postal Questionnaire). None of these studies reported calibration, or reclassification data for these tools. The GDG agreed that in order to recommend the use of a tool in clinical practise sensitivity and specificity data must be available. Of the tools that reported sensitivity and specificity data, only the VES-13 tool and eFI tool demonstrated moderate discrimination as assessed using AUC. The GDG considered that the VES-13 tool performed best in terms of specificity at thresholds over 4 or less and the eFI tool could be used with either the moderate or severe threshold depending on whether sensitivity or specificity was the priority. Inpatient Of the tools evaluated in an inpatient population, the GDG considered the following tools to demonstrate the highest performance as assessed using AUC: MPI; PROFUND index; BISCEP; PI; CCI (Romano, ICD-9-CM codes). The GDG agreed that in order to recommend the use of a tool in clinical practise sensitivity and specificity data must be available. The GDG noted that evidence on sensitivity and specificity was only available for 1 tool in the inpatient population; evidence indicated that the HOPE tool has a good sensitivity but a poor specificity. The HOPE tool as demonstrated poor discrimination as assessing using AUC. The GDG agreed that this tool was not accurate enough to be used in current practice, due to having low specificity which may lead to patients who are not at risk may be labelled inappropriately, be referred unnecessarily or have treatment withdrawn unnecessarily. Data on calibration was only available for tools assessed in the inpatient population. Evidence from 1 study demonstrated that the PI tool explained 20.9% of the variance in life expectancy (pseudo R^2). Data from a single study demonstrated that the BISEP tool explained 17.2% of the variance in life expectancy. Results using the Hosmer–Lemeshow test demonstrated that the PROFUND index and the unnamed (Chan 2012) tool both showed good fit to the data. Summary Overall the GDG felt there was a high level of uncertainty around the evidence on prediction accuracy for the different tools. The GDG felt that no tool demonstrated both high sensitivity and specificity for predicting mortality. The GDG agreed that this was an area where additional research is required as a tool that performed well would inform future recommendations on identifying people with multimorbidity who have reduced life expectancy in order to inform decisions on optimising care.

Economic considerations	No relevant economic evaluations were identified. The GDG considered the cost associated with using risk tools to identify people who are at risk of reduced life expectancy and noted that these are generally not associated with any licencing cost although some of them may require a specific software installation with its associated costs. The main cost associated with using this risk tool is additional healthcare professional time associated with completing them. If the specificity of a tool is low, there is a risk that a large number of people will be triggered for further care that they do not require (over-treatment), which would make the tool unlikely to be not cost-effective. Conversely, if the tool has low sensitivity then a large number of people will not be identified as being at risk of reduced life expectancy, and therefore not receive the additional care they could benefit from. The GDG decided not to make a recommendation as the evidence did not identify any particular tool which was sufficiently accurate to use in clinical practice.
Quality of evidence	The GDG noted that the majority of data included in this review was of low or very low quality. In the majority of cases the studies had a high or very high risk of bias, which was often due to concerns about sampling methodology, having a reasonable number of outcome events, handling of missing data and the evaluation and reporting of relevant performance measures. In addition, the majority of studies were from an indirect older adult population. The VES-13 tool emerged as the most accurate tool identified in the review for identifying people at risk of mortality; however the GDG did not feel able to make a recommendation on the use of this tool as they were concerned that the sensitivity and specificity data for this tool was taken from only 1 study set in the USA. The number of people who scored above the each threshold was not reported; however the GDG noted that the mean score of those in the study who died (mean = 6) was much higher than optimum threshold reported (9 and above). The GDG expressed concern about the lack of availability of sensitivity and specificity data for assessing the accuracy of risk tools to predict life expectancy in a population of people with multimorbidity. The majority of studies only reported AUC values, which provide an overall estimate of the accuracy of the tool, but does not give an indication of the number of false positive and false negative diagnoses that will be made if the tool was used in practice. The GDG noted that the majority of studies were conducted with an older adult population; the GDG downgraded these studies for indirectness due to uncertainty that the same tools would be accurate at predicting mortality in a population of older people with multimorbidity compared to the a general population of people with multimorbidity. This is because some of the items included in the tools may be more or less prevalent or important in people with multimorbidity, or because of variation in people's risk of mortality. The GDG also discussed differences in life expectancy between younger and older populations. The GDG discussed whether the evidence in this review, taken from an older adult population, could be generalised to a younger multimorbid adult population, particularly where risk tools include specific factors that may be more or less prevalent in an older adult population. The GDG felt that the disease score tools were more easily generalisable to a younger population as they covered a wide range of conditions that both younger and older adults may have. However, , the GDG noted that disease score tools do not capture other factors (for example functionality) which may have an impact on life expectancy and so may not capture differences in life expectancy between a younger and an older person with the same number of conditions. The GDG felt that some of the mixed factor tools were too specific to older adults as they included factors usually more prevalent in the older adult population (for example, number of falls, walking speed, number of ADLs/IADLs, incontinence), and so would not necessarily generalise to a younger population.

	The GDG noted that all of the included studies validated the risk tools in countries outside of the UK. The GDG discussed how tools developed and validated only abroad may have limited applicability to UK practice. The GDG discussed the studies conducted in Europe, Australia and Canada, and agreed that these countries has similar health systems to the UK and so did not downgraded these studies for indirectness. In particular, the GDG was concerned about using data from the included studies that were conducted in USA and China due to differences in life expectancy and health systems. Studies that were conducted in USA and China have been downgraded for applicability.
Other considerations	The GDG were aware of initiatives such as the Gold Standards framework which aim to improve care in the last year of life. These include identifying people who may be in the last year of life and include asking the 'Surprise' question about whether a healthcare professional would be surprised if a person were to die in the near future. However, the GDG were clear that this review was not intended to identify a risk tool that was able to identify people in the last year of life. The GDG discussed the relative benefit of using a risk tool to identify people with multimorbidity with a reduced life expectancy rather than encouraging clinicians to base their judgement solely on a conversation with the person with multimorbidity about their care and using clinical judgement. The GDG questioned whether a tool could sufficiently encompass the complexity of multimorbidity to accurately quantify the risk of reduced life expectancy. The GDG noted that the relationships between the conditions of people with multimorbidity were often not fully understood and therefore the health outcomes of people with multimorbidity were difficult to predict. In addition the GDG noted that clinicians should interpret quantified estimates of life expectancy with caution, as these will be based on populations and may not be sufficiently accurate for individual patients. The GDG considered that this was an important area that needed more evaluation and they developed a research recommendation in this area. They considered that primary care data could be used to explore this area. The GDG was interested in a tool to identify people with reduced life expectancy, which they defined in terms of risk of mortality within 10 years. The GDG noted that in older adults a shorter duration may be considered as a reduced life expectancy, but less than 10 years in younger adults with multimorbidity would be considered a reduced life expectancy. Of importance they also noted that while in older people the outcome of this type of tool might be to reduce preventative medicines that the person might not benefit from; the actions in younger multimorbid people might differ in that they might benefit from the addition of preventative medicines or other treatments if their true state of health was understood. More details on this research recommendation can be found in Appendix O.

7.6 Polypharmacy: unplanned hospital admissions, health related quality of life, mortality and admission to care facilities

The evidence reviews in this chapter were planned to allow the GDG to make recommendations for tools or other measures that would identify people with multimorbidity at highest risk of adverse outcomes. One of the areas of interest that was identified at scoping stage was polypharmacy. The priority was to identify prognostic studies but where these were not available the GDG agreed the inclusion of studies that reported on associations of polypharmacy with adverse outcomes. Polypharmacy in used in this context as simply the number of medicines a person is taking. In the context of people with multimorbidity who are on multiple medicines and who perceive a treatment burden, optimisation of medicines may involve the removal of medicines that are considered necessary therapeutic interventions in guidelines and according to medical experts so additional

descriptors such as appropriate polypharmacy are not used. The purpose of the review was to explore whether the number of medicines was a simple but useful way of identifying people who were at risk of poor outcomes.

7.7 Review question: Is polypharmacy associated with a greater risk of unplanned hospital admissions amongst people with multimorbidity?

For full details see review protocol in Appendix C.

Table 51: Characteristics of review question

Population	Adults (aged >17 years) with multimorbidity
Prognostic variable/s under consideration	Polypharmacy
Outcomes	Unplanned admissions at ≥ 1 year Statistical outputs may include: Sensitivity, specificity, AUC, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
Study design	Prognostic studies

7.7.1 Clinical evidence

We searched for prospective cohort studies investigating the association of polypharmacy with unplanned hospital admissions in people with multimorbidity, in order to identify if polypharmacy could be used to identify people with multimorbidity who are at high risk of adverse outcomes. Only papers published after year 2000 were included in this review. This is because the GDG believed that in general, the number of medications patients receive has increased over time and the relationship between polypharmacy and adverse outcomes may be different in older papers. We prioritised studies that evaluated the prognostic accuracy of polypharmacy for predicting unplanned hospital admissions (that is, discrimination and calibration data); however, no studies were identified. As a consequence, the GDG chose to also search for evidence that evaluated whether the risk of unplanned hospital admission increased with polypharmacy. As the GDG wished to use polypharmacy as an isolated identifier of people with multimorbidity who are at risk for adverse outcomes, the GDG chose to only include studies that reported the risk of unplanned admission at increased levels of polypharmacy where this was unadjusted for other factors (including known confounders, such as number of conditions and illness severity).

Two studies[190,222] were included in the review; these are summarised in Table 52 below. Evidence from these studies is summarised in the clinical evidence profiles. See also the study selection flow chart in Appendix C, forest plots in Appendix K, Grade tables in appendix J, study evidence tables in Appendix H and exclusion list in Appendix L.

One of the included studies[190] was conducted with an older adult population living in the community, and the other was conducted with an older adult population living in a care facility. Neither study provided information on the prevalence of the sample who were multimorbid. Both studies compared the risk of unplanned hospital admission at specific thresholds of polypharmacy compared to no polypharmacy as defined by less than 5 drugs.

Table 52: Summary of studies included in the review

Study	Population	Analysis	Prognostic variable(s)	Outcomes
Pozzi 2010[190]	n=788 Italy Older adults (aged 65 years or over; mean age 73±6.8) Living in the community Number of events = 634 (80.5%)	Cox proportional hazard regression model	Polypharmacy (≥5 drugs) vs. no polypharmacy (<5 drugs)	Hospitalisation (4-8 years)
Spector 2013[222]	n=62 745 USA Older adults (aged 65 years or over; 46% over 85) Living in care facility Number of events = not stated	Fine and Grey competing risks proportional hazards regression	Polypharmacy (5-9 drugs) vs. no polypharmacy (<5 drugs) Polypharmacy (10-14 drugs) vs. no polypharmacy (<5 drugs) Polypharmacy (≥15 drugs) vs. no polypharmacy (<5 drugs)	Ambulatory care sensitive hospitalisation (3 – 25 months) Nursing home sensitive hospitalisation (3 – 25 months) Unavoidable hospitalisation (3 – 25 months)

Community data

7.7.2 Prognostic accuracy data

No relevant data identified.

7.7.3 Unadjusted data

Table 53: Clinical evidence profile: risk of hospitalisation at various thresholds of polypharmacy

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (≥5 drugs) vs. no polypharmacy (<5 drugs) for predicting hospitalisation (unadjusted HR) [older adults, living in the community]	1	788	LOW[a]	Not applicable	Serious indirectness[b]	Serious imprecision[c]	Unadjusted HR [95% CI]: 1.00 [0.78 – 1.28]	LOW

(a) Risk of bias was assessed using QUIPS for more details, please see the 'comments' section in the clinical evidence tables
(b) Downgraded once as the majority of the evidence included an indirect population
(c) Downgraded once as the 95% CI crosses the null line

Living in a care facility

7.7.4 Prognostic accuracy data

No relevant data identified.

7.7.5 Unadjusted data

Table 54: Clinical evidence profile: risk of hospitalisation at various thresholds of polypharmacy

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (5-9 drugs) vs. no polypharmacy (<5 drugs) for predicting ambulatory care sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	6165	LOW[a]	Not applicable	Very serious indirectness[b]	Serious imprecision[c]	Sub-hazard RR [95% CI]: 1.10 (0.96 – 1.25)	VERY LOW
Polypharmacy (5-9 drugs) vs. no polypharmacy (<5 drugs) for predicting nursing home sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	7595	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.19 (1.07 – 1.33)	LOW
Polypharmacy (5-9 drugs) vs. no polypharmacy (<5 drugs) for predicting 'unavoidable' hospitalisation (sub-hazard RR) [older	1	9320	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.21 (1.09 – 1.33)	LOW

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
adults, living in care facility]								
Polypharmacy (10-14 drugs) vs. no polypharmacy (<5 drugs) for predicting ambulatory care sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	6165	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.24 (1.09 – 1.42)	LOW
Polypharmacy (10-14 drugs) vs. no polypharmacy (<5 drugs) for nursing home sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	7595	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.33 (1.19 – 1.49)	LOW
Polypharmacy (10-14 drugs) vs. no polypharmacy (<5 drugs) for 'unavoidable' hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	9320	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.39 (1.25 – 1.54)	LOW

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (≥15 drugs) vs. no polypharmacy (<5 drugs) for predicting ambulatory care sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	6165	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.41 (1.22 – 1.63)	LOW
Polypharmacy (≥15 drugs) vs. no polypharmacy (<5 drugs) for predicting nursing home sensitive hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	7595	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.42 (1.26 – 1.61)	LOW
Polypharmacy (≥15 drugs) vs. no polypharmacy (<5 drugs) for predicting 'unavoidable' hospitalisation (sub-hazard RR) [older adults, living in care facility]	1	9320	LOW[a]	Not applicable	Very serious indirectness[b]	No serious imprecision	Sub-hazard RR [95% CI]: 1.38 (1.23 – 1.54)	LOW

(a) Risk of bias was assessed using QUIPS for more details, please see the 'comments' section in the clinical evidence tables
*(b) Downgraded twice as the majority of the evidence included an indirect population and the outcome included unplanned admissions within 1 year of baseline Downgraded once as the 95% CI crosses the null line*Click here to enter text.

7.7.6 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F

7.7.7 Evidence statements

Clinical

- Across the 2 studies included in the review, the evidence suggested that polypharmacy was associated with an increased risk of unplanned hospital admission, without adjusting for confounding variables. This association was observed for people taking 5 or more drugs, with greater risk of unplanned hospital admission at increased levels of polypharmacy. In particular, the GDG noted that the risk of unplanned hospital admission was particularly high in people taking 15 or more drugs. The evidence ranged from low to very low quality due to serious indirectness and serious imprecision for lower levels of polypharmacy.

Economic

- No relevant economic evaluations were identified.

7.7.8 Recommendations and link to evidence

See – Recommendations and link to evidence in polypharmacy review on admission to care facilities section 7.10.7.

7.8 Polypharmacy: health-related quality of life

7.8.1 Review question: Is polypharmacy associated with a greater risk of reductions in health-related quality of life amongst people with multimorbidity?

For full details see review protocol in Appendix C.

Table 55: Characteristics of review question

Population	Adults (aged >17 years) with multimorbidity
Prognostic variable/s under consideration	Polypharmacy
Outcomes	Health-related quality of life at ≥ 1 year Statistical outputs may include: Sensitivity, specificity, AUC, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
Study design	Prognostic studies

7.8.2 Clinical evidence

No relevant clinical studies investigating the prognostic accuracy of polypharmacy for predicting reductions in health-related quality of life were identified.

7.8.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.8.4 Evidence statements

Clinical

No clinical evidence was identified.

Economic

- No relevant economic evaluations were identified.

7.8.5 Recommendations and link to evidence

See Recommendations and link to evidence in polypharmacy: admission to care facilities review 7.10.7.

7.9 Polypharmacy: mortality

7.9.1 Review question: Is polypharmacy associated with a greater risk of mortality amongst people with multimorbidity?

For full details see review protocol in Appendix C.

Table 56: Characteristics of review question

Population	Adults (aged >17 years) with multimorbidity
Prognostic variable/s under consideration	Polypharmacy
Outcomes	Mortality at ≥ 1 year Statistical outputs may include: Sensitivity, specificity, AUC, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
Study design	Prognostic studies

7.9.2 Clinical evidence

We searched for prognostic studies investigating the association of polypharmacy with mortality in people with multimorbidity, in order to identify if polypharmacy could be used to identify people

with multimorbidity who are at high risk of adverse outcomes. Ten studies were included in the review[4,77,94,96,119,127,146,190,194,241] these are summarised below.

We prioritised studies that evaluated the prognostic accuracy of polypharmacy for predicting mortality (that is, discrimination and calibration data). However, only 1 study[94], which evaluated the prognostic accuracy of polypharmacy (threshold 5 or less drugs) for predicting mortality in people with multimorbidity, reported this data. No studies were identified that evaluated the prognostic accuracy of polypharmacy for predicting mortality in a non-multimorbid population. As a consequence, the GDG chose to also search for evidence that evaluated whether the risk of mortality increased with polypharmacy. As the GDG wished to use polypharmacy as an isolated identifier of people with multimorbidity who are at risk for adverse outcomes, the GDG chose to only include studies that reported the risk of mortality at increased levels of polypharmacy where this was unadjusted for other factors (including known confounders, such as number of conditions and illness severity).

Several studies included in the review compared the risk of mortality at specific thresholds of polypharmacy compared to no polypharmacy as defined by less than 5 drugs. A range of thresholds were used in the studies; 5 or more medications (n=4); 6 or less medications (n=1); 6-9 medications (n=1); 10 or more medications (n=1). Additionally, some studies assessed polypharmacy as a continuous risk factor for mortality (n=6). One of these studies[127] assessed polypharmacy as a continuous predictor as assessed using the number of drug classes, as opposed to the number of medications. This study was analysed separately as the number of drug classes is a less accurate measure of polypharmacy; for example, this does not distinguish between a person who takes multiple drugs within a single drug class and someone who takes only 1 drug within the same class. To aid interpretation of the continuous risk factor data alongside the threshold data, as well as being presented as risk differences per single additional drug, the continuous risk factor data was extrapolated to estimate the risk difference per 5 additional drugs. This data is presented in the footnotes of the clinical summary tables (Table 60 and Table 61).

All of the evidence is taken from studies conducted with an older adult population where the level of multimorbidity in the sample was unknown. All studies were conducted with a majority outpatient population.

Evidence from these are summarised in the clinical evidence profile below (Tables 58 – 61). See also the study selection flow chart in Appendix C, forest plots in Appendix K, Grade tables in appendix J, study evidence tables in Appendix H and exclusion list in Appendix L.

Table 57: Summary of studies included in the review

Study	Population	Analysis	Prognostic variable(s)	Outcomes
Ahmad 2005[4]	n=1042 England Older adults (aged 65 years or over; mean 75.21) Living in the community Number of events = 741 (71%)	Cox regression function of CoRGA	Number of drugs (continuous)	Mortality (15 year)
Espino 2006[77]	n=3050 USA	Cox proportional hazards regression models	Polypharmacy (≥ 5 drugs) vs. no	Mortality (8 years)

Study	Population	Analysis	Prognostic variable(s)	Outcomes
	Older adults (aged 65-99) Living in the community Number of events = 950 (30.8%)		polypharmacy (<5 drugs)	
Gnjidic 2012[94]	n=1705 Australia Older males (aged 70 years or over) Living in the community Number of events n=305 (17.9%)	Risk prediction model	Polypharmacy (≥5 drugs) vs. no polypharmacy (<5 drugs) Number of drugs (continuous)	Mortality (6 years)
Gomez 2015[96]	n=5052 Spain Older adults (aged 65 or older, mean ages ranging from 72.7 to 75.4 years between groups, SDs of 6.7 to 6.9) Number of events = 334 (6.6%)	Cox proportional hazards model	Polypharmacy (≥6 drugs) vs no medication (0 drugs) Number of drugs (continuous)	Mortality (median follow-up 6.5 years)
Jyrkka 2009[119]	n=601 Finland Older adults (aged 75 years or older) Living in the community (86%) or living in care facility (14%) Number of events = 358 (59.6%)	Cox proportional hazards model	Polypharmacy (6-9 drugs) vs. no polypharmacy (<5 drugs) Polypharmacy (≥ 10 drugs) vs. no polypharmacy (<5 drugs)	Mortality (4 years)
Krause 2007[127]	n=5888 USA Older adults (aged 65 years or over) Living in the community Number of events = not	Cox proportional hazards regression	Number of drug classes (continuous)	Mortality (8 years)

Study	Population	Analysis	Prognostic variable(s)	Outcomes
	stated			
Md Yusof 2010[146]	n=113 England Older adults (aged 64 years or over) Living in the community Number of events = 20 (17.7%)	Cox regression method	Number of drugs (continuous)	Mortality (7 years)
Pozzi 2010[190]	n=788 Italy Older adults (aged 65 years or over; mean age 73±6.8) Living in the community Number of events = 271 (34.4%)	Cox proportional hazard regression model	Polypharmacy (≥5 drugs) vs. no polypharmacy (<5 drugs)	Mortality (4-8 years)
Richardson 2011[194]	n=12423 England and Wales Older adults (aged 65 years or over; 10% aged over 85) Living in the community (96%) or living in care facility (4%) Number of events = 9225 (75%)	Cox proportional hazard regression model	Polypharmacy (≥ 5 drugs) vs. no polypharmacy (<5 drugs)	Mortality (18 years)
Wang 2015[241]	n=1562 China Older adults (aged 80 years or over; mean age 85.2, range 80-104) Living in the community Number of events not reported	Logistic regression	Number of drugs (continuous)	Mortality (5 years)

7.9.3 Prognostic accuracy data

Table 58: Clinical evidence profile: prognostic accuracy of polypharmacy for predicting mortality

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity	Specificity	Area Under Curve Pooled/Median (range)	R2	Quality
Polypharmacy (≥ 5 drugs) vs. no polypharmacy (<5 drugs) for predicting mortality	1	1705	HIGH[a]	Not applicable	Serious indirectness[b]	Not estimable	0.51	0.65	0.61	-	LOW

(a) Risk of bias was assessed using the PROBAST checklist
(b) Downgraded once as the majority of the evidence included an indirect population

7.9.4 Unadjusted data

Table 59: Clinical evidence profile: risk of mortality at various thresholds of polypharmacy

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (≥ 5 drugs) vs. no polypharmacy (<5 drugs) for predicting mortality (unadjusted HR)	2	15473	LOW[a]	No serious inconsistency	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 1.87 [1.77 - 1.98]	MODERATE

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (6-9 drugs) vs. no polypharmacy (<5 drugs) for predicting mortality (unadjusted HR)	1	601	LOW[a]	Not applicable	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 1.50 [1.14 - 1.98]	MODERATE
Polypharmacy (≥10 drugs) vs. no polypharmacy (<5 drugs) for predicting mortality (unadjusted HR)	1	601	LOW[a]	Not applicable	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 2.87 [2.20 - 3.74]	MODERATE
Polypharmacy (≥6 drugs) vs no medication (0 drugs) for predicting mortality (unadjusted HR)	1	5052	LOW[a]	Not applicable	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 2.78 [2.36 – 3.27]	MODERATE

(a) Risk of bias was assessed using QUIPS for more details, please see the 'comments' section in the clinical evidence tables
(b) The majority of the evidence included an indirect population

Table 60: Clinical evidence profile: risk of mortality with increasing polypharmacy (polypharmacy as a continuous predictor)

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Number of drugs for predicting mortality	2	6094	LOW[a]	No serious inconsistency	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 1.16 [1.14 – 1.18][c]	MODERATE

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
(unadjusted HR)								
Number of drugs for predicting mortality (unadjusted OR)	2	3267	LOW[a]	No serious inconsistency	Serious indirectness[b]	No serious imprecision	Unadjusted OR [95% CI]: 1.16 [1.13 – 1.20][d]	MODERATE
Number of drugs for predicting mortality (unadjusted OR)	1	113	LOW[a]	Not applicable	Serious indirectness[b]	Not estimable	Unadjusted OR: 1.26 [not reported][e,f]	MODERATE

(a) Risk of bias was assessed using QUIPS for more details, please see the 'comments' section in the clinical evidence tables
(b) Downgraded once as the majority of the evidence included an indirect population
(c) Based on a mean difference of 5 drugs between those with and without polypharmacy HR = 2.10, 95% CI = 1.92 – 2.29
(d) Based on a mean difference of 5 drugs between those with and without polypharmacy OR = 2.10, 95% CI = 1.84 – 2.49
(e) Based on a mean difference of 5 drugs between those with and without polypharmacy; OR = 3.18
(f) OR calculated by Exp(β coefficient)

Table 61: Clinical evidence profile: risk of mortality with increasing polypharmacy (polypharmacy as a continuous predictor as assessed using number of drug classes)

Risk factor	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Number of drug classes for predicting mortality (unadjusted HR)	1	5888	VERY HIGH[b]	Not applicable	Serious indirectness[b]	No serious imprecision	Unadjusted HR [95% CI]: 1.19 [1.15 – 1.22][c]	VERY LOW

(a) Risk of bias was assessed using QUIPS; downgraded twice as the majority of evidence was at high risk of bias for more details, please see the 'comments' section in the clinical evidence tables
(b) Downgraded once as the majority of the evidence included an indirect population
(a) Based on a mean difference of 5 drugs between those with and without polypharmacy; HR = 2.39, 95% CI = 2.01 – 2.70

7.9.5 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.9.6 Evidence statements

Clinical

- Across the 10 studies included the review, the evidence suggested that polypharmacy was associated with an increased risk of mortality without adjusting for confounding variables. This association was observed for people taking 5 or more drugs, with greater risk of mortality at increased levels of polypharmacy. In particular, the GDG noted that the risk of mortality was high in people taking 15 or more drugs. The evidence ranged from moderate to very low quality due to risk of bias and serious indirectness.

Economic

- No relevant economic evaluations were identified.

7.9.7 Recommendations and link to evidence

See Polypharmacy – Admission to Care facility review for the Recommendations and link to evidence in section 7.10.7.

7.10 Polypharmacy: admission to care facilities

7.10.1 Review question: Is polypharmacy associated with a greater risk of admission to care facility amongst people with multimorbidity?

For full details see review protocol in Appendix C.

Table 62: Characteristics of review question

Population	Adults (aged >17 years) with multimorbidity
Prognostic variable/s under consideration	Polypharmacy
Outcomes	Admission to care facility at ≥ 1 year Statistical outputs may include: Sensitivity, specificity, AUC, R^2, beta coefficients, OR/RR, HR, MD will be extracted if no sensitivity/specificity data
Study design	Prognostic studies

7.10.2 Clinical evidence

We searched for studies investigating the association of polypharmacy with admission to care facility in people with multimorbidity, in order to identify if polypharmacy could be used to identify people

with multimorbidity who are at high risk of adverse outcomes. One study was included in the review[249], this is summarised below.

We prioritised studies that evaluated the prognostic accuracy of polypharmacy for predicting admission to care facility (that is, discrimination and calibration data). However, no studies were identified that evaluated this were identified. As a consequence, the GDG chose to also search for evidence that evaluated whether the risk of admission to care facility increased with polypharmacy. As the GDG wish to use polypharmacy as an isolated identifier of people with multimorbidity who are at risk for adverse outcomes, the GDG chose to only include studies that reported the risk of admission to care facility at increased levels of polypharmacy where this was unadjusted for other factors (including known confounders, such as number of conditions and illness severity).

The included study assessed the risk of admission to care facility with excessive polypharmacy (less than or equal to 13 drugs) compared to no polypharmacy (0 drugs) amongst older adults living in the community.

Evidence is summarised in the clinical evidence profile below. See also the study selection flow chart in Appendix E, forest plots in Appendix K, Grade tables in appendix J, study evidence tables in Appendix H and exclusion list in Appendix L.

Table 63: Summary of studies in the review

Study	Population	Analysis	Prognostic variable(s)	Outcomes	Limitations/comments
Zuckerman 2006[249]	n= 487 383 USA Older adults (aged 65 years or over; 7.9% over 85) Living in the community Number of events = 22 042 (4.5%)	Continuous time proportional hazards model for interval-censored data	Polypharmacy (≥13 drugs) vs. no polypharmacy (0 drugs)	Admission to care facility (3 years)	*comparison is excessive polypharmacy vs. no medication

Community dwelling

7.10.3 Prognostic accuracy data

No relevant data identified.

7.10.4 Unadjusted data

Table 64: Clinical evidence profile: risk of mortality at various thresholds of polypharmacy

Risk factor	Number of studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Pooled effect with 95% CIs [if meta-analysed] OR Effect and CI in single study	Quality
Polypharmacy (≥13 drugs) for predicting admission to care facility (unadjusted RR) [older adults, community dwelling]	1	487 383	LOW[a]	Not applicable	Serious indirectness[b]	No serious imprecision	Unadjusted RR [95% CI]: 3.31 [3.16 – 3.46]	MODERATE

(a) Risk of bias was assessed using QUIPS for more details, please see the 'comments' section in the clinical evidence tables
(b) The majority of the evidence included an indirect population.

7.10.5 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

7.10.6 Evidence statements

Clinical

- One study comprising 487 383 people demonstrated that polypharmacy (≥13 drugs) is associated with an increased risk admission to care facility compared to no polypharmacy (<13 drugs) with no adjustment for confounders. This evidence was of moderate quality due to serious indirectness.

Economic

- No relevant economic evaluations were identified.

7.10.7 Recommendations and link to evidence

Recommendations	**11. Use an approach to care that takes account of multimorbidity for adults of any age who are prescribed 15 or more regular medicines, because they are likely to be at higher risk of adverse events and drug interactions.** **12. Consider an approach to care that takes account of multimorbidity for adults of any age who:** • **are prescribed 10 to 14 regular medicines** • **are prescribed fewer than 10 regular medicines but are at particular risk of adverse events.**
Relative values of different outcomes	The GDG considered mortality, unplanned admissions, admission to care facility and health-related quality of life at 1 year or less to be critical outcomes in evaluating whether polypharmacy could be used to identify people with multimorbidity who were at risk of adverse outcomes in primary care. Prognostic accuracy data (for example, sensitivity, specificity and AUC) were identified as the best evidence available for this review, as this is the only data that can tell you how accurate polypharmacy is at identifying the people you want; including how many at risk people you will miss (false negatives) and how many will be identified unnecessarily (false positives). This data can therefore give you a sense of the clinical implications of recommending a specific threshold of polypharmacy in practice. R^2, beta coefficients, OR/RR, HR and MD data were also extracted. Only

	unadjusted data was extracted (that is, data that reported the association between polypharmacy and the outcome where this was not adjusted for other covariates) because the GDG was interested in recommending the number of drugs as a marker to identify people at risk of adverse events, and were not concerned with establishing a causal relationship between polypharmacy and adverse outcomes.
Trade-off between clinical benefits and harms	<u>Unplanned admissions</u> No prognostic accuracy data were identified. Unadjusted data demonstrated that older adults living in a care facility with polypharmacy (both those who were taking 10 to 14 drugs or 15 or more drugs) were at increased risk of hospitalisation compared to those without polypharmacy (taking fewer than 5 drugs). <u>Health-related quality of life</u> No data was identified. <u>Mortality</u> Prognostic accuracy data (sensitivity, specificity and AUC) were available for 1 study of older adults living in the community, which showed that polypharmacy, defined as having recent prescription of 5 or more drugs, performed poorly as an indicator of increased risk of morality compared to those taking 5 or fewer drugs. In this 1 study, presence of polypharmacy was ascertained by trained personnel interviewing subjects and taking a medication inventory. Only regular prescription medication was included. Unadjusted data (risk ratios, odds ratios and hazard ratios) demonstrated that people with polypharmacy living in the community (both those who are taking 5 or more drugs or 10 or more drugs) were at greater risk of mortality compared to those without polypharmacy (taking less than 5 drugs). <u>Admission to a care facility</u> No prognostic accuracy data was identified. Unadjusted data demonstrated that older adults with polypharmacy living in the community (taking 13 or more drugs) were at increased risk of admission to care facility compared to those taking no drugs. <u>Threshold of polypharmacy that indicates a high risk of adverse events</u> The GDG noted that there was a continuous gradient between the increasing number of drugs and an increased risk of adverse events. The GDG recognised the need for a threshold of polypharmacy to operationalise its use as a tool in clinical practice, although the GDG expressed concern about the lack of prognostic accuracy data (sensitivity and specificity) to indicate the clinical implications of recommending a particular threshold. The GDG considered the use of 5, 10 and 15 drugs as cut-offs for polypharmacy, as these thresholds were used in several studies and the GDG were aware that all 3 are commonly used to define polypharmacy. The GDG noted that a threshold of 5 or more drugs was too low as a marker of people who are at risk of adverse outcomes, as they were aware that many people taking 5 or more drugs do not need additional care. The GDG noted that some people taking 5 or more drugs will be at risk of adverse outcomes but these will be influenced by additional factors (for example, the type of drugs they are taking and the severity of their conditions). Therefore the GDG agreed that healthcare professionals should use their clinical judgement and consider the wider context of the person at lower thresholds The GDG noted that of the evidence demonstrated that people who are taking 10 or more drugs are at higher risk of adverse events than people who are taking 5 or more drugs The GDG noted that the evidence demonstrated

	that people prescribed 15 or more drugs may be at significantly higher risk of unplanned hospital admissions and agreed via consensus that they may also be at increased risk of mortality. On this basis the GDG agreed that people prescribed 15 or more drugs would benefit from a multimorbidity approach to care and this can be considered on the basis of the number of drugs alone, independent of other risk factors. The approach should be consider for people prescribed 10-14 medicines, and also in those prescribed less than 10 particularly when other factors are taken into consideration.
Economic considerations	No economic evidence was found on prognostic factors. The prognostic reviews aimed at identifying individuals at higher risk of mortality, unplanned hospital admissions, admission to care facilities and poor health-related quality of life, as a multimorbidity approach to care may be beneficial in this population Based on the polypharmacy prognostic reviews and GDG consensus, people taking 15 or more drugs were judged to be at higher risk of unplanned admission and mortality, while no additional care was considered necessary for people taking less than 10 drugs. For the group between 10 and 14 drugs, the GDG decided that clinical judgement would be required. No additional costs are expected to be associated with a multimorbidity approach to care itself. Currently patients have their medications reviewed every year or more often; this recommendation aims at changing the content of these discussion rather than changing the quantity or intensity of the reviews. The GDG agreed that the majority of these conversations would take place within usual consultation time with no associated increased in cost of GP's time. In addition, a multimorbidity approach to care aims at reducing treatment burden, adverse events, and unplanned or uncoordinated care, which would create some cost savings to the NHS.
Quality of evidence	<u>Unplanned admissions</u> The quality of the evidence was low to very low. All of the studies were indirect as they were conducted in an older adult population. Furthermore the evidence for people living within a care facility was also indirect as the outcome included unplanned admissions within 1 year of baseline. Additionally, some of the studies demonstrated a serious imprecision. <u>Mortality</u> The quality of the evidence was moderate to very low. All of the studies were indirect as they were conducted in an older adult population. Additionally some of the studies were at a serious risk of bias. <u>Admission to care facility</u> The quality of the evidence was moderate. The evidence was indirect as the study was conducted in an older adult population. The GDG noted that evidence from one study that compared taking 13 drugs was compared to taking no drugs at all. The GDG noted that the use of this comparison may exaggerate the risk associated with polypharmacy, as people taking no drugs at all may be a much healthier population. <u>General points</u> The GDG noted that prognostic accuracy data was only available for 1 of the outcomes (mortality), and the bulk of the evidence included in the review demonstrated the raw association between polypharmacy and adverse outcomes. This data provides an indication of the risk of adverse outcomes with increasing levels of polypharmacy , however, it is not able to demonstrate the number of false positive and false negative 'diagnoses' you may get by recommending a particular threshold of polypharmacy for use in clinical practice. The GDG discussed whether the evidence in this review taken from an older adult population could be generalised to a population of

	younger adults with multimorbidity, particularly as taking a larger number of drugs is more prevalent in an older adult population. The GDG felt that although age may affect the magnitude of risk associated with polypharmacy, in general the direction of effect would be maintained.
Other considerations	The GDG discussed how polypharmacy should be assessed. The method of assessing the number of medicines taken and the definition of a medicine varied across the 12 studies included in this review. Four studies only counted prescription medicines, 6 studies counted both prescription and non-prescription medicines and the remaining 2 studies did not specify. One study only counted regular medicine, 3 studies counted regular and PRN medicine and the remainder did not specify. Six studies gathered data from interviews or questionnaires, 2 studies checked patient records and the remainder used a combination of methods. Overall, the GDG agreed that there was no evidence that any 1 method gave a more reliable prediction of risk or a produced a larger estimate of risk. However, the GDG noted that regular prescription data would generally be the most available to healthcare professionals, and therefore may be easier to use in clinical practice. The GDG recognised that primary care records in particular provided a means to easily quantify the number of medicines a person was taking and flag those who may benefit from a multimorbidity approach to care. The GDG discussed whether polypharmacy assessments should include creams and ointments. The GDG agreed that it should, as the use of these can be onerous for people. The GDG noted the lack of risk prediction data available for using polypharmacy as a predictor of adverse clinical outcomes and the consequence that it is not possible to estimate the likelihood of false negatives or false positives. The GDG noted that the impact of false negatives is likely to be lessened by the concurrent use of other tools to identify people who would benefit from an individualised approach, as per the other recommendations in this guideline. The GDG considered that even if an individual was not at risk of an adverse outcome, there would still be some benefit from at least a medication review for those people on more than 15 drugs. By using unadjusted data in this review, polypharmacy is being assessed in isolation. The GDG felt this was appropriate for the purposes of this review. One of the advantages of using the numbers of medicines is the ease of assessing this. The GDG also considered that it has face validity as a potential signifier of concern as many healthcare professionals have experience of practical problems in managing large number of drugs. There is no evidence for the cumulative effect of many medicines, and interactions and adverse events are likely to be greater the larger number of drugs people are taking. The GDG noted that the relationship between polypharmacy and mortality may vary considerably between different populations and settings, where the other factors associated with mortality (for example, age, number of conditions and illness severity) differ from this review population. Therefore, clinicians should consider other risks alongside polypharmacy and be aware of the setting in which they are assessing polypharmacy. The GDG had some concern that there is a perception that multimorbidity is a problem of older people. Younger people with multimorbidity may therefore not be identified as needing a more holistic approach. The GDG therefore included the tern 'people of any age' when developing these recommendations to emphasise this point. The GDG noted that over the counter medications that people are taking may be missed in consultations or in record checks. Healthcare professionals may want to ask specifically about this category of medications.

8 Frailty

8.1 Introduction

Increased attention has been drawn to the concept of frailty in recent years. Frailty can be considered a condition characterised by reduced biological reserves which puts an individual at risk when facing minor stressors. A minor stress puts a frail person at risk of falls and fluctuating disability which may increase care needs, hospital admission and care home admission. Identifying someone as frail may be a useful way of identifying those people with multimorbidity who would particularly benefit from optimising medicines and treatments.

8.2 Review question: What is the most accurate tool for assessing frailty?

For full details see review protocol in Appendix C.

Table 65: Characteristics of review question

Population	Adults (aged 18 years and over) with multimorbidity
Target condition	Frailty
Index test	Tools and brief assessments identified in the literature for assessing frailty; including: • Abbreviated Comprehensive Geriatric Assessment (aCGA) • Vulnerable Elders-Survery-13 (VES-13) • Groningen Frailty Indicator (GFI) • Geriatric 8 (G8) • Tilburg frailty indicator • PRISMA 7 • Timed up and go test (TUG) • Edmonton frail scale • Brief assessments (for example, gait speed, grip strength)
Reference standard	• Cardiovascular health study (CHS) phenotype model • Comprehensive geriatric assessment (CGA) • Cumulative deficit model
Statistical measures Outcomes	Sensitivity, specificity, AUC
Study design	Diagnostic test accuracy studies Stratify by reference standard If heterogeneity is found, the following subgroup analyses will be conducted: Population (<65 years versus > 65 years)
Other exclusions	Children and young people with multiple morbidity (aged <18 years) Adults with more than 1 mental health condition and no physical condition Adults with cancer

We sought studies that examined the diagnostic accuracy of tools or brief assessments to diagnose frailty. The GDG identified 3 reference standards that are accepted in current practice. A brief explanation of these reference standards is provided below (page 191). Although all 3 are perceived as the gold standard for the assessment of frailty, evidence in the wider literature indicates that the 3 are only moderately correlated, and may therefore represent different definitions of frailty. As a

consequence, the GDG decided to stratify by the reference standard used by studies. The GDG identified both sensitivity and specificity of the tests as being critical to decision-making for this review. Sensitivity is important as not treating individuals with frailty may mean that vulnerable people do not receive the additional support they need, and may miss out on assessment to optimise their care (including withdrawing individuals from treatments in order to reduce burden). However, if specificity is too low, then people who receive a false positive diagnosis may experience harm if treatments are withdrawn, and if underlying conditions are undiagnosed due to symptoms being attributed to frailty. Furthermore, some people may be unhappy with being 'labelled' as frail, and there are also significant cost implications of treating individuals unnecessarily. The GDG therefore sought tests with both high sensitivity and specificity.

8.3 Clinical evidence

Thirteen studies (reported in fourteen publications) were included in the review[12,33,38,58,60,61,112,182,191,210,213,219,234,235].

Evidence was identified for each of the references standards, and several studies used more than 1 reference standard. In summary; 2 studies used the Cumulative deficit model as a reference standard, 1 study used the Comprehensive Geriatric Assessment (CGA) as a reference standard, and 13 studies used Fried's phenotype as a reference standard. The majority of evidence was conducted with older adults (that is, adults 65 or more years of age). No studies included in the review clearly identified the study sample as being multimorbid. Two papers reported that the sample was not multimorbid. Prevalence rates of frailty varied widely between studies.

Evidence from the study using Cumulative deficit model as a reference standard to diagnose frailty is summarised in section 8.3.2. Evidence from studies using CGA as a reference standard to diagnose frailty is summarised in section 8.3.3. Evidence from studies using Fried as a reference standard to diagnose frailty is summarised in section 8.3.4. See also the study selection flow chart in Appendix F, and sensitivity and specificity forest plots in Appendix K, study evidence tables in Appendix H and exclusion list in Appendix L.

8.3.1 Reference standards

Cumulative deficit model: There are a number of different scales assessing frailty according to the Cumulative deficit model frailty definition. In these scales, people are assigned a score representing the ratio of deficits they have from a broad list of deficits. For example, in 1 of the studies included in this review, participants were assigned a score between 0-1 based on a 45-item list of deficits taken from the CGA. This list includes presence of health diagnoses (for example, vascular, respiratory and kidney problems), neurological and psychological difficulties, functional difficulties (for example, requiring help with personal care, housework, finance) and social factors (for example, loneliness). These were assessed using a combination of interviews with a geriatrician and geriatric nurse, and scale such as mini-mental state exam (MMSE). Frailty was defined as impairment in 25% or more of deficits.

Comprehensive Geriatric Assessment: This is a multidimensional assessment, treatment plan and regular review delivered by a MDT that usually includes doctors, nurses, physiotherapists, occupational therapists and social workers. The format and content of the CGA will vary between community and inpatient settings. In the study included in this review, the CGA assessed 5 domains; functional status, cognition, depression, nutritional status, and medication use. Frailty was defined as impairment in 2 or more of these domains.

Fried's phenotype: The majority of evidence in this review uses Fried's phenotype model as the reference standard for diagnosing frailty. This approach defines frailty as impairment in 3 or more of the following physical domains; nutritional status, exhaustion, weakness, low physical activity, walking speed. Across the studies included in this review, different tests were used to assess each of the components, with the thresholds chosen to indicate impairment varying according to the population under study.

Table 66: Summary of studies included in the review

Study	Population	Index tool	Reference standard	Comments
Auyeung 2014[12]	Older adults (mean age = 72 years) Community living Prevalence of frailty = 5.7% (males) and 5.2% (females) Unclear if multimorbid population N = 4000	• BMI ≤18.5 • Physical activity as assessed with the Physical activity scale for the elderly (PASE) • Grip strength • Walking speed • Self-reported exhaustion (yes/no)	Fried's phenotype	Hong Kong
Boxer 2008a[33]	Older adults with congestive heart failure (mean age = 77 years) Outpatient Prevalence of frailty = 27% Unclear if multimorbid population N = 60	6-minute walking test	Fried's phenotype	USA
Castell 2013[38]	Older adults (mean age = 75.4 years) Community living Prevalence of frailty = 11.2% Not a multimorbid population (33.8% of participants were diagnosed with ≥2 comorbid conditions) N = 1327	Walking speed	Fried's phenotype	Spain
Da Camara 2013[58]	Older adults (mean age 69.5 years)	Short physical performance battery, including	Fried's phenotype	Reports data from 2 cohorts (Canada and

Study	Population	Index tool	Reference standard	Comments
	Community living Prevalence of frailty = 19.4%. There was a higher prevalence of frailty in the high deprivation cohort (Brazil; 28.1%) than in the low deprivation cohort (Canada; 10%). Unclear if multimorbid population N = 124	gait, balance, and chair stand		Brazil). Data varied between cohorts, and is presented for each site individually.
Dent 2012[60]	Older adults (mean age = 85.2 years) Inpatients admitted with acute illness Prevalence of frailty = 66% Unclear if multimorbid population N = 100	Mini-nutritional assessment – short form	Fried's phenotype	Australia
Dibari 2014[61]	Older adults ≥70 years Community living Prevalence of frailty = 36.6% Unclear if multimorbid population N = 1037	Postal questionnaire	Fried's phenotype	Italy
Hoogendijk 2013[112]	Older adults (mean age = 78.6 years) Primary care Prevalence of frailty = 11.6% Possibly a multimorbid population (mean number of chronic conditions = 2.9; ±1.9)	• Clinical judgement of the GP (yes/no) • Self-rating • Polypharmacy • GFI • PRISMA7	Fried's phenotype	Netherlands

Study	Population	Index tool	Reference standard	Comments
	N = 102			
Nunes 2015[182]	Older adults (mean age = 85.7 years) Community living Prevalence of frailty = 37% Significant minority of the population not multimorbid (63.5% of sample with ≥2 conditions) N = 433	Self-report questionnaire	Fried's phenotype	Brazil
Purser 2006[191]	Older adults with significant coronary artery disease (mean age = 77 years) Inpatient Prevalence of frailty = 27% (Fried); 63% (Cumulative deficit model) Likely to be a multimorbid population (mean number of comorbidities = 3.8 ±1.6); 80.3% hypertension, 75.4% hyperlipidemia 36.6% diabetes 29.4% congestive heart failure 41.7% myocardial infarction 24.9% depression 29.4% congestive heart failure 16.8% COPD 19.1% cerebrovascular disease N = 309	• Grip strength • Gait speed • 30-second chair stand	• Fried's phenotype • Cumulative deficit model	USA
Savva 2013[210]	Older adults (median age = 70 years) Community living Prevalence of frailty = 4.5%; prevalence of frailty	Timed up and go test (TUG)	Fried's phenotype	Ireland

Study	Population	Index tool	Reference standard	Comments
	Unclear if multimorbid population N = 1814			
Schoon 2014[213]	Older adults (mean age = 76.8 years) Community living Prevalence of frailty = 10% (Fried); prevalence of frailty according to Cumulative deficit model not reported Unclear if multimorbid population N = 593	• Gait speed • Maximum step length • Chair lift	• Fried's phenotype • Cumulative deficit model	Netherlands
Smets 2014[219]	Older adults (median age = 78 years) Community living Prevalence of frailty = 48% Unclear if multimorbid population N = 290	• Abbreviated comprehensive geriatric assessment (aCGA) • VES-13 • GFI	CGA	Netherlands
Tribess 2012[235]; Tribess 2013[234]	Older adults (mean age = 71.1 years) Community living Prevalence of frailty = 20% Unclear if multimorbid population (95.3% of participants had ≥1 diseases) N = 624	• Age • Physical activity (International physical activity questionnaire; IPAQ)	Fried's phenotype	Brazil

8.3.2Tests for identifying frailty (Cumulative deficit model as reference standard)

Table 67: Clinical evidence profile: diagnostic test accuracy for chair stand (time taken to rise 5 times from chair without arms)

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Chair lift	2	849	Serious [a]	Serious [b]	None [c]	None [d]	0.57 - 0.76 (CI = 0.71 – 0.80)[e]	LOW

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

(e) CI only reported for highest AUC

Table 68: Clinical evidence profile: diagnostic test accuracy for gait speed

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Gait speed	2	827	Serious[a]	None[b]	None[c]	Serious[d]	0.70 - 0.81 (CI = 0.76 – 0.85)[e]	LOW

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the

threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

(e) CI only reported for highest AUC

Table 69: Clinical evidence profile: diagnostic test accuracy for grip strength

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Grip strength	1	309	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.66	MODERATE

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

Table 70: Clinical evidence profile: diagnostic test accuracy for maximum step length

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Step length	1	547	Serious[a]	Not applicable	None[c]	Serious[d]	0.77 (CI = 0.72 – 0.81)	LOW

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was

downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

8.3.3Tests for identifying frailty (CGA as reference standard)

Table 71: Clinical evidence profile: diagnostic test accuracy for abbreviated CGA (aCGA)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at impairment in 1 domain	1	290	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.87	0.64	-	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 72: Clinical evidence profile: diagnostic test accuracy for Groningen Frailty Indicator (GFI)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥4	1	290	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.74	0.73	-	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 73: Clinical evidence profile: diagnostic test accuracy for Geriatric 8 (G8)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤14	1	290	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.75	0.69	-	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).
(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).
(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.
(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 74: Clinical evidence profile: diagnostic test accuracy for Vulnerable elders survey 13 (VES-13)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥3	1	290	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.82	0.79	-	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making
(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).
(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).
(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.
(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

8.3.4Tests for identifying frailty (Fried's phenotype as reference standard)

Table 75: Clinical evidence profile: diagnostic test accuracy for age

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥67 years (males)	1	218	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.32	0.977	0.59	MODERATE
Index threshold at ≥72 years (females)	1	406	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.814	0.844	0.72	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 76: Clinical evidence profile: diagnostic test accuracy for polypharmacy

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥5 medications	1	102	None[a]	Not applicable	None[c]	Not estimable[d]	0.70	0.73	0.71	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 77: Clinical evidence profile: diagnostic test accuracy for clinical judgement (GP rating; yes/no)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Clinical judgement	1	102	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.70	0.77	0.73	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 78: Clinical evidence profile: diagnostic test accuracy for self-report ('how would you rate your health status on a scale from 0 to 10?')

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Self-report	1	102	None[a]	Not applicable	None[c]	Not estimable[d]	0.85	0.73	0.79	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 79: Clinical evidence profile: diagnostic test accuracy for self-report questionnaire (Dibari 2014)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Self-report (postal questionnaire) threshold at ≥4 frailty items and no disability	1	1037	None[a]	Not applicable	None[c]	Not estimable[d]	0.93	0.27	0.695	HIGH
Self-report (postal questionnaire) threshold at ≥5 frailty items and no disability	1	1037	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.71	0.58	0.695	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 80: Clinical evidence profile: diagnostic test accuracy for self-report questionnaire (Nunes 2015)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Self-report (Nunes 2015) threshold at deficit of ≥3 domains	1	433	Very Serious [a]	Not applicable	Serious [c]	Not estimable [d]	0.632	0.716	-	VERY LOW

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 81: Clinical evidence profile: diagnostic test accuracy for self-reported exhaustion (yes/no)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Self-reported exhaustion (males)	1	2000	None[a]	Not applicable	None[c]	Not estimable[d]	0.385	0.955	0.670	HIGH

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Self-reported exhaustion (females)	1	2000	None[a]	Not applicable	None[c]	Not estimable[d]	0.283	0.951	0.617	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 82: Clinical evidence profile: diagnostic test accuracy for BMI

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥18.5 (males)	1	2000	None[a]	Not applicable	None[c]	Not estimable[d]	0.317	0.957	0.637	HIGH
Index test threshold at ≥18.5 (females)	1	2000	None[a]	Not applicable	None[c]	Not estimable[d]	0.222	0.959	0.591	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 83: Clinical evidence profile: diagnostic test accuracy for Mini-Nutritional Assessment

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤7	1	100	None[a]	Not applicable	None[c]	Not estimable[d]	0.636	0.794	0.802	HIGH
Index test threshold at ≤8	1	100	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.803	0.765	0.802	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 84: Clinical evidence profile: diagnostic test accuracy for 30-second chair stand

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤7 stands	1	309	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.79	0.79	0.802	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 85: Clinical evidence profile: diagnostic test accuracy for chair stand (time taken to complete 5 chair stands; SPPB scale 0-4)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤2 (Brazil)	1	64	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.81	0.58	0.64 (CI 0.48 – 0.81)	LOW
Index test threshold at ≤2 (Canada)	1	60	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.92	0.70	0.82 (CI 0.72 – 0.93)	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables)The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 86: Clinical evidence profile: diagnostic test accuracy for chair stand (time taken to complete 5 chair stand without arms)

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Chair stand	1	540	Very serious[a]	Not applicable	None [c]	Not estimable [d]	0.81 (CI = 0.75 – 0.88)	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

Table 87: Clinical evidence profile: diagnostic test accuracy for gait speed

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at 0.65 m/s	1	309	Very serious[a]	Not applicable	None[c]	Not estimable [d]	0.82	0.82	0.89	LOW
Index test threshold at 0.76 – 0.78 m/s	2	2518	Very serious[a]	None[b]	None[c]	None[d]	0.91	0.80	0.90 (CI 0.87 – 0.96)	LOW
Index test threshold at 0.8 m/s	2	1843	Very serious[a]	None[b]	None[c]	Serious[d]	0.85 -0.99	0.64 -0.91	0.92 (CI 0.87 – 0.96)	VERY LOW

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test threshold at 0.89 – 0.9 m/s	2	2518	Very serious[a]	None[b]	None[c]	Serious[d]	0.61 -0. 827	0.83 – 0.96	0.83 - 0.92 (CI 0.87 – 0.96)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 88: Clinical evidence profile: diagnostic test accuracy for gait speed (SPPB scale 0-4)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤4 (Brazil)	1	64	Very serious[a]	Not applicable	None[c]	Serious[d]	0.54	0.47	0.58 (CI 0.42 – 0.75)	VERY LOW

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test threshold at ≤4 (Canada)	1	60	Very serious[a]	Not applicable	None[c]	Serious[d]	0.84	0.75	0.69 (CI 0.56 – 0.83)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 89: Clinical evidence profile: diagnostic test accuracy for walking speed (distance achieved in 6-minutes)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≤300m	1	60	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.94	0.75	-	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 90: Clinical evidence profile: diagnostic test accuracy for maximum step length

Index test	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Area under the curve, median (range)	Quality
Step length	1	547	Very serious[a]	Not applicable	None[c]	Serious[d]	0.84 (CI = 0.77 – 0.90)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(e) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(f) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(g) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(h) The judgement of precision was based on the median AUC value and its 95% CI. The evidence was downgraded by 1 increment if the CI varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0–0.5, 0.5-0.8 and 0.8-1).

Table 91: Clinical evidence profile: diagnostic test accuracy for the Timed up and go test (TUG)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
TUG threshold at >8	1	1814	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.97	0.18	0.87	MODERATE
TUG threshold at >9	1	1814	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.95	0.42	0.87	MODERATE
TUG threshold at >10	1	1814	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.93	0.62	0.87	MODERATE
TUG threshold at >11	1	1814	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.80	0.78	0.87	MODERATE
TUG threshold at >12	1	1814	Serious[a]	Not applicable	None[c]	Not estimable[d]	0.72	0.86	0.87	MODERATE

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 92: Clinical evidence profile: diagnostic test accuracy for grip strength

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at 18kg	1	2000	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.845	0.819	0.844	LOW
Index test threshold at 25kg	1	308	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.72	0.72	0.83	LOW
Index test threshold at 28kg	1	2000	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.895	0.806	0.862	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 93: Clinical evidence profile: diagnostic test accuracy for Physical activity (IPAQ)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at 140 minutes/week (males)	1	218	Very serious[a]	Not applicable	None[c]	None[d]	0.731	0.977	0.90 (CI = 0.86 – 0.94)	LOW
Index test threshold at 145 minutes/week (females)	1	406	Very serious[a]	Not applicable	None[c]	None[d]	0.814	0.844	0.86 (0.85 – 0.92)	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 94: Clinical evidence profile: diagnostic test accuracy for Physical activity (PASE; 0-361)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test threshold at ≤56.4 (males)	1	2000	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.837	0.835	0.849	LOW
Index test threshold at ≤58.8 (females)	1	2000	Very serious[a]	Not applicable	None[c]	Not estimable[d]	0.828	0.847	0.857	LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 95: Clinical evidence profile: diagnostic test accuracy for Physical activity (SPPB; 0-12)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test threshold at <9 (Brazil)	1	64	Very serious[a]	Not applicable	None[c]	Very serious[d]	0.81	0.52	0.67 (CI = 0.49 – 0.84)	VERY LOW
Index test threshold at <9 (Canada)	1	60	Very serious[a]	Not applicable	None[c]	Serious[d]	0.92	0.80	0.81 (CI = 0.70 – 0.92)	VERY LOW

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 96: Clinical evidence profile: diagnostic test accuracy for Groningen Frailty Indicator (GFI)

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥4	1	102	None[a]	Not applicable	None[c]	Not estimable[d]	0.57	0.72	0.64	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

Table 97: Clinical evidence profile: diagnostic test accuracy for PRISMA-7

Index test (Threshold)	Number of studies	n	Risk of bias	Inconsistency	Indirectness	Imprecision	Sensitivity (median/ range/ 95% CI)	Specificity (median/ range/ 95% CI)	Area under the curve, median (range)	Quality
Index test										
Index test threshold at ≥3	1	102	None[a]	Not applicable	None[c]	Not estimable[d]	0.86	0.83	0.85	HIGH

The assessment of the evidence quality was conducted with emphasis on test sensitivity/specificity as this was the primary measure discussed in decision-making

(a) Risk of bias was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables).

(b) Inconsistency was assessed by inspection of the sensitivity and specificity forest plots (based on the primary measure), or summary area under the curve (sROC) plots across studies, using the point estimates and confidence intervals. Particular attention was placed on values above or below 0.5 (diagnosis based on chance alone) and the threshold set by the GDG (the threshold above which would be acceptable to recommend a test) – for example, the GDG might set a threshold of 0.8 an acceptable level to recommend a test. The evidence was downgraded by 1 increment if the individual studies varied across 2 areas (for example, 0.5-0.8 and 0.8-1) and by 2 increments if the individual studies varied across 3 areas (for example, 0-0.5, 0.5-0.8 and 0.8-1).

(c) Indirectness was assessed using the QUADAS-2 checklist (please see the 'General limitations' section in the clinical evidence tables) items referring to applicability.

(d) The judgement of precision was based on visual inspection of the confidence region in the diagnostic meta-analysis if a diagnostic meta-analysis was conducted. Where a diagnostic meta-analysis was not conducted imprecision was assessed according to the range of point estimates. As a rule of thumb (after discussion with the GDG) a range of 0–0.2 of differences in point estimates of sensitivity was considered not imprecise, 0.2-0.4 serious imprecisions, and >0.4 very serious imprecision. Imprecision was assessed on the primary measure for decision-making

8.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

8.5 Evidence statements

Clinical

A large number of tools to diagnose frailty were identified and included in the review. Each tool was evaluated in a small number of studies (1 or 2 studies only, at each threshold). These tools varied widely in their complexity; some tools were relatively simple (for example, clinician judgement or self-reported exhaustion) and others were more complex (distance walked in 6-minutes, assessments of physical activity within the previous weeks). The evidence indicated that very simple tests demonstrated high sensitivity and specificity. No inconsistency in the findings was identified across different reference standards, although the vast majority of the evidence compared the accuracy of tools against the reference standard of the Fried's phenotype model.

- Walking (gait) speed was evaluated in 6 studies, comprising of a total of 6413 people. This evidence was not pooled due to differences between the studies in the method of assessing walking speed and the threshold used to indicate frailty. This evidence suggested that walking speed has high sensitivity and high specificity for identifying frailty, although accuracy may be lower if using a threshold consistent with a faster walking speed (≥0.8 – 0.9 metres per second). This evidence was of low or very low quality.
- Self-reported health status ('how would you rate your health status on a scale from 0 to 10?') was evaluated in 1 study, comprising 102 people. This evidence demonstrated that self-reported health status had a sensitivity of 0.85 and specificity of 0.73 (AUC 0.79). This evidence was of high quality.
- PRISMA-7 was evaluated in 1 study, comprising 102 people. This evidence demonstrated that PRISMA-7 had a sensitivity of 0.86 and specificity of 0.83 (AUC 0.85). This evidence was of high quality.
- The timed up and go test (TUG) was evaluated in 1 study, comprising 1814 people. This evidence demonstrated that the TUG had a high sensitivity (0.72 – 0.97) for thresholds between >8 and >12, and high specificity at thresholds of >11 and >12 (0.78 – 0.86; AUC 0.87). This evidence was of moderate quality.
- The physical activity scale for the elderly (PASE) was evaluated in 1 study comprising 4000 people. This evidence demonstrated that the PASE had a high sensitivity (0.84 in males and 0.83 in females) and high specificity (0.84 in males and 0.85 in females) at a threshold of ≤56.4 in males and ≤58.8 in females (AUC 0.85 in makes and 0.86 in females). This evidence was of low quality.
- The mini-nutritional assessment short form (MNA) was evaluated in 1 study comprising 100 people. This evidence demonstrated that the MNA-short form had a high sensitivity (0.80) and high specificity (0.77) at a threshold of ≤8 (AUC 0.80). This evidence was of moderate quality.

Economic

- No relevant economic evaluations were identified.

8.5.1 Recommendations and link to evidence

Recommendations	**13. Consider assessing frailty in people with multimorbidity.** **14. Be cautious about assessing frailty in a person who is acutely unwell.** **15. Do not use a physical performance tool to assess frailty in a person who is acutely unwell.** **<u>Primary care and community care settings</u>** **16. When assessing frailty in primary and community care settings, consider using 1 of the following:** • **an informal assessment of gait speed (for example, time taken to answer the door, time taken to walk from the waiting room)** • **self-reported health status (that is, 'how would you rate your health status on a scale from 0 to 10?', with scores of 6 or less indicating frailty)** • **a formal assessment of gait speed, with more than 5 seconds to walk 4 metres indicating frailty** • **the PRISMA-7 questionnaire, with scores of 3 and above indicating frailty.** **<u>Hospital outpatient settings</u>** **17. When assessing frailty in hospital outpatient settings, consider using 1 of the following:** • **self-reported health status (that is, 'how would you rate your health status on a scale from 0 to 10?', with scores of 6 or less indicating frailty)** • **the 'Timed Up and Go' test, with times of more than 12 seconds indicating frailty** • **a formal assessment of gait speed, with more than 5 seconds to walk 4 metres indicating frailty** • **the PRISMA-7 questionnaire, with scores of 3 and above indicating frailty** • **self-reported physical activity, with frailty indicated by scores of 56 or less for men and 59 or less for women using the Physical Activity Scale for the Elderly.**
Relative values of different outcomes	The aim of this review was to identify tools that could be used in clinical practice to diagnose frailty. The GDG identified both sensitivity and specificity as important outcomes in the evaluation of tools to diagnose frailty in adults with multimorbidity, and therefore prioritised those tools that report both high sensitivity and specificity. Sensitivity is important as not treating individuals with frailty may mean that vulnerable people do not receive the additional support they need, and they may miss out on assessments to optimise their care (including withdrawing individuals from treatments in order to reduce burden). Specificity is also important as people

	who receive a false positive diagnosis may experience harm if treatments are withdrawn, and if underlying conditions are undiagnosed due to symptoms being attributed to frailty. Furthermore, some people may be unhappy with being 'labelled' as frail, and there are also significant cost implications of treating individuals unnecessarily. Many of the studies included in the review reported the area under the curve (AUC). The GDG felt that this metric was important for comparing the overall accuracy of the tools, but in itself was unlikely to provide enough information to make a recommendation.
Trade-off between clinical benefits and harms	The GDG did not identify any 1 of the 3 reference standards as being more relevant to adults with multimorbidity, and so considered the evidence for index tests compared to any of the 3 reference standards. CSHA as reference standard The evidence indicated that gait speed and step length both demonstrated moderate accuracy as assessed using AUC, the chair lift demonstrated very poor to moderate accuracy as assessed using AUC, and grip strength demonstrated poor accuracy as assessed using AUC for diagnosing frailty. CGA as reference standard When compared to this reference standard, the VES-13 demonstrated the highest balance between sensitivity and specificity. This evidence indicated that this test will miss 18% of true cases and will falsely diagnose 21% of non-frail people as frail. The abbreviated CGA demonstrated slightly higher sensitivity (indicating that this will miss 13% of true cases), but reduced specificity (will falsely diagnose 36% of non-frail individuals as frail). The GFI and G8 both demonstrated moderate sensitivity and specificity, but resulted in a greater number of false negative and false positive diagnoses than the other tests. Phenotype model as reference standard The GDG identified the following index tests as having moderate or high accuracy for diagnosing frailty compared to the reference standard: polypharmacy (sens 70, spec 73, AUC .71); clinical judgement (sens 70, spec 77, AUC .73); self-report (how would you rate your health on a scale of 1-10?; sens 85, spec 73, AUC 79); mini-nutritional assessment at ≤8 (sens 80, spec 76.5, AUC 80); 30-second chair stand (sens 79, spec 79, AUC .80); 5 chair stand (AUC .81); gait speed for thresholds between 0.65 – 0.8 m/s (sens range 82 – 91.9; spec range 76 – 91; AUC range .89 - .92); walking distance with threshold at ≤ 300m (sens 94, spec 75); step length (AUC 0.84); timed up and go test at a threshold of >11 (sens 80, spec 78) and >12 (sens 72, spec 86); grip strength at a threshold of <18kg for women and <28kg for men (sens 84.5, spec 81.9, AUC 0.84 and sens 89.5, spec 80.6, AUC 0.86 respectively); physical activity (IPAQ) with a threshold of <140 minutes/week for men and <145 minutes/week for women (sens 73.1, spec 97.7, AUC 0.90 and sens 81.4, spec 84.4, AUC 0.86 respectively); physical activity (PASE) with a threshold of ≤56.4 for men and ≤58.8 for women (sens 83.7, spec 83.5, AUC 0.85 and sens 82.8, spec 84.7, AUC 0.857 and sens 82.8, spec 84.7, AUC 0.857); and PRISMA-7 (sens 86, spec 83, AUC 0.85). Summary Overall, the GDG agreed that the evidence indicated that assessments of physical activity, in addition to a person's self-reported health, were useful for diagnosing frailty in older adults. The GDG felt that it was important that an index test to diagnose frailty in clinical practice should be quick and feasible to conduct in clinical practice. They discussed how some tests, such as the timed get up and go test, may be feasible within specialist clinics where appointments are usually longer. The GDG therefore chose to consider both the accuracy and the feasibility of the index test when making their recommendation. The GDG chose to recommend a number of index tests, so clinicians are able to choose a test that is easiest for a particular patient or setting. In primary care and community settings, the GDG decided to recommend gait speed, self-reported health status, and PRISMA-7. These tools were identified as being accurate as well as relatively easy to conduct during a routine

	healthcare appointment. The GDG chose to recommend an informal assessment of gait speed in addition to a formal assessment. This is because the evidence demonstrated a consistent relationship between slow walking speed and higher risk of frailty, and the GDG felt that slow walking speed may be identified by clinicians easily and quickly if done informally (for example, when walking from the waiting room into the consultation room) as well as through formalised walking tests. In hospital outpatient settings, the GDG felt that there was more time for more elaborate tests of frailty, and so in addition to these tests, the GDG also decided to recommend clinicians consider using the timed get up and go test and reported physical activity (self-reported or using PASE). The GDG discussed whether the recommendation should include the mini-nutritional assessment (short form) as a tool to identify frailty. This tool demonstrated high accuracy for diagnosing frailty relative to the reference standard (the phenotype model). However, the GDG opted not to include this tool in the final recommendation as there were other tools of similar accuracy and these were available without cost, and would therefore were considered to be more cost-effective.
Economic considerations	No relevant economic evaluations were identified. The GDG considered the clinical evidence for the different tools and highlighted the importance of selecting frailty tools that are easy, quick and cheap to apply as well as effective. None of the tools considered requires any additional equipment, although the GDG was mindful that, in more time constrained primary care settings, tools had to be particularly quick and easy to apply to be usable. This area was deemed to have no major economic implications as no significant costs are expected to be associated with healthcare professionals assessing frailty; this would be usually in the format of assessing gait speed which would require a negligible time. However assessing frailty could generate health benefits as it would help pick up people at risk of adverse effects from medication and prevent those effects and their associated costs and health burden.
Quality of evidence	The GDG noted that the reference standards may each assess different definitions of frailty, and that greater research is needed to identify which definition and reference standard is most appropriate for adults with multimorbidity. The GDG noted that the vast majority of data in the review were taken from single studies, and so there is limited data to support any of the index tests. Furthermore, many of the studies used a threshold that has not been validated in another sample (for example, they selected a threshold representing the lowest quintile of the study sample), and therefore it is unclear whether the accuracy data will remain the same when applied in wider practice. The evidence included in the review was all with a population of older adults, and no evidence was found evaluating the diagnostic accuracy of tools to diagnose frailty in younger adults with multimorbidity. The GDG also noted that the vast majority of evidence included in this review was conducted with people living in the community. The GDG agreed that performance-based tools, such as walking speed, would not be appropriate for use in people with multimorbidity who are acutely unwell, such as those in hospital. One study included in the review was conducted with an inpatient population, which demonstrated similar accuracy of walking speed in identifying frailty as studies conducted with a population of people living in the community. However, the GDG noted that the sample in this study may be more mobile than many people in hospital, and therefore did not think that this study could be generalised to the wider inpatient population. Due to the lack of evidence with an inpatient population, the GDG agreed that they were not able to make a recommendation on the use of a tool to identify frailty in an inpatient population. CSHA The evidence for gait speed and step length was of low quality due to risk of bias and imprecision; the evidence for chair lift was of low quality due to risk of bias and inconsistency; and the evidence for grip strength was of moderate quality due to risk

	of bias. CGA The evidence for VES-13 was of moderate quality due to risk of bias; the evidence for the abbreviated CGA was of low quality due to risk of bias; and the evidence for the GFI and G8 was of moderate quality due to risk of bias. Phenotype model The evidence included in the review had differing quality ratings which are listed below: • Polypharmacy, self-report (how would you rate your health on a scale of 1-10?) and PRISMA-7 were high quality. • clinical judgement, mini-nutritional assessment at ≤8 , 30-second chair stand and timed up and go test at a threshold of >11 and >12 all were moderate quality with remaining tools low or very low quality, that is, 5 chair stand, gait speed for thresholds between 0.65 – 0.8 m/s walking distance with threshold at ≤ 300mstep length; grip strength at a threshold of <18kg for women and <28kg for men; physical activity (IPAQ) with a threshold of <140 minutes/week for men and <145 minutes/week for women; physical activity (PASE) with a threshold of ≤56.4 for men and ≤58.8 for women
Other considerations	The GDG felt that it was important that GPs are aware of frailty and consider assessing this in people with multimorbidity. Rather than provide a formal diagnosis, for which the GDG felt would require a reference standard assessment, the GDG felt that identifying people in primary care who may be vulnerable using 1 of these tests may be useful for informing decisions on care. For example, it may inform whether to refer a person for a more formal holistic assessment of their needs or identify them as requiring particular attention to burden of treatment The GDG agreed that the evidence indicated that simple tools could provide useful indications of frailty in primary care and that more formal assessment tools would be more appropriate in specialist settings. The GDG also agreed that many of these tools could be applied easily in a relatively informal way in primary and community care. How long someone takes to walk from a general practice waiting room to a clinical room, or how long someone takes to open their door when being visited at home are aspects of functioning that are readily assessed during normal encounters, and already regularly recognised and commented on by healthcare professionals. The GDG considered that practitioners could do formal testing but should also be empowered to record and act on common clinical observations about gait speed. The GDG agreed that it was particularly inappropriate to use tests of frailty which assess physical performance when people are acutely unwell as frailty may be conflated with effects of acute illness. The GDG discussed whether the data from the review could be generalised to a younger population of adults with multimorbidity. The GDG believed that some frailty tools may be less accurate at identifying frailty in younger adults with multimorbidity, due the lower prevalence rates of frailty in this population. However, the GDG felt that younger adults with multimorbidity who exhibit reductions in walking speed or functioning may require additional assessment or support, and therefore decided that this recommendation should apply to adults with multimorbidity of all ages.

9 Delivering an approach to care that takes account of multimorbidity

9.1 Introduction

The development of the guideline included a number of evidence reviews which examined specific interventions such as holistic assessments of care, collaborative care and self-management. These are discussed in later chapters. There is a lack of evidence to support such interventions but the GDG considered that evidence from other reviews (such as those reviews examining barriers to optimising care) provided useful insights to the difficulties faced by people with multimorbidity (see section 6.3). That evidence prompted them to outline explicit steps that would allow an individualised approach to care and this chapter outlines the steps involved in that approach in more detail. These include

9.2 Approach to the patient; 9.3 Treatment burden; 9.4 Establishing patient preferences, values and priorities; 9.5 Effectiveness of interventions from condition specific guidance.

9.2 Approach to the patient

9.2.1 Recommendations and link to evidence

Recommendations	**18. Follow the recommendations in the NICE guideline on patient experience in adult NHS services which provides guidance on knowing the patient as an individual, tailoring healthcare services for each patient, continuity of care and relationships, and enabling patients to actively participate in their care.** **<u>Discussing the purpose of an approach to care that takes account of multimorbidity</u>** **19. Discuss with the person the purpose of the approach to care, that is, to improve quality of life. This might include reducing treatment burden and optimising care and support by identifying:** • **ways of maximising benefit from existing treatments** • **treatments that could be stopped because of limited benefit** • **treatments and follow-up arrangements with a high burden** • **medicines with a higher risk of adverse events (for example, falls, gastrointestinal bleeding, acute kidney injury)** • **non-pharmacological treatments as possible alternatives to some medicines** • **alternative arrangements for follow-up to coordinate or optimise the number of appointments.**
Relative values of different outcomes	The GDG were interested in ensuring that people with multimorbidity who might benefit from an approach to care that takes account of multimorbidity were fully involved in the process and understood its aims.
Trade-off between clinical benefits and harms	The GDG considered that people could only be fully involved if they understood the process of a multimorbidity approach to care. Central to achieving this is treating the patient as outlined in the NICE patient experience guideline. The GDG considered

	that were unlikely to be harms if people were treated in this way.
Economic considerations	The GDG considered the potential resource costs of these recommendations. In general the recommendations emphasise the importance of recommendations from generic NICE guidelines around patient involvement and experience, for this specific population. The GDG noted that this may result in additional time required in consultations but that this will vary significantly between people and it is difficult to quantify. The GDG considered that the delivery of a multimorbidity approach to care could be carried out as part of usual medical practice when providing and reviewing care for people with multimorbidity.
Quality of evidence	These recommendations were informed by other NICE guidance and evidence reviews in chapter 6 and this chapter where quality of evidence contributing to the recommendations is discussed.
Other considerations	The GDG considered that the recommendations in NICE guideline on patient experience in adult NHS services outlined important areas for the care of all adults but particularly for the care of people with multimorbidity. The GDG considered it important to emphasise cross-referral to that guideline. People with multimorbidity will have unique combination of conditions and characteristics and knowing the patient as an individual, tailoring healthcare services for each patient, establishing continuity of care and relationships, and enabling patients to actively participate in their care as outlined in that guideline are important for this group. As discussed in section 6.2 the experience of multimorbidity can be one of confusion for people with a lack of clear direction and co-ordination of their care. The GDG considered that healthcare practitioners therefore needed to be explicit and clear if they were offering people a multimorbidity approach to care. This should explain to the person that the purpose of an approach to care that takes account of multimorbidity is to find ways of reducing treatment burden and optimising care and what this might involve such as identifying treatments that could be stopped because of limited benefit, medicines with a higher risk of adverse events and changes to follow up and co-ordination of care. The GDG considered that such a conversation while clear and explicit needs to be done sensitively to ensure people understand that the aim is to improve quality of life and not save costs.

9.3 Treatment Burden

9.3.1 Review question: How can treatment burden be assessed?

For full details see review protocol in Appendix C.

Table 98: PICO characteristics of review question

Population	Adults (18 years and over) with multimorbidity
Intervention	Questionnaires identified in the literature that aim to assess people's experience on treatment burden
Statistical Measures	Reliability Validity Reproducibility Responsiveness Interpretability Time to complete

	User friendliness
Study design	Questionnaire validation studies

9.3.2 Clinical evidence

We searched for studies that developed and assessed instrument(s) to measure treatment burden in adults with multimorbidity. Three studies were included in the review[89,231,232] these are summarised in Table 99 below. Evidence from these studies is summarised in the clinical evidence profile below (Table 100). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H and excluded studies list in Appendix L.

Four tools were included in the review: the Treatment Burden subscale[89]; the Activity limitation subscale[89]; the Treatment Burden Questionnaire 2012 (French version)[232]; and the Treatment Burden Questionnaire 2014 (English version)[231]. The data from the 4 questionnaires was not pooled due to variation in the content of the tools, and so data for each of the tools is reported separately. Two tools[89] were validated in a population of people with multimorbidity. The other 2 tools[231,232] were validated in an adult population where the number of people with multimorbidity was not reported.

Table 99: Summary of studies included in the review

Study	Questionnaire	Population	Setting	Comments
Gibbons 2013[89]	Two subscales relevant to the review were extracted from a broader questionnaire. These were the Treatment Burden subscale and the Activity Limitation subscale. <u>Treatment Burden subscale</u> 6 items. All items were rated and scored on a 4-point scale from 1 'strongly disagree' to 4 'strongly agree'. 1. Taking medications for each of my conditions has caused me problems 2. Having more than one condition makes my treatments less effective 3. It is difficult to take all of the medications the way I am supposed to 4. Having more than one condition makes it difficult to get the best available treatment 5. I don't like mixing medications for different conditions 6. I feel so overwhelmed by the treatment for one condition that it is hard to manage any others <u>Activity Limitation subscale</u> 3 items. All items were rated and scored on a 6-point scale from 0 'strongly disagree' to 6 'strongly agree'. 1. Time spend managing my condition has made it more difficult to carry out my usual	n=490 Adults (mean 70±10 years); outpatient Gender (M:F): 49:51 Multimorbidity: 100% (number of conditions 2-5 34.2%, 6-10 50.2%, 11 or more 15.6%; mean number of conditions 7.3±3.2)	4 GPs, Greater Manchester, England Assessment format and setting not reported	Part of the Multimorbidity Illness Perceptions Scale (MULTIPLEs) Construct validity assessed using: Brief Illness Perception Questionnaire (BIPQ); Hospital Anxiety and Depression Scale (HADS)

Study	Questionnaire	Population	Setting	Comments
	activities 2. Time spent managing my conditions has reduced my social life 3. Spending time managing my conditions has limited my activities			
Tran 2012[232]	<u>Treatment Burden Questionnaire (TBQ)</u> Aims to measure the extent to which healthcare impacts on the functioning and wellbeing of people with chronic condition(s), apart from specific treatment side effects. 7 constructs (13 items) assessing the extent to which patients believed each item caused them 'burden'. All items were rated and scored on a 10-point scale ranging from 0 'no burden' to 10 'considerable burden') Items in TBQ (translated from French to English): 1. Medication: – 1a. Taste, shape or size of your tablets and/or inconvenience caused by your injections (for example, pain, bleeding, scars) – 1b. Number of times you have to take your medication daily – 1c. Things you do to remind yourself to take your daily medication and/or to manage your treatment when not at home – 1d. Specific conditions when taking your medication (for example, taking it at a specific time of day or meal, not being able to do certain things after taking them like driving or lying down) 2. Assessments/appointments: – 2a. Lab tests and other exams (frequency, time spent and inconvenience of these exams) – 2b. Self-monitoring (for example, taking your blood pressure or measuring your blood sugar yourself: frequency, time spent and inconvenience of this surveillance) – 2c. Doctors' visits (frequency and time spent for visits) – 2d. Arrange appointments and schedule doctors' visits and lab tests 3. How would you rate the burden associated with taking care of paperwork from health insurance agencies, welfare organisations, hospitals and/or social care? 4. How would you rate the constraints associated with your diet (for example, not	n=502 Adults (mean 59.3± 17 years); inpatient (51.2%) Gender (M:F): 47:53 Multimorbidity: number of people with multimorbidity not reported	6 hospitals, Paris, France Assessment format and setting not reported	Developed and conducted in French Construct validity assessed using: Treatment Satisfaction Questionnaire for Medication (TSQM)

Study	Questionnaire	Population	Setting	Comments
	being able to eat certain foods)? 5. How would you rate the burden associated with the recommendations from your doctors to practise regular physical exercises? 6. What is the impact of your healthcare on your social relationships (for example, need for assistance, being ashamed to take your medication in front of people)? 7. 'Frequent healthcare reminds me of my health problems'			
Tran 2014[231]	<u>Treatment Burden Questionnaire (TBQ)</u> Aims to measure the 'work' of being a person with chronic condition(s) (that is, challenges associated with everything patients have to do to take care of themselves) and its effect on quality of life. 15 items assessing the extent to which patients believed each item caused them problems. All items rated and scored on a 10-point scale ranging from 0 'not a problem' to 10 'large problem'. 1. Taste, shape or size of your tablets and/or the annoyances caused by your injections (for example, pain, bleeding, bruising or scars) 2. Number of times you should take your medication daily 3. Efforts you make not to forget to take your medications (for example, managing your treatment when you are away from home, preparing and using pillboxes) 4. Necessary precautions when taking your medication (for example, taking them at specific times of the day or meals, not being able to do certain things after taking medications such as driving or lying down) 5. Lab tests and other exams (for example, blood tests or radiology): frequency, time spent and associated nuisances or inconveniences 6. Self-monitoring (for example, taking your blood pressure or checking your blood sugar): frequency, time spent and associated nuisances or inconveniences 7. Doctor visits and other appointments: frequency and time spent for these visits and difficulties finding healthcare providers 8. Difficulties you could have in your relationships with healthcare providers (for example, feeling not listened to enough or not taken seriously) 9. Arranging medical appointments and/or transportation (doctors' visits, lab tests and other exams) and reorganizing your schedule	n=610 Adults (mean 51.5 ± 12.4 years); outpatient Gender (M:F): 23:77 Multimorbidity: number of people with multimorbidity not reported (mean number of chronic conditions 2.9 ± 1.9)	Mainly USA, UK, Canada, Australia, New Zealand Assessment format: self-report , online	TBQ (2012) translated into England by forward-backward translation method. Addition of items: financial burden of healthcare; relationships with healthcare providers Construct validity assessed using: Patients Like Me Quality of Life (PLMQOL) scale; Morisky's Medication Adherence Scale 8 (MMAS-8); patient's knowledge of their conditions and treatments; clinical variables (for example, number of conditions)

Study	Questionnaire	Population	Setting	Comments
	around these appointments 10. Administrative burden related to healthcare (for example, all you have to do for hospitalizations, insurance reimbursements and/or obtaining social services) 11. Financial burden associated with your healthcare (for example, out-of-pocket expenses or expenses not covered by insurance) 12. Burden related to dietary changes (for example, avoiding certain foods or alcohol, having to quit smoking) 13. Burden related to doctors' recommendations to practice physical activity (for example, walking, jogging, swimming) 14. How does your healthcare impact your relationships with others (for example, being dependent on others and feeling like a burden to them, being embarrassed to take your medications in public) 15. 'The need for medical healthcare on a regular basis reminds me of my health problems'			

Table 100: Clinical evidence profile

No. of studies	n	Risk of bias	Indirectness	Internal reliability	Construct validity	Reproducibility	Responsiveness	Interpretability
Treatment Burden subscale								
1	490	HIGH[a]	No indirectness	Person Separation Index (PSI) 0 = 0.7 Cronbach's alpha = 0.9 Factor loadings range 0.5 – 0.84 Unidimensionality: t-test 1.2%	Total scores on the Treatment Burden subscale were correlated with individual items on the HADS and BIPQ. Authors state that correlations >0.5 would be indicative of construct validity of the subscale, not indication of direction of effect was given. Hospital Anxiety and Depression Scale (HADS) Anxiety, Spearman's Rho (rs) = 0.52 Depression rs = 0.53 Psychological distress rs = 0.55 Brief Illness Perception Questionnaire (BIPQ) Emotional affect rs = 0.44 Impact of illness rs = 0.32 Concern rs = 0.28 Experience of symptoms rs = 0.25 Perceived control of illness rs = -0.16 Efficacy of treatment rs = -0.16 Understanding of illness rs= 0.11	rs = 0.63 (Retest after 1 month)	Not assessed	Not assessed
Activity Limitation subscale								
1	490	HIGH[a]	Serious Indirectness	PSI = 0.65 Cronbach's alpha = 0.8	Total scores on the Activity Limitation subscale were correlated with individual items on the HADS and BIPQ. Authors state that correlations >0.5 would be indicative	rs = 0.6 (Retest after 1 month)	Not assessed	Not assessed

No. of studies	n	Risk of bias	Indirectness	Internal reliability	Construct validity	Reproducibility	Responsiveness	Interpretability
				Factor loadings range 0.59-0.79 Unidimensionality: t-test 29%	of construct validity of the subscale, not indication of direction of effect was given. HADS Anxiety rs = 0.52 Depression rs = 0.53 Psychological distress rs = 0.55 BIPQ Emotional affect rs = 0.46 Impact of illness rs = 0.45 Experience of symptoms rs = 0.38 Concern rs = 0.35 Perceived control of illness rs = -0.19 Efficacy of treatment rs = -0.17			
Treatment Burden Questionnaire (French language) (2012)								
1	501	LOW	Serious indirectness	Cronbach's alpha = 0.89	Hypothesis: negative correlation between treatment burden and treatment satisfaction. Total scores on the Treatment Burden Questionnaire (French language) were correlated with the total scores on the Treatment Satisfaction Questionnaire for Medication (TSQM), rs= -0.41. Authors state that correlations >0.5 would be considered high, and 0.35-0.50 moderate.	Intraclass correlation coefficient (ICC) 0.76 (95% CI 0.67 to 0.83) (Retest after 2 weeks (36%) or 1 month (6%))	Not assessed	Mean total scores (SD) of the TBQ different subgroups Age Aged <60 (n= 243): 38.4± 26.7 Aged >60 (n= 259): 24.0± 21.9 Setting Inpatients (n=257): 34.7±27.7

No. of studies	n	Risk of bias	Indirectness	Internal reliability	Construct validity	Reproducibility	Responsiveness	Interpretability
								Outpatients (n= 245): 27.1± 23.1 Reported main chronic condition Diabetes (n=81): 46.4±28.6 Rheumatologic diseases (n=59): 28.6± 26.3 High blood pressure and dyslipidemia (n=44): 18.5± 19.9 Systemic diseases (n=43): 39.0± 26.3 Pulmonary diseases (other than asthma) (n=40): 24.8± 17.5 Heart diseases (n=37): 29.3± 23.7
Treatment Burden Questionnaire (2014)								
1	610	MODERATE [a]	Serious indirectness	Not reported	Total scores on the Treatment Burden Questionnaire (English language) were correlated with various measures to test the following hypothesis: Quality of life Hypothesis: negative correlation between treatment burden (as measured by the TBQ global score) and quality of life.	ICC 0.77 (95% CI 0.70 to 0.82) (Retest after 2 weeks)	Not assessed	TBQ validated in different subgroups (mean TBQ score ±SD) Chronic condition(s) Gastrointestinal diseases (n=128): 65.4±32.5 Skin diseases (n=68): 64.9±30.8 Fibromyalgia (n=77):

No. of studies	n	Risk of bias	Indirectness	Internal reliability	Construct validity	Reproducibility	Responsiveness	Interpretability
					PatientsLikeMe Quality of Life (PLMQOL) scale Total rs = −0.5 Range rs = −0.39 to rs = −0.5 Treatment burden Hypothesis: the greater the treatment burden, the lower the adherence to treatment. Morisky's Medication Adherence Scale 8 (MMAS-8) (mean TBQ score ± SD) High/moderate adherence 37.7 ± 27.5 Low adherence 61.8 ± 30.5 Patient knowledge Hypothesis: the greater the patient's knowledge of their conditions and treatments, the lower the treatment burden Sufficient knowledge of conditions 49.3 ± 30.7 (mean TBQ score ± SD) Insufficient knowledge of conditions 63.0 ± 31.6 Sufficient knowledge of treatments 47.8 ± 30.4 Insufficient knowledge of treatments 62.3 ± 31.3			64.7±32.2 Lung diseases (n=90): 64.3±35.0 Rheumatologic diseases (n=201): 62.2±31.7 Psychiatric diseases (n=245): 61.3±32.7 Diabetes (n=42): 60.1±35.6 Other endocrine disorders (n=119): 57.8±32.8 Heart diseases (n=34): 57.8±38.7 Kidney diseases (n=37): 57.7±36.8 Vision problems (n=83): 57.5±36.0 Cancer or malignant blood diseases (n=30): 57.4±36.3 Hearing problems (n=48): 55.9±30.4 High blood pressure (n=153): 51.9±31.3 Neurologic diseases (n=270): 51.8±30.1 Infectious diseases (n=18): 51.2±28.1 Stroke or cerebrovascular diseases (n=17): 50.3±39.2

No. of studies	n	Risk of bias	Indirectness	Internal reliability	Construct validity	Reproducibility	Responsiveness	Interpretability
					Clinical variables Hypothesis: positive correlation between treatment burden and the following clinical variables Number of conditions 1: 44.3±29.1 (mean TBQ score ± SD) Number of conditions 2-3: 49.7±29 Number of conditions >4: 65.4±33 Number of tablets and pills/day rs=0.2 Number of injections/week rs=0.11 Number of drug administration(s)/day rs=0.25 Number of different doctors patient sees regularly rs=0.21 Number of appointments/month rs=0.25 Number of hospitalization/year rs=0.11			

PSI, Person Separation Index; rs, Spearman's Rho; ICC, Intraclass correlation coefficient

(a) Risk of bias was assessed using the Q-BAST checklist. Moderate risk of bias: downgraded by 1 increment as 1 item was at high risk of bias. High risk of bias: downgraded by 2 increments as 2 items were at high risk of bias

9.3.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

9.3.4 Evidence statements

Clinical

- One questionnaire development study with 490 participants demonstrated that a 6-item questionnaire performed highly in terms of internal reliability and construct validity, and moderately in terms of reproducibility. No data were provided on the responsiveness, interpretability, time to complete or user friendliness of the scale. This evidence was at high risk of bias.
- One questionnaire development study with 490 participants was at high risk of bias and demonstrated that a 3-item questionnaire performed highly in terms of internal reliability and construct validity, and moderately in terms of reproducibility. No data were provided on the responsiveness, interpretability, time to complete or user friendliness of the scale. This evidence was at high risk of bias.
- One questionnaire development study with 502 participants was at low risk of bias and demonstrated that a 13-item questionnaire performed highly in terms of internal reliability, reproducibility and interpretability, and low in terms of construct validity. No data were provided on the responsiveness, time to complete or user friendliness of the scale. This evidence was at low risk of bias.
- One questionnaire development study with 610 participants was at moderate risk of bias and demonstrated that a 15-item questionnaire performed highly in terms of internal reliability, construct validity, reproducibility and interpretability. No data were provided on the responsiveness, time to complete or user friendliness of the scale. This evidence was at moderate risk of bias.

Economic

- No relevant economic evaluations were identified.

9.3.5 Recommendations and link to evidence

Recommendations	**20. Establish disease burden by talking to people about how their health problems affect their day-to-day life. Include a discussion of:** • **mental health** • **how disease burden affects their wellbeing** • **how their health problems interact and how this affects quality of life.** **21. Establish treatment burden by talking to people about how treatments for their health problems affect their day-to-day life. Include in the discussion:** • **the number and type of healthcare appointments a person has and where these take place** • **the number and type of medicines a person is taking and how often** • **any harms from medicines** • **non-pharmacological treatments such as diets, exercise programmes and psychological treatments** • **any effects of treatment on their mental health or wellbeing.** **22. Be alert to the possibility of:** • **depression and anxiety (consider identifying, assessing and managing these conditions in line with the NICE guideline on common mental health disorders)** • **chronic pain and the need to assess this and the adequacy of pain management.**
Relative values of different outcomes	The GDG considered internal validity, construct validity, reproducibility, responsiveness and interpretability as metrics for evaluating treatment burden questionnaires. For assessing construct validity, the GDG considered the evidence and agreed with the associations for the following measures: negative correlation with quality of life, medication adherence and knowledge about conditions; positive correlation with number of long term conditions, number of medications and use of healthcare resources. For the evaluation of reproducibility, the GDG considered whether treatment burden was consistent over time and therefore whether stability of people's scores was necessary criteria for evaluating tools to assess treatment burden. The GDG believed that the treatment of people with multimorbidity was unlikely to change over a short period of time, and they expected that self-reported treatment burden should remain stable over that time. The GDG therefore decided that treatment burden questionnaires should have good test-retest reliability over short periods (1 month or less). The GDG felt that responsiveness was an important metric to assess the ability of treatment burden questionnaires to detect change in treatment burden, for example following an intervention. The GDG also thought that it was important for treatment burden questionnaires to report data to aid interpretation of scores on the tool, so as to inform a recommendation and to provide guidance for clinicians using treatment burden questionnaires.

Trade-off between clinical benefits and harms	The GDG noted that all tools performed adequately in all of the domains that were reported in published papers. However, limited data was reported on responsiveness and interpretability. The Treatment Burden subscale performed highly in terms of internal reliability and construct validity, and moderately in terms of reproducibility. The Activity Limitation subscale performed highly in terms of internal reliability and construct validity, and moderately in terms of reproducibility. The TBQ (French version) performed highly in terms of internal reliability, reproducibility and interpretability, and low in terms of construct validity. The TBQ (English version) performed highly in terms of internal reliability, construct validity, reproducibility and interpretability. The GDG agreed that none of the tools performed highly in all domains and that no one tool outperformed the others. Further, the GDG noted that data on interpretability was only reported for 2 tools, which would be needed to aid interpretation of the results in clinical practice, and that no data on responsiveness was reported, and so it would be difficult to use the tool in clinical practice to assess how a patient's burden has changed over time. Therefore the GDG did not feel that they could recommend the use of 1 particular tool in clinical practice. However, the GDG felt that some of the items in the tools were important factors to consider when assessing treatment burden, for example: the number of medicines being taken; frequency of medicines being taken; psychological treatments; number of appointments; dietary requirements; exercise requirements; and how treatment impacts social relationships. The GDG agreed that assessing treatment burden should include a discussion of such factors, alongside what matters most to the person with multimorbidity, for example, health priorities and treatment preferences. The GDG agreed that assessing treatment burden would not cause any harm and in most cases the assessment of treatment burden was likely to benefit people by increasing their awareness of the burden that their treatment may cause and through initiating a conversation about these issues. However, using an inappropriate tool or using a tool without proper patient engagement may cause harm as there would be a risk that treatment burden would not be assessed accurately. Additionally, the use of the treatment burden questionnaire may cause harm through the process potentially becoming a 'tick-box' exercise where the clinician does not actively engage with people about their treatment burden. The GDG noted that no data was provided on time to complete and on user friendliness.
Economic considerations	No relevant economic evaluations were identified. The GDG considered that, although this recommendation may have cost implications as a result of additional health care professional time, they felt that discussion with people about their perceived treatment burden was important as it could help facilitate discussion regarding treatment and identify any of change in treatments required.
Quality of evidence	The risk of bias for the studies ranged from high to low. The evidence for the Treatment Burden subscale was at high risk of bias due to evidence of floor and ceiling effects and the rate of missing data from responders. The evidence for the Activity Limitation subscale was at high risk of bias due to the rate of missing data from responders. The evidence for the Treatment Burden Questionnaire (French version) was at low risk of bias. The evidence for the Treatment Burden Questionnaire (English version) was at moderate risk of bias due to not all relevant data being reported. No studies provided responsiveness data. Correspondingly, the GDG expressed concern about using any of the treatment burden questionnaires in clinical practice to assess change in treatment burden over time and following intervention. The GDG also expressed concern about the lack of data on the interpretability of the questionnaires. Interpretability data were only provided for 2 questionnaires. The

	GDG noted that even if 1 of the tools had performed highly, it would be difficult to identify how clinicians should use this in practice without information on what score is deemed a 'high' treatment burden score. Only 1 of the studies reported that the population had multimorbidity. The studies that did not report the number of people with multimorbidity were downgraded for indirectness as the GDG was unsure whether the performance of tools may be different in a population of people with multimorbidity.
Other considerations	The GDG agreed that there was insufficient evidence to recommend the use of a treatment burden questionnaire in clinical practice. While the GDG felt that their recommendation would facilitate greater discussion of treatment burden in practice, the GDG believed that the use of a formal treatment burden questionnaire could be beneficial, but that further research is needed to support this. This research should include provide guidance on how scores should be interpreted and an assessment of whether questionnaires can capture change in treatment burden over time. The GDG felt that clinicians should regularly discuss treatment burden with people with multimorbidity. In particular, the GDG thought that it was important for clinicians to initiate discussion of treatment burden with their patients when introducing new treatments and when conducting the annual medication review. The GDG noted that the question 'how do you rate your treatment burden?' alone would not be sufficient to prompt this discussion as patients may not be able to easily interpret the concept of treatment burden or know of its components. The GDG agreed that clinicians should explore the patient's perspective of their treatment burden using a series of prompts which highlight important components that may be part of the person's treatment burden (for example, number of pharmacological and non-pharmacological treatments or appointments, and the person's perception of impact on their lives). The GDG also wished to note that clinicians should ask people with multimorbidity about their social circumstances, which may also impact on treatment burden. The GDG used the content of the questionnaires identified in this review and their own professional and personal experience to develop a list of the areas they thought should be covered in conversations about treatment burden. The GDG used the evidence review examining barriers to optimising care (see section 6.2) to develop the recommendation on assessing disease burden. That review indicated that the presence of multimorbidity itself can be a burden but that people also report an effect of this on their mental health and wellbeing. The GDG considered that the treatment burden questionnaire and questions developed from it did not adequately cover issues related to the burden of people's conditions and that a true understanding of people's experience could not be achieved without exploring this. The GDG also chose to explicitly cross-refer to NICE guideline on common mental health disorders which include appropriate ways of case finding depression and anxiety which are common problems and often co-exist with physical disease. The known presence of mental health disorders or the new identification of mental health disorders may indicate that additional time is required to deliver a multimorbidity approach to care. Working across physical and mental health service boundaries may also be more difficult when considering reducing or discontinuing medicines. The GDG were also aware that epidemiological studies indicate the high prevalence of pain in people with multimorbidity. This can be easily overlooked as it may not always be related to a specific diagnosis or is related to musculoskeletal problems which are not appropriately recorded or coded in patient records. Pain can however be a significant cause of morbidity and associated with polypharmacy.

9.4 Establishing patient preferences, values and priorities

9.4.1 Recommendations and link to evidence

Recommendations	**23. Clarify with the patient whether and how they would like their partner, family members and/or carers to be involved in key decisions about the management of their conditions. Review this regularly. If the patient agrees, share information with their partner, family members and/or carers. [This recommendation is adapted from the NICE guideline on patient experience in adult NHS services.]** **24. Encourage people with multimorbidity to clarify what is important to them, including their personal goals, values and priorities. These may include:** • **maintaining their independence** • **undertaking paid or voluntary work, taking part in social activities and playing an active part in family life** • **preventing specific adverse outcomes (for example, stroke)** • **reducing harms from medicines** • **reducing treatment burden** • **lengthening life.** **25. Explore the person's attitudes to their treatments and the potential benefits and harms of those treatments. Follow the recommendations on patient involvement in decisions about medicines and understanding the patient's knowledge, beliefs and concerns about medicines in the NICE guideline on medicines adherence.**
Relative values of different outcomes	The GDG were interested in ensuring that people who might benefit from a multimorbidity approach to care were fully involved in the process and that the process conformed to the principles of shared decision making and good practice on prescribing.
Trade-off between clinical benefits and harms	The GDG did not consider there were any harms likely from ensuring people are asked about how their family or carers should be involved. They were aware that raising the issue of people's priorities and preferences might be a sensitive issue but as long as this is done sensitively it should allow decisions about medicines and treatments to be made in line with the person's values.
Economic considerations	The GDG considered the potential resource costs of these recommendations. In general the recommendations emphasise the importance of recommendations from generic NICE guidelines around patient involvement and experience, for this specific population. The GDG noted that although this may result in additional time required in consultations this will vary significantly between people, should generally already be considered best practice and is very difficult to model. The recommendations provide guidance on ensuring that family members and carers are involved in the way the person with multimorbidity would like them to be, that their preferences and priorities are recognised and that their attitudes to and views about medicines are included in the conversation. The GDG considered that the delivery of an approach to care that takes account of multimorbidity could be carried out as part of usual medical practice.

Quality of evidence	These recommendations were informed by other NICE guidance and evidence reviews in chapter 6 and this chapter where quality of evidence contributing to the recommendations is discussed.
Other considerations	The GDG considered that it was important to emphasise the importance of ensuring that family members of the person with multimorbidity are involved in decisions if that is what the person wants. Many people who will benefit from a multimorbidity approach to care may be elderly and frail and will have significant support from families in their lives. They therefore agreed to explicitly cross refer to this recommendation from the NICE guideline on Patient Experience in NHS adult services. One of the important principles in shared decision making is ensuring that people's decisions are in line with their values. The GDG considered that healthcare professionals often are not explicit in their discussions with people about what particular medicines may achieve. If people have significant issues with treatment burden it is particularly important to explore what is important to them to achieve by their treatments. For some people, preventing a specific outcome may be of great importance because a family member may have suffered for example from stroke. For others the side effects of treatments or even of taking medicines at all may not fit with people's priorities. This discussion is a necessary as a basis for later discussion of treatment's someone is taking. NICE guideline on Medicines Adherence includes more detailed recommendations on exploration of people's understanding of how their medicines work and what they will achieve and both about how to support people to take the medicines they wish to take and support them if they do decide not to take all their medicines. The GDG wished to make explicit cross referral to the Medicine Adherence guideline.

9.5 Effectiveness of interventions from condition-specific guidance

9.5.1 Review question: How might data from condition-specific guidance best be used and presented to inform a ranking of treatments based on absolute risk and benefit and time to achieve benefits?

For full details see review protocol in Appendix C.

Table 101: PICO characteristics of review question

Objective	To develop an example of how data from condition-specific guidance may be presented to inform a ranking of treatments as part of decisions to optimise care amongst people with multimorbidity.
Conditions and interventions	• Hyperlipidemia (statins) • Hypertension (ACE inhibitors, beta blockers, calcium channel blockers, thiazides, angiotensin receptor blockers) • Type II diabetes (Metformin hydrochloride, sulfonylureas, DPP4 inhibitors) • Chronic heart failure (ACE inhibitors, beta blockers) • Atrial fibrillation (anticoagulants) • Chronic kidney disease (ACE inhibitors, angiotensin receptor blockers, spironolactone) • Angina (aspirin) • Depression (antidepressants) • Schizophrenia (anti-psychotics) • Migraine (prophylaxis)

Outcomes	The following metrics will be reported/calculated: • Demographics of trial participants • Duration of treatment • Outcome (critical outcomes; including mortality and serious adverse events) • Length of follow-up • Event rate as reported/calculated • Relative risk (95% CI) • Absolute benefit (95% CI) • Annualised absolute benefit (95% CI) • Number needed to treat (95% CI) • Annualised number needed to treat (95% CI)
Study design	Published NICE guidelines. Quality assessment of data will not be conducted.

This review sought to develop a method for re-presenting data from single-condition NICE guidelines in a format that could enable clinicians and people with multimorbidity to rank treatments according to their effectiveness so that this could be used to inform decisions about medicines in line with people's own values and preferences. The GDG believed that this tool could inform treatment decisions between healthcare professionals and people with multimorbidity. In particular, the GDG felt that this resource may be most useful in cases where people with multimorbidity are experiencing treatment burden and would like to discuss withdrawing treatment(s), and in instances where people with multimorbidity may not be expected to experience the full benefit of prophylactic treatment (for example, due to reduced life expectancy).

Within the timeframe of this guideline, the GDG decided to prioritise a limited number of conditions and interventions in the resource. The GDG used the following criteria as inclusion criteria:

- *Guidelines for conditions that commonly occur amongst people with multimorbidity*
- *Chronic conditions*
- *Conditions where the effect of the treatment is not observable*
- *Treatments aimed at preventing the onset or exacerbation of existing conditions (that is, primary and secondary prevention)*
- *First line treatments*

Recent epidemiological data was used to identify conditions commonly occurring in people with multimorbidity[101] and this was cross checked with information on commonly prescribed drugs from the Health and Social Care Information Centre. Data is taken from evidence used in relevant guideline and includes both single-condition and multi-morbid populations.

9.5.2 Clinical evidence

Ten NICE clinical guidelines were included in the review[156-159,161-164,166,168]; these are summarised in Table 102 below.

The GDG prioritised the inclusion of first line treatments for each of the included conditions. All treatment effectiveness data extracted is for intervention versus placebo comparisons. The GDG also prioritised the inclusion of outcomes specified as critical for decision-making by the expert GDGs for each of the guidelines, where these were stated. The GDG noted that this resource should be used to

inform decisions on treatment in addition to discussions between healthcare professionals and people with multimorbidity about other effects of treatment (for example, side effects of treatment that are important to a person but not represented in the resource). Evidence is missing where insufficient information has been reported in the guideline to enable calculation (for example, in older guidelines where event rate data is not reported).

Table 102: Summary of guidelines included in the review

Guideline	Population	Intervention	Outcomes	Comments
Stable Angina: management (2011)	Adults (mean age range 63 – 67 years)	Aspirin	Mortality; fatal myocardial infarction; non-fatal myocardial infarction	
Atrial fibrillation: management (2014)	Adults (mean age range 67 – 74 years)	Anticoagulants	Mortality; ischaemic stroke;	Majority of evidence excluded people who had previously experienced stroke or transient ischaemic attack
Cardiovascular disease: risk assessment ad reduction, including lipid modification (2014)	Adults	Statins	Mortality; cardiovascular mortality; non-fatal myocardial infarction; stroke	Separate data reported for primary and secondary prevention
Chronic Heart Failure in adults: management (2010)	Older adults (age >65 years)	Beta-blockers	Mortality; sudden death; hospitalisation; number of people who experience adverse event	
Chronic Kidney Disease in adults: assessment and management (2014)	Adults (mean age range 55 – 70 years)	ACE inhibitors; Angiotensin II receptor blockers; Spironolactone	Mortality, cardiovascular events; progression of CKD (change in eGFR); progression of CKD (change in ESRD); acute kidney injury	Majority of evidence included people with type I or type II diabetes
Depression in adults: recognition and management (2009)	Adults	Antidepressants	Relapse	Participants in approximately half of the included studies received treatment for <6 months prior to randomisation
Hypertension in adults: diagnosis and management (2011)	Adults with isolated systolic hypertension (SBP 160 – 219 mmHg and DBP <90 mmHg)	Antihypertensive drug therapy (all); Bendroflumethiazide; Indapamide; Chlorthalidone, ACE inhibitors,	Mortality, myocardial infarction; stroke; coronary heart disease event; quality of life (no	Some data was extracted from the 2006 guideline [165], where analyses had not been updated in the most recent

Guideline	Population	Intervention	Outcomes	Comments
		angiotensin II receptor blockers, beta blockers, calcium channel blockers	limitations in daily activities)	guideline. The evidence for some of the more commonly used treatments for hypertension (for example, ACE inhibitors) was only available for limited outcomes.. This is because these recommendation were published in the 2004 guideline [181]), and limited data from that guideline was repeated in the 2011 guideline.
Headaches in over 12s: diagnosis and management	Children and adults (mean age range 14 – 41 years) experiencing recurrent migraine	Beta-blockers; Topiramate	People with >50% reduction in migraine days	
Psychosis and schizophrenia in adults: prevention and management (2014)	Adults	Second generation antipsychotics	Relapse	Mixed inpatient and community settings
Type II diabetes in adults: management (2015)	Adults (mean age range 51 – 72 years) receiving intervention as first line/monotherapy	Metformin; Pioglitazone; Sulfonylurea; Linagliptin; Saxagliptin; Sitagliptin; Vildagliptin	Hypoglycemia; stopping treatment due to adverse events	

9.5.3 Economic evidence

This question is considered to have no economic implications as it is about the format for presenting data on treatment ranking.

9.5.4 Recommendations and link to evidence

Recommendations	**26. When reviewing medicines and other treatments, use the database of treatment effects to find information on:** • **the effectiveness of treatments** • **the duration of treatment trials** • **the populations included in treatment trials.** **27. Consider using a screening tool (for example, the STOPP/START tool in older people) to identify medicine-related safety concerns and medicines**

	the person might benefit from but is not currently taking. [This recommendation is adapted from the NICE guideline on medicines optimisation.] **28. When optimising treatment, think about any medicines or non-pharmacological treatments that might be started as well as those that might be stopped.** **29. Ask the person if treatments intended to relieve symptoms are providing benefits or causing harms. If the person is unsure of benefit or is experiencing harms from a treatment:** • **discuss reducing or stopping the treatment** • **plan a review to monitor effects of any changes made and decide whether any further changes to treatments are needed (including restarting a treatment).** **30. Take into account the possibility of lower overall benefit of continuing treatments that aim to offer prognostic benefit, particularly in people with limited life expectancy or frailty.** **31. Discuss with people who have multimorbidity and limited life expectancy or frailty whether they wish to continue treatments recommended in guidance on single health conditions which may offer them limited overall benefit.** **32. Discuss any changes to treatments that aim to offer prognostic benefit with the person, taking into account:** • **their views on the likely benefits and harms from individual treatments** • **what is important to them in terms of personal goals, values and priorities (see recommendation 24).**
Relative values of different outcomes	The GDG discussed what metrics should be used to compare treatment effectiveness data. Currently, full NICE guidelines report information about the trial populations and settings (included studies and clinical evidence tables), follow-up time, baseline risk/event rate data, relative effect data, and absolute effect data (clinical evidence summary tables). The GDG felt that this information was informative and should be reproduced in the resource. The GDG also agreed to add several additional measures that they felt may facilitate the ranking of treatments and decisions between healthcare professionals and people with multimorbidity. These included: • Details on the duration of treatment in trials. While this information is provided in the clinical evidence tables of NICE guidelines, the GDG felt that including these in the resource would inform clinicians and people with multimorbidity on the length of treatment that corresponds with the treatment effects in trials • Numbers needed to treat (NNT). The GDG were aware of evidence that some people find NNT easier to interpret than relative and absolute measures of effect. The GDG felt that this data should be available to facilitate discussions between clinicians and people with multimorbidity who prefer NNT • Annualised absolute effect and NNT. Current NICE guidelines present treatment effect data alongside the follow-up time used in the included trials. The GDG noted that the size of the treatment effect will be influenced by the length of

	follow-up. In cases where the length of follow-up varies between different treatments, the GDG felt that this may therefore make it difficult for clinicians and people with multimorbidity to easily compare and rank treatment effects. Annualised data standardises the effect of treatment to one year, and therefore can be used as a method for clinicians and people with multimorbidity to compare treatment effects more easily. The GDG agreed that the resource should be interactive, so that healthcare professionals and people with multimorbidity can alter the data displayed according to need and their own preferences.
Quality of the clinical evidence	The GDG decided not to present data alongside GRADE quality ratings from the original guideline. This is because the primary aim of this resource is to inform treatment decisions for people with multimorbidity, which is not the population included in the trial populations. As a consequence, some aspects of the quality rating may not be applicable. There are several limitations of the data included in the resource. Firstly, data for some metrics is missing for a small number of conditions. This is because insufficient data was provided in the original guideline for calculation. This was mostly relevant to older guidelines, where raw event rate data was not provided. Where information about follow-up time was not reported for each individual outcome in the original guideline, these were estimated based on information in the included studies table or extrapolated from other outcomes. In some cases, limited data about the populations included in the studies was available. While annualised data may be useful for comparing treatment effectiveness across trials with different follow-up times, the GDG note that annualised data is calculated with the assumption that the effect of treatment will be stable across time; that is, the number of events will be the same after 1 year, 2 years, 3 years and so on. In many cases this may not be the case (for example, for treatments with augmentative effects over time, and in cases where the risk of events may increase as a person ages and has the condition for longer). As a consequence, the GDG suggested that healthcare professionals and people with multimorbidity who use the resource may wish to consider the likely effect of time and consider additional metrics alongside annualised data.
Economic considerations	The GDG considered the potential resource costs of these recommendations. In general the recommendations emphasise the importance of recommendations from generic NICE guidelines around patient involvement and experience, for this specific population. The GDG noted that this may result in additional time required in consultations but that this will vary significantly between peopleand it is difficult to quantify. Further reviewing medicine prescriptions should lead to an optimisation of care and could reduce treatment burden and costs.
Other considerations	Resource: The GDG recognise that this resource only includes a small number of conditions and prescribed interventions that are received by people with multimorbidity in the UK. The GDG hoped that this resource may expand over time, with the publication of new NICE guidelines. In the meantime, the GDG felt that this resource will be useful to compare treatment effectiveness data for a small number of commonly occurring conditions and treatments, which can be used in addition to single-condition NICE guidelines for other conditions not included. The GDG felt that while this resource may be accessed by interested people, it is more likely to be used by clinicians to access information outside consultations. The GDG noted that healthcare professionals can tailor the data presented in the table according to preference and to choice of outcomes and conditions.

	The GDG believed that data in this resource should accompany, and not replace, discussion of treatment decisions between healthcare professionals and people with multimorbidity. These discussions should include discussion of a person's values and preferences towards different outcomes. For example, a person may prefer to reduce their risk of stroke rather than CV mortality. Healthcare professionals should also be aware that a person may wish to prioritise other outcomes not reported in this resource (for example, other side effects of medication or medicines that are unpleasant in taste). The GDG highlighted that clinicians should consider the applicability of the evidence in the resource to each person with multimorbidity. For example, healthcare professionals should consider whether the baseline risk of participants (that is, the event rate in people who did not receive the intervention) in the included trials indicates that the study populations were similarly at risk of an event occurring as the person in consultation. This is because the absolute benefit or harm of an intervention will vary depending on how likely a person is to experience the event; that is, fewer people at high risk of an event occurring need to be treated to avoid one adverse outcome, whereas more people at low risk of an event occurring will need to be treated to avoid one event. Clinicians should consider whether the person they are treating is at a higher or lower risk of an event occurring than the populations included in the evidence to inform treatment decisions. The GDG wished to highlight the importance of considering the duration of treatment of people included in the trials in the resource, and consider whether the effectiveness data may vary in a person with multimorbidity who has been receiving the intervention for a shorter or longer duration. For example, people who have been receiving treatments that are thought to have an augmentative effect for a longer duration than the study populations may be at a lower risk of adverse events if withdrawn from treatment. The GDG noted that healthcare professionals and people with multimorbidity may wish to consult the full evidence and supporting documents for recommendations in the original single-condition guidelines. The GDG noted that, in some cases, treatment effectiveness data may vary in people with multimorbidity compared to people with single-conditions only. This may because treatments may be less likely to lead to clinical benefit, or may lead to less clinical benefit, in people with other health conditions. People with multimorbidity may also be more likely to experience adverse outcomes; for example, because of interactions between medications and conditions. The GDG are aware of evidence that the relative effect of treatments has been shown to be consistent across populations in the majority of cases. However, as people with multimorbidity may not be represented in many clinical intervention trials, the GDG felt that healthcare professionals should be aware that actual treatment effectiveness in people with multimorbidity may vary from the data reported in this resource. The GDG discussed whether to include effectiveness data for non-pharmacological interventions in this resource. Within the timeframe of this guideline, the GDG agreed to restrict the resource to pharmacological interventions, so as to inform a ranking of treatments in people taking by multiple medications. However, the GDG note that non-pharmacological interventions may also be associated with treatment burden in people with multimorbidity (for example, restrictive diets and exercise programmes). The GDG believed that healthcare professionals should also discuss treatment burden with respect to non-pharmacological treatments with people with multimorbidity, and may wish to discuss stopping or adapting these treatments if appropriate. The GDG agreed not to report confidence intervals for NNT and aNNT data. This is because it is difficult to interpret the meaning of confidence intervals where the lower confidence interval is negative and the upper confidence interval is positive (therefore being consistent with a number needed to benefit and a number needed to harm). This was the case for several of the analyses included in the resource. The

	GDG felt that in cases where healthcare professionals or people with multimorbidity wished to use NNT, they may wish to consider other metrics of imprecision in the resource, so as to consider the uncertainty of the data. Healthcare professionals need to take into consideration the health literacy and numeracy of each person when discussing evidence for treatments. STOPP/START tool and other recommendations The resource developed in this guideline provides detail on benefits and risk of medicines extracted from other NICE guidance. Other tools are available and the GDG agreed to include a recommendation to use of the STOPP/START tool which is recommended in the NICE medicines optimisation guideline. This tool is recommended in that guideline as a possible screening tool to identify potential medicines-related safety incidents in groups such as adults, children and young people taking multiple medicines, adults, children and young people with chronic or long-term conditions and older people. The GDG reviewed the evidence in the Medicines Optimisation guideline which found low quality evidence from one RCT on use of STOPP/START tool which supported its use in elderly hospitalised patients but that it has also been used in other settings. The tool however also identifies medicines that people might benefit from which they are not already taking and the GDG adapted the wording to emphasise this aspect. The GDG considered it important that while reviewing medicines and their efficacy might need to reduction in treatments, it was important to be aware that the aim of review was not about reducing medicines per se but about optimising treatments. The GDG developed a separate recommendation to make explicit that optimising medicines might result in starting of some medicines as well as stopping of medicines. The GDG added consensus recommendations to outline that medicines and treatments that aim for preventative or prognostic benefit should be reviewed particularly in light of a person's life expectancy and their priorities and preferences as discussed in section 9.4. Symptomatic treatments The GDG agreed that their experience was that people are often started on treatments for symptom control and they remain on these treatments without adequate review. They considered that discussion about reviewing the efficacy of symptomatic treatments should be explored and that individual trials of stopping or reducing treatments might be appropriate.

9.6 Stopping drugs: antihypertensive treatment

The scope for the guideline included reviewing evidence for the effect of stopping drugs. It had been envisaged that given the large number of people taking medicines such as statins and antihypertensives that such evidence would be available. As part of guideline development initial review protocols were developed to examine the effect of stopping antihypertensives, statins and drugs for treatment of osteoporosis. The paucity of evidence available caused the GDG to agree to complete these reviews but not to look for evidence for other possible topics. A research recommendation for stopping drugs was however developed.

9.6.1 Review question: What is the clinical- and cost-effectiveness of stopping antihypertensive treatment?

For full details see review protocol in Appendix C.

Table 103: PICO characteristics of review question

Population	People taking antihypertensive drugs as primary or secondary prevention for at least 1 year
Intervention	Stopping anti-hypertension agents (thiazides, beta blockers, alpha blockers, calcium-channel blockers, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers)
Comparison	Continuing anti-hypertensive agents
Outcomes	**Critical:** • All-cause mortality • Cardiovascular mortality • Non-fatal myocardial infarction • Stroke • Quality of life (QoL) • Hospitalisation • Admission to care facility **Important:** • Blood pressure • Falls
Study design	Randomised clinical trials (RCTs); cohort studies if no RCTs retrieved

9.6.2 Clinical evidence

We searched for studies comparing the outcomes for people with multimorbidity who have stopped antihypertensive treatment versus people who continued on antihypertensive treatment.

Three randomised controlled trials (RCTs)[81,99,140] that evaluated the effect of stopping antihypertensive treatment for primary prevention were included. None of the studies reported the proportion of people in the sample who had multimorbidity. The included studies are summarised in Table 104 below. Evidence from these studies is summarised in the clinical evidence profile below (Table 105). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L. No studies evaluating the effect of stopping antihypertensive treatment for secondary prevention were identified.

Table 104: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments
Freis 1975[81]	Intervention (n=60): discontinuation of hypertension medication (hydrochlorothiazide, reserpine or hydralazine) and allocated to placebo Comparison (n=26): continuation of anti-hypertensives (hydrochlorothiazide, reserpine or hydralazine)	Adult (average age intervention 52.2 years, control 52.8 years) veterans hospitalised prior to treatment for hypertension, with normal blood pressure (average diastolic blood pressure <95mm Hg for 2 or more years) and on anti-hypertensives for primary prevention for 2 or more years Male to female ratio 1:0	Cardiovascular mortality (18 months) Non-fatal congestive heart failure (18 months) Atrial fibrillation (18 months) Right bundle block (18 months)	51 people (85%) in the intervention group were removed from the trial; 42 because of return to increased arterial pressures and 6 because of major cardiovascular complications

Study	Intervention and comparison	Population	Outcomes	Comments
		Multimorbidity: number of people with multimorbidity not reported USA	Return to hypertension (diastolic blood pressure ≥95 mm Hg) (14 months)	
Greenberg 1986[99]	Intervention (n=783): discontinuation of anti-hypertensives (bendrofluazide or propanololol) Comparison (n=837): continuation of anti-hypertensives (bendrofluazide or propanololol)	Adult (range 35-64 years), living in the community, with mild hypertension (diastolic blood pressure 90-109mm Hg) and on anti-hypertensives for primary prevention for 5.5 years Male to female ratio 1418:1347 Multimorbidity: number of people with multimorbidity not reported England	Maintained target blood pressure (diastolic blood pressure <90 mm Hg) (2 years)	396 people (24.4%) completed the 2 year follow up
Maland 1983[140]	Intervention (n=31): discontinuation of anti-hypertensives (chlorthalidone, hydrothiazide, triamterene) and allocated to placebo Comparison (n=31): continuation of anti-hypertensives (chlorthalidone, hydrothiazide, triamterene)	Adult (aged 30 years or over; mean age 60.3 years), living in the community, with mild hypertension (diastolic blood pressure average 90mm Hg or less for 1 year) and on anti-hypertensives for primary prevention for 1 or more years Male to female ratio 1:1 Multimorbidity: number of people with multimorbidity not reported USA	Cardiovascular mortality (1 year) Non-fatal myocardial infarction (1 year) Transient ischaemic attack (1 year) Return to hypertension (diastolic blood pressure >95 mm Hg) (1 year)	59 people (95.2%) completed the 1 year follow up

Table 105: Clinical evidence summary: Stopping versus continuing antihypertensive treatment

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with continuing antihypertensive treatment	Risk difference with stopping (95% CI)
Cardiovascular mortality	148 (2 studies) 13-18 months	VERY LOW[a,b,c] due to risk of bias, inconsistency, imprecision	OR 0.65 (0.04 to 11.68)[d]	16 per 1000	6 fewer per 1000 (from 17 fewer to 155 more)
Non-fatal myocardial infarction	62 (1 study) 1 years	VERY LOW[a,c] due to risk of bias, imprecision	OR 7.39 (0.15 to 372.38)[d]	0 per 1000	32 more per 1000 (from 53 fewer to 117 more)
Transient ischaemic attack	62 (1 study) 18 months	VERY LOW[a,c] due to risk of bias, imprecision	OR 0.14 (0 to 6.82)[d]	32 per 1000	28 fewer per 1000 (from 32 fewer to 153 more)
Non-fatal congestive heart failure	86 (1 study) 18 months	VERY LOW[a,c] due to risk of bias, , imprecision	OR 4.34 (0.36 to 52.52)[d]	0 per 1000	83 more per 1000 (from 5 fewer to 171 more)
Atrial fibrillation	86 (1 study) 18 months	VERY LOW[a,c] due to risk of bias, imprecision	OR 4.19 (0.06 to 299.15)[d]	0 per 1000	17 more per 1000 (from 47 fewer to 81 more)
Right bundle block	86 (1 study) 18 months	VERY LOW[a,c] due to risk of bias, imprecision	OR 4.19 (0.06 to 299.15)[d]	0 per 1000	17 more per 1000 (from 47 fewer to 81 more)
Return to hypertension (diastolic blood pressure ≥95 mm Hg)	146 (2 studies) 12-14 months	MODERATE[a] due to risk of bias	RR 7.66 (2.97 to 19.71)	71 per 1000	476 more per 1000 (from 141 more to 1000 more)
Maintained target blood pressure (diastolic blood pressure <90 mm Hg)	333 (1 study) 2 years	LOW[a,b] due to risk of bias, imprecision	RR 0.61 (0.5 to 0.76)	721 per 1000	281 fewer per 1000 (from 173 fewer to 360 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 or 2 increments because the point estimates varied widely across studies
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(d) Peto OR

9.6.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

9.6.4 Evidence statements

Clinical

- Very low quality evidence from 2 RCTs comprising of 148 participants demonstrated a clinical benefit of stopping anti-hypertensive medication compared to continuing with regards to cardiovascular mortality.
- Very low quality evidence from 1 RCT comprising of 62 participants demonstrated a clinical harm of stopping anti-hypertensive medication compared to continuing with regards to non-fatal myocardial infarction.
- Very low quality evidence from 1 RCT comprising of 62 participants demonstrated no clinical difference between stopping anti-hypertensive medication and continuing with regards to transient ischemic attack.
- Very low quality evidence from 1 RCT comprising of 86 participants demonstrated a clinical harm of stopping anti-hypertensive medication compared to continuing with regards to non-fatal congestive heart failure.
- Very low quality evidence from 1 RCT comprising of 86 participants demonstrated no clinical difference between stopping anti-hypertensive medication and continuing stopping anti-hypertensive medication compared to continuing with regards to atrial fibrillation.
- Very low quality evidence from 1 RCT comprising of 86 participants demonstrated no clinical difference between stopping anti-hypertensive medication and continuing anti-hypertensive medication compared to continuing with regards to right bundle block.
- Moderate quality evidence from 2 RCTs comprising of 146 participants demonstrated a clinical harm of stopping anti-hypertensive medication compared to continuing with regards to return to hypertension.
- Low quality evidence from 1 RCT comprising of 333 participants demonstrated a clinical harm of stopping anti-hypertensive medication compared to continuing with regards maintaining target blood pressure

Economic

- No relevant economic evaluations were identified.

9.6.5 Recommendations and link to evidence

Recommendations	No recommendations made.
Research recommendation	**2. What is the clinical and cost effectiveness of stopping preventive medicines in people with multimorbidity who may not benefit from continuing them?**
Relative values of different outcomes	The GDG identified all-cause mortality, cardiovascular mortality, non-fatal myocardial infarction, stroke, quality of life, hospitalisation and admission to care facility as critical outcomes in evaluating the clinical and cost-effectiveness of withdrawing anti-hypertensive medication. They also identified blood pressure and falls as important outcomes.
Trade off between clinical benefits and harms	The evidence indicated that there was a clinical benefit of stopping treatment compared to continuing antihypertensive treatment in terms of cardiovascular mortality (critical outcomes). The evidence also indicated that stopping treatment was associated with a clinical harm for return to hypertension (that is, a rise in blood pressure to above the threshold for diagnosing hypertension, which was an important outcome) and a clinical benefit for maintaining target blood pressure. However, the GDG noted that, within this analysis in people who were not excluded from the trial (for example due to high blood pressure, previous cardiovascular events), the majority of people who stopped taking antihypertensives did not return to hypertension. Therefore while the evidence indicates a significant harm of stopping for some people, there may be a significant proportion of people taking antihypertensives who may be able to stop taking them without returning to hypertension. The GDG believed that stopping antihypertensives may be of clinical benefit for reducing people's treatment burden and side effects, and for increasing quality of life; however, there was no evidence identified for these outcomes.
Economic considerations	No relevant economic evaluations were identified. Starting treatment with anti-hypertensives has already been judged to be clinically and cost-effective. There is a trade-off between the possible benefits of stopping treatment, as it could reduce treatment-related adverse events and cost of treatment, and the possible harm of stopping an effective treatment. The GDG considered the clinical evidence and found it inconclusive as some benefits were shown in people who stopped treatment (fewer cardiovascular mortality and transient ischaemic attack events) but a clinical harm was found in this group for return to hypertension. Considering the uncertainty in the clinical evidence, the GDG concluded there was high uncertainty on the cost effectiveness of stopping treatment too and they decided not to make a recommendation. A research recommendation was recommended in this area and details can be found in Appendix O.
Quality of evidence	The evidence was of very low quality. The included studies had small sample sizes with very low event rates. All of the evidence was at serious risk of bias due to selection bias and high rates of missing data. In addition to all of the studies being selective in their inclusion criteria to exclude high risk patients, they removed people after randomisation who were deemed high risk (for example when blood pressure rose above a certain level). None of the studies reported the proportion of people in the sample who had multimorbidity but studies were not downgraded for indirectness because evidence was deemed applicable to people with multimorbidity as the GDG believed that the clinical outcome of stopping antihypertensive treatment would be the same in people with and without multimorbidity. The majority of studies also showed very serious imprecision. The GDG considered that this evidence came from 2 studies that only included participants who have 'mild' hypertension (diastolic blood pressure levels

	below 90-95mm Hg for 1-2 year years) who did not have a history of major cardiovascular events (for example stroke, myocardial infarction, congestive heart failure, renal failure).The GDG thought that such 'low risk' people who stop antihypertensive medication may be at a reduced risk for returning to hypertension and for the associated harms (that is, mortality or cardiovascular events). The GDG noted that therefore the evidence was not applicable to a higher risk population. The GDG noted that all of the studies used anti-hypertensive drugs that were discontinued (for example, reserpine) or were not current standard practice, and that this may limit the applicability of the evidence as the GDG noted that these drugs were less effective than currents one and therefore there may be a greater benefit with continuing current medication. The GDG agreed that the evidence was sparse and of low quality and so was insufficient to make a specific recommendation on the stopping of anti-hypertensive drugs.
Other considerations	The GDG agreed that stopping anti-hypertensive medication would be suitable in some cases, for example, in 'low risk' people who have maintained blood pressure at normal levels for a long period of time or whose treated blood pressure has fallen (for example, because of lifestyle change or weight loss) and have no history of cardiovascular events, and in people with a high disease or treatment burden and limited life expectancy. The GDG noted that it is common practice to review medication in all people and to stop medication in some cases. The GDG felt that a regular review of medication of people with multimorbidity, where stopping medications such as antihypertensives is considered, would be beneficial. The GDG noted that changes in circumstances may prompt a medication review of people with multimorbidity. Lifestyle changes, such as changes in diet, or physiological changes, such as weight loss, may mean that a person's need for medication is different than when the medication was started. Additionally, the GDG noted that increases in the number or the severity of conditions, or increases in the number or intensity of treatment a person is taking, may lead to greater disease or treatment burden on the person. The GDG agreed that decisions about stopping treatment should be discussed with the person, and their carer where appropriate, and that they should be fully informed of the expected benefits and risks of medication withdrawal before making this decision. The GDG noted that decisions on stopping medication might be informed by information on absolute effects and on the timeframe expected to experience benefit (see sections 9.6 to 9.8). The GDG recognised the importance of regularly monitoring blood pressure in people who have stopped taking anti-hypertensive medication and felt that in some cases it may be appropriate to restart medication if hypertension returns.

9.7 Stopping drugs: treatments for osteoporosis

9.7.1 Review question: What are the effects of stopping common drug treatments (drugs for osteoporosis)?

For full details see review protocol in Appendix C.

Table 106: PICO characteristics of review question

Population	People taking drugs for osteoporosis for at least 1 year
Intervention	Stopping: Drugs affecting bone metabolism (a) Bisphosphonates: ○ Alendronate ○ Sodium clodronate ○ Etidronate ○ Risedronate ○ Ibandronate ○ Zoledronate ○ Pamidronate (b) Other drugs affecting bone metabolism used for treatment of osteoporosis: ○ Strontium ranelate ○ Denosumab Other drugs : Teriparatide
Comparison	Continuing drugs for osteoporosis
Outcomes	Critical: Health related quality of life Functional outcomes (for example, mobility, activities of daily living, FIM, or Barthel index, performance status) Fracture Falls Pain Hospitalisation Admission to care facility Important: GI bleed Atypical fracture Osteonecrosis jaw Discontinuation of medication due to side effects
Study design	Study designs: RCTs; Cohort studies if no RCTs Stratification ○ Bisphosphonates vs. other drugs affecting bone metabolism vs. other drugs ○ Primary prevention of fragility fracture vs. secondary prevention of fragility fracture

9.7.2 Clinical evidence

Five RCTs were included in the review[21-23,75,149,150]; these are summarised in Table 107 below. Evidence from these studies is summarised in the clinical evidence summary below (evidence summary).

See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

All studies evaluated the effect of stopping bisphosphonate treatment; no studies evaluated the effect of stopping other drugs for osteoporosis. The majority of studies either did not specify whether bisphosphonate treatment was being used for the primary or secondary prevention of fractures, or included patients using bisphosphonates for both. One study[150] evaluated the effect of stopping bisphosphonate treatment used for the secondary prevention of fractures, and was pooled with all other studies.

Table 107: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments
Black 2006[23] (Ensrud 2004[75])	Stopping bisphosphonate treatment (placebo); duration = 5 years Continuing bisphosphonate treatment (alendronate 5mg or 10mg/day); duration = 5 years	Postmenopausal women aged 55-81 years with low femoral neck BMD (<0.68 g/cm2) who had taken either 5mg or 10mg/day alendronate for 3 years. N = 1099	Vertebral fracture; Non-vertebral fracture; Morphometric vertebral fracture, hospitalisation, discontinuation of drug due to side effects	All participants advised to take daily supplement containing calcium (500mg) and vitamin D (250 IU)
Black 2012[21]	Stopping bisphosphonate treatment (placebo); duration = 3 years Continuing bisphosphonate treatment (Zolendronate 5mg intravenous infusion, annually); duration = 3 years	Osteoporotic women (mean age = 75.5 years, SD = 4.9) who had received annual intravenous zolendronate 5mg for 3 years. N = 1233	Fracture (any); vertebral fracture; non-vertebral fracture; morphometric vertebral fracture; atypical femur fracture; discontinuation of drug due to side effects	All participants received daily oral calcium (1000 to 1500 mg) and vitamin D (400 to 1200 IU)
Black 2015[22]	Stopping bisphosphonate treatment (placebo); duration = 3 years Continuing bisphosphonate treatment (Zolendronate 5mg intravenous infusion, annually); duration = 3 years	Osteoporotic women (mean age = 78 years, SD = 4.8) who had received annual intravenous zolendronate 5mg for 6 years. N = 190	Fracture (any); morphometric vertebral fracture	All participants received daily oral calcium (1000 to 1500 mg) and vitamin D (400 to 1200 IU) This is an extension of the Black 2012 trial[21], including only participants who had continued to receive bisphosphonate treatment for 6 years
Michalska 2006[149]	Stopping bisphosphonate treatment (placebo for 1 year then open label no treatment for 1 year); duration = 2 years	Ambulatory postmenopausal women, 50–80 years of age, who had taken alendronate (10 mg/d) for ≥3 years	Non-vertebral fracture; discontinuation of drug due to side effects	All patients received supplemental calcium (500 mg/d) and vitamin D (800 IU/d).

Study	Intervention and comparison	Population	Outcomes	Comments
	Continuing bisphosphonate treatment (Alendronate 10mg/day); duration = 2 years	N = 66		
Miller 1997[150]	Stopping bisphosphonate treatment (cyclical placebo daily for 14 days, followed by elemental calcium 500mg/day for 74 days); duration = 2 years Continuing bisphosphonate treatment (cyclical Etidronate treatment; 2mg phosphate for 3 days, followed by etidronate 400mg/day for 14 days, followed by elemental calcium 500mg/day for 74 days. Cycle repeated every 90 days); duration = 2 years	Women with post-menopausal osteoporosis (mean age = 70.4 years) who had experienced between 1-4 vertebral fractures and had received intermittent cyclical etidronate treatment for ≥ 1 year	Non-vertebral fracture; discontinuation of drug due to side effects	

Table 108: Clinical evidence summary: Stopping versus continuing bisphosphonate treatment

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Continuing bisphosphonates	Risk difference with Stopping (95% CI)
Clinical fracture (any) Time to any fracture	1145 (2 studies) 3 years	VERY LOW[b,c] due to risk of bias, imprecision	HR 0.95 (0.67 to 1.35)	Study population -[a]	-[a]
Clinical vertebral fracture Time to vertebral fracture	955 (1 study) 3 years	VERY LOW[b,c] due to risk of bias, imprecision	HR 0.55 (0.16 to 1.89)	Study population -[a]	-[a]
Clinical vertebral fracture	1099 (1 study) 5 years	MODERATE[c] due to imprecision	RR 2.22 (1.18 to 4.17)	Study population -[a]	-[a]
Clinical non-vertebral fracture Time to non-vertebral fracture	955 (1 study) 3 years	VERY LOW[b,c] due to risk of bias, imprecision	HR 1.01 (0.67 to 1.52)	Study population -[a]	-[a]
Clinical non-vertebral fracture	1331 (3 studies) 2-5 years	LOW[c,e] due to inconsistency, imprecision	RR 0.98 (0.76 to 1.27)	Study population -[a]	-[a]
Morphometric vertebral fracture	2244 (3 studies) 3-5 years	LOW[b,c] due to risk of bias, imprecision	OR 1.36 (0.97 to 1.91)	Study population -[a]	-[a]
Hospitalisation	1099 (1 study) 3 years	HIGH	RR 1.03 (0.85 to 1.25)	276 per 1000	8 more per 1000 (from 41 fewer to 69 more)
Atypical femur fracture	955 (1 study)	MODERATE[b]	See comment	0 per 1000	0 more per 1000

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Continuing bisphosphonates	Risk difference with Stopping (95% CI)
	3 years	due to risk of bias	4		(from 0 fewer to 0 more)
Discontinuation of study due to side effects	2587 (4 studies) 2-3 years	VERY LOW[c,e] due to inconsistency, imprecision	RR 0.96 (0.71 to 1.29)	67 per 1000	3 fewer per 1000 (from 19 fewer to 19 more)
1					

(a) Not calculated as (adjusted) raw data was not reported
(b) Downgraded once if the majority of evidence was at high risk of bias and twice if the majority of the evidence was at very high risk of bias
(c) Downgraded once if the CI crossed one MID and twice if the CI crossed two MIDs
(d) Not calculated as zero events in both groups
(e) Downgraded once if I2 >50% and/or there was serious variation in point estimates, and twice if I2 >75% and/or there was very serious variation in point estimates

9.7.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

9.7.4 Evidence statements

Clinical

- Very low quality evidence from 2 RCTs comprising of 1145 participants demonstrated no clinical difference between stopping and continuing bisphosphonate treatment with regards to time to clinical facture (any).
- Very low quality evidence from 1 RCT comprising of 955 participants demonstrated a clinical benefit of stopping bisphosphonate treatment compared to continuing bisphosphonate treatment with regards to time to clinical vertebral fracture.
- Moderate quality evidence from 1 RCT comprising of 1099 participants demonstrated a clinical harm of stopping bisphosphonate treatment compared to continuing bisphosphonate treatment with regards to clinical vertebral fracture.
- Very low quality evidence from 1 RCT comprising of 955 participants demonstrated no clinical difference between stopping and continuing bisphosphonate treatment with regards to time to clinical non-vertebral fracture.
- Low quality evidence from 3 RCTs comprising of 2244 participants demonstrated a clinical harm of stopping bisphosphonate treatment compared to continuing bisphosphonate treatment with regards to morphometric vertebral fracture.
- High quality evidence from 1 RCT comprising of 1099 participants demonstrated no clinical difference between stopping and continuing bisphosphonate treatment with regards to hospitalisation.
- Moderate quality evidence from 1 RCT comprising of 955 participants demonstrated no clinical difference between stopping and continuing bisphosphonate treatment with regards to atypical femur fracture.
- Very low quality evidence from 4 RCTs comprising of 2587 participants demonstrated no clinical difference between stopping and continuing bisphosphonate treatment with regards to discontinuation of study due to side effects.

Economic

- No relevant economic evaluations were identified.

9.7.5 Recommendations and link to evidence

Recommendations	**33.Tell a person who has been taking bisphosphonate for osteoporosis for at least 3 years that there is no consistent evidence of:** • **further benefit from continuing bisphosphonate for another3 years** • **harms from stopping bisphosphonate after 3 years of treatment.** **Discuss stopping bisphosphonate after 3 years and include patient choice, fracture risk and life expectancy in the discussion.**
Relative values of different outcomes	The GDG identified health related quality of life, functional outcomes, fracture, falls, pain, hospitalisation, and admission to care facility as critical outcomes for evaluating the effect of stopping drugs to treat osteoporosis. GI bleed, atypical fracture, osteonecrosis jaw, and discontinuation of medication due to side effects as important outcomes.
Trade-off between clinical benefits and harms	The evidence demonstrated a clinical benefit of stopping treatment for osteoporosis for time to clinical vertebral fracture compared to continuing treatment. There was a clinical harm of stopping treatment for osteoporosis for overall incidence of clinical vertebral fracture and morphometric vertebral fracture. There was no clinical difference between stopping and continuing treatment for osteoporosis for incidence of any clinical fracture, time to clinical non-vertebral fracture, incidence of clinical non-vertebral fracture, hospitalisation, incidence of atypical femur fracture, and discontinuation of study due to side effects.
Economic considerations	No relevant economic evaluations were identified on stopping treatment for osteoporosis. The GDG considered the trade-off between the cost of the treatment itself against the possible consequences of stopping treatment (occurrence of fractures and falls) in terms of both their costs (cost of managing fractures) and health burden. The GDG considered that stopping treatment may lead to improvements in health related quality of life and savings to the NHS as a result of fewer treatment-related adverse events and reduction in pill burden.
Quality of evidence	The evidence for stopping treatment for osteoporosis had the following quality ratings: time to any clinical fracture at very low quality due to risk of bias and imprecision; time to vertebral fracture at very low quality due to risk of bias and imprecision; incidence of clinical vertebral fracture at moderate quality due to imprecision; time to non-vertebral fracture at very low quality due to risk of bias and imprecision; incidence of clinical non-vertebral fracture at low quality due to inconsistency and imprecision; morphometric vertebral fracture at low quality due to risk of bias and imprecision; hospitalisation at high quality; atypical femur fracture at moderate quality due to risk of bias; discontinuation of study due to side effects at very low quality due to inconsistency and imprecision.
Other considerations	While the evidence indicated that there was no difference between stopping and continuing bisphosphonates for the outcomes of osteonecrosis or GI bleed, the GDG suggested that people at risk of these outcomes may have had treatment withdrawn within 3 years, and therefore may not be represented in these trials. The GDG believed that clinicians should instigate discussion of stopping bisphosphonate treatment, as some patients may be unlikely to suggest this. The GDG recognised that there are some people at particularly high risk of fracture where continuation of treatment may be benefical and so worded the recommendation to ensure this was considered in any discussion. The GDG noted that the evidence included in the review evaluated the impact of stopping bisphosphonate treatment for up to 3 years, and demonstrated no consistent evidence of harm in stopping treatment during this time. Consequently, the GDG did not believe that there was a need to routinely review this decision within this time period. However, the GDG agreed that clinicians will wish to review the decision to stop treatment if the person's circumstances changes; for example if

	a clinician believes that a person's risk of fracture has increased.

9.8 Stopping drugs: statins

9.8.1 Review question: What is the clinical and cost effectiveness of stopping statin treatment?

For full details see review protocol in Appendix C.

Table 109: PICO characteristics of review question

Population	People taking statins as primary or secondary prevention for at least1 year
Intervention(s)	Stopping statins (all)
Comparison(s)	Continuing statins
Outcomes	**Critical:** Quality of life Hospitalisation All-cause mortality Cardiovascular (CV) mortality Stroke Non-fatal myocardial infarction (MI) Institutionalisation **Important:** Myalgia
Study design	RCTs, Cohort studies if RCTs not retrieved (confounders: MM, age, reason for stopping)

9.8.2 Clinical evidence

We searched for studies comparing outcomes for stopping statin treatment versus continuing statin treatment in people who had been taking statins for 1 year or more for either primary or secondary prevention of cardiovascular events. Our objective was to assess the clinical and cost impact of patients stopping long term statin treatment. We sought evidence for all populations older than 18, regardless of multimorbidity status, as the GDG felt that findings in a general population could be generalised to a population of individuals with multimorbidity. We pooled evidence from all different statin treatments, as the GDG felt there was unlikely to be a difference between different statins in the impact of stopping.

One RCT evaluating the effect of stopping statin treatment was included.[128] This study is summarised in Table 110 below. Evidence from this study is summarised in the clinical evidence summary below (111). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Table 110: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments
Kutner 2015[128]	(n=189) Stop statins. Discontinued statins	Mean age 74.1 years (SD 11.6)	All-cause mortality (time to event) at	Patients were taking statins for either

Study	Intervention and comparison	Population	Outcomes	Comments
	at time of randomisation. (n=192) Continue statins. No change to statin therapy.	Included: adults aged >18 years, receiving a statin for 3 months or longer for primary or secondary prevention of cardiovascular disease, diagnosis of "advanced, life-limiting illness", predicted life expectancy between 1 month and 1 year Excluded: Physician opinion that the patient had active CVD requiring ongoing therapy with statin medications, symptoms of myositis/deranged LFTs or other contraindications to continuing statin therapy USA	end of follow-up (median 18 weeks, IQR 8-36 weeks) Cardiovascular-related events at end of follow-up (median 18 weeks , IQR 8-36 weeks) MacGill Quality of life as assessed as mean score across multiple time-points between 0 and 20 weeks	primary or secondary prevention (58% of patients had a history of cardiovascular disease) RCT without blinding Patients were on statins for at least 3 months prior to trial (1.6% on statins for <1 year, 26.5% on statins for 1-5 years, 69% on statins for>5 years, 2.9% unknown)

Table 111: Clinical evidence summary – Stopping statins versus continuing statins

Outcomes	Number of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with continue statins	Risk difference with stop statins (95% CI)
Quality of life – Total (assessed as the mean quality of life across multiple time-points between 0 and 20 weeks (baseline, 4, 8, 12, & 20 weeks; AUC); MacGill, 0-10, higher indicates a better income[c]	381 (1 study) Up to 20 weeks	LOW[a,b] due to risk of bias, imprecision	-	The mean quality of life (AUC) between 0 and 20 weeks in the group continuing statins was 6.85	The mean quality of life (AUC) between 0 and 20 weeks in the group stopping statins was 0.26 higher (0.02 to 0.50 higher)
All-cause mortality (time to event)	381 (1 study) median follow-up 18 weeks, IQR 8-36 weeks	VERY LOW[a,b] due to risk of bias, imprecision	HR 0.95[d] (0.7 to 1.28)	510 per 1000	18 fewer per 1000 (from 117 fewer to 89 more)
New cardiovascular event/invasive procedure with hospital/emergency department admission	381 (1 study) median follow-up 18 weeks	VERY LOW[a,b] due to risk of bias, imprecision	RR 1.23 (0.56 to 2.67)	58 per 1000	13 more per 1000 (from 26 fewer to 97 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
(c) Availability of data diminished over course of 20 weeks, mean is across all 4 time-points where data was complete for each person
(d) Hazard ratio estimated from Kaplan-Meier curve and follow-up times

9.8.3 Economic evidence

Published literature

No relevant economic evaluations were identified.

See also the economic article selection flow chart in Appendix F.

9.8.4 Evidence statements

Clinical

- Low quality evidence from 1 RCT comprising 381 patients demonstrated no clinical difference between stopping statins and continuing statins with regards to mean total scores on the MacGill Quality of life scale between 0 and 20 weeks. The evidence was at serious risk of bias and demonstrated serious imprecision.
- Very low quality evidence from 1 RCT comprising 381 patients demonstrated a clinical benefit of stopping statins compared to continuing statins with regards to all-cause mortality (time to event). The evidence was at serious risk of bias and demonstrated very serious imprecision.
- Very low quality evidence from 1 RCT comprising 381 patients demonstrated no clinical difference between stopping statins and continuing statins with regards to cardiovascular-related events. The evidence was at serious risk of bias and demonstrated very serious imprecision.

Economic

- No relevant economic evaluations were identified.

9.8.5 Recommendations and link to evidence

Recommendations	**No recommendation**
Research recommendation	**What is the clinical and cost effectiveness of stopping preventive medicines in people with multimorbidity who may not benefit from continuing them?**
Relative values of different outcomes	The GDG identified all-cause mortality, CV mortality, non-fatal MI, stroke, quality of life, hospitalisation and admission to care facility as critical outcomes for evaluating the effect of stopping statins. In addition, the GDG identified myalgia as an important outcome.
Trade off between clinical benefits and harms	The evidence demonstrated a clinical benefit of stopping statins for all-cause mortality compared to continuing statins in people who had been taking statins for a median of 18 weeks (IQR 8-36 weeks) for either primary or secondary prevention of cardiovascular events. There was no clinical difference between stopping and continuing statins for quality of life or cardiovascular related events. No evidence was identified to evaluate the effect of stopping on statins on cardiovascular mortality, non-fatal MI, stroke, hospitalisation, admission to care facility, and myalgia.
Economic considerations	No relevant economic evaluations were identified on stopping treatment with statins. The GDG considered the trade-off between the cost of the treatment itself against the possible consequences of stopping treatment (occurrence of cardiovascular events) in terms of both their costs (cost of managing cardiovascular events) and health burden. The GDG considered that stopping treatment may lead to improvements in health related quality of life and savings to the NHS as a result of fewer treatment-related adverse events and

	reduction in pill burden. However no clinical evidence has been identified to support this. The GDG concluded that for patients in whom the risk of cardiovascular events over their lifetime is low, stopping treatment with statins would lead to lower costs with no detriment in their quality of life, and therefore this option should be discussed.
Quality of evidence	All of the evidence in this review was identified from a single randomised-controlled trial. The evidence for stopping statins had the following quality ratings: all-cause mortality at very low quality due to risk of bias and imprecision; quality of life at low quality due to risk of bias and imprecision; cardiovascular related events at very low quality due to risk of bias and imprecision. The study used a relatively a short average follow-up (median 18 weeks, IQR 8-36 weeks). The GDG concluded that this may be an insufficient timeframe to identify the longer term clinical benefits and harms of stopping statins.
Other considerations	All evidence came from 1 study with a mixed primary and secondary prevention population. The GDG noted that stopping statins may have different effects on these groups. The GDG expected that the risks of withdrawing statin treatment may be greater in people who have previously experienced a cardiovascular event, and further research is needed to evaluate the effect of stopping statins in this group separately. The study population was defined as having a limited life expectancy of less than 1 year and people were excluded from the study if a physician was of the opinion that they had active cardiovascular disease requiring ongoing therapy with statin medications. The GDG noted the study did not clearly define "active cardiovascular disease" and that this may be difficult to do in clinical practice. However the GDG felt that this was an appropriate population in which there may be a benefit of stopping statins. Clinicians should use their judgement when discussing stopping statin treatment, taking into account the nature and severity of a person's cardiovascular disease. The GDG chose to develop a research recommendation for this area and details on this can be found in Appendix O.

9.9 Developing an individualised management plan

Recommendations and link to evidence

Recommendations	**34. After a discussion of disease and treatment burden and the person's personal goals, values and priorities, develop and agree an individualised management plan with the person. Agree what will be recorded and what actions will be taken. These could include:** • **starting, stopping or changing medicines and non-pharmacological treatments** • **prioritising healthcare appointments** • **anticipating possible changes to health and wellbeing** • **assigning responsibility for coordination of care and ensuring this is communicated to other healthcare professionals and services** • **other areas the person considers important to them**

	• **arranging a follow-up and review of decisions made.** **Share copies of the management plan in an accessible format with the person and (with the person's permission) other people involved in care (including healthcare professionals, a partner, family members and/or carers).**
Relative values of different outcomes	The GDG considered that the review on barriers to optimising care and the review on principles for treating people with multimorbidity established that both practitioners and patients understanding of conditions and treatments can be suboptimal and that an important outcome from any approach to people with multimorbidity should involve increased clarity around decisions made about optimising treatments and how those conditions will be communicated and who will take responsibility to co-ordinate care.
Trade off between clinical benefits and harms	The GDG did not consider there were likely to be any harms from agreeing an plan with each patient
Economic considerations	The GDG did not consider that agreeing a plan with the person with multimorbidity would result in additional resource costs. Review of medicines and treatments is considered a core part of the delivery of medical care and already part of the role of healthcare practitioners. It is likely that the discussions involved will be spread over several consultations and the GDG considered that the delivery of an approach to care that takes account of multimorbidity could be carried out as part of usual medical practice when providing and reviewing care. Reviewing medicines prescribed may generate some cost savings if unnecessary treatments are discontinued.
Quality of evidence	The recommendation was informed by qualitative reviews reported in chapter 6 where the quality of evidence is discussed.
Other considerations	The GDG agreed this recommendation to ensure there is clarity about what is involved in agreeing a plan with person who might benefit from a multimorbidity approach to their care. Any decisions should build on previous steps where people's possible treatment burden is explored, their preferences and priorities are elicited and their treatments, including healthcare appointments are discussed in the context of these. The GDG were clear that the outcomes may include stopping or changing medicines and non-pharmacological treatments, decisions about prioritising some healthcare appointments and not others. One of the more difficult tasks, but an essential one can be agreeing responsibility for coordination of care and ensuring this is communicated to other healthcare professionals and services. In most cases this is likely to be the GP taking this role but methods of communication may need to be developed to ensure this is facilitated in an increasingly complex health service. One of the ways of doing this might be to provide the person with copies of any specific management plan that is made although the GDG were sceptical about the value plans generated from computer systems at present. Appropriate follow up to review any decisions made is important. The GDG chose specifically to call the plan an individualised management plan and understood that the emphasis of the plan is on clinical care of the patient. They were aware of inconsistency generally in language around plans but agreed that 'care' plans have a specific meaning in social services and that care plans may include a wider range of issues than the intention with this plan which is concerned with decisions around clinical management and particularly reduction in treatment burden forpeople with multimorbidity.

10 Interventions to improve care for people with multimorbidity

Introduction

Modern medical care has become increasingly specialised. Many people are seen in tertiary centres for highly specialised care where one aspect of their condition is reviewed in isolation from other aspects of their health or other circumstances. At the same time services such as primary care continue with a model of care where contact with a practitioner is often reactive to patient request for appointment and appointments are short and not readily available.

The scope for the guideline therefore included a number of possible interventions that might be considered to improve care for people with multimorbidity. Self –management programmes and formats of encounters are discussed in chapters 13 and 14 respectively.

This chapter includes reviews of interventions that we have called 'models of care'. The terminology in this area is confusing and overlapping and interventions overlap in their components or use the same term to describe different interventions. To make sense of the evidence we have not used the terms in the papers but have extracted the descriptions of the interventions and provided as much detail as possible about these. The GDG used consensus to decide on the terms to use in the review and the components are described in section 12.1.2.The second part of this chapter discusses holistic assessment programmes developed around the model of Comprehensive Geriatric Assessment (CGA). The reason that this is presented separately is explained in section 12.2. CGA and similar models do overlap with interventions included in section 12.1 and the separate presentation is for ease of presentation and analysis.

10.1 Models of care

10.1.1 Review question: What models of care improve outcomes in people with multimorbidity?

For full details see the review protocol in Appendix C.

Table 112: PICO characteristics of review question

Population	Adults (18 years and over) with multimorbidity
Interventions	Interventions targeted at improving outcomes and continuity of care for people with multimorbidity. Examples may include: • Collaborative care • Integrated care • Case management • Provider continuity • Care plan • Patient held records • Multi-professional working • Interventions to improve continuity of information • Medication management • A combination of above
Comparison	Standard care

Outcomes	**Critical** • Health-related quality of life • Mortality • Functional outcomes • Patient and carer satisfaction • Length of hospital stay • Unscheduled care • Admission to care facility **Important** • Continuity of care • Patient/carer treatment burden
Study design	Randomised controlled trials (RCTs)

We sought studies evaluating interventions aimed at improving outcomes for people with multimorbidity. Studies were included if the intervention was delivered to people with multimorbidity and the intervention targeted more than 1 of the person's health conditions. As a consequence, interventions targeted at improving patient outcomes for a single condition amongst people with multimorbidity were excluded.

10.1.2 Clinical evidence

Models of care (including and not including a self-management component):

Twenty randomised clinical trials reported in 28 publications were included in the review[5,16,17,19,29-32,34,41,45,51,72,73,91-93,111,122,132,138,145,148,176,209,218,220,237]. Evidence from these studies is summarised below (Tables 113-134). Please see also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

One additional study was identified as relevant for the review; however, the trial methodology was assessed as being at an unacceptably high risk of bias and so it was excluded[133]. The risk of bias was due to an "attempt to manipulate the assignment to groups by Customer Centre Representatives" as reported by the authors. This led to significant differences between the 2 groups at baseline. Although the authors performed an adjusted analysis, this analysis did not include certain characteristics that were deemed obligate in order to accept the outcomes as sufficiently free of bias.

The GDG chose to stratify papers into those evaluating a model of care and those evaluating a model of care that contained a self-management component. This review is presented as follows; evidence from papers evaluating models of care for people with multimorbidity is first presented followed by evidence from papers evaluating models of care with a self-management component. Please note that papers evaluating models of care with a self-management component were not included in the self-management review as the self-management component was felt to be the smaller component of the overall intervention.

Of the 20 models of care studies: 10 were carried out in the USA or Canada; 8 were carried out in Europe; 1 was carried out in Australia and 1 was carried out in Hong Kong.

Analysis

The studies included in the review evaluated the efficacy of complex interventions aimed at improving outcomes for people with multimorbidity. Studies contained multiple components and varied in terms of their duration (they ranged from 30 days to 6.2 years), comparator (usual care and enhanced usual care), and population (multimorbid and older adults). Due to the complexity of

interventions and the lack of commonality across studies, the GDG chose not to pool data from different studies, and evidence from these studies is presented individually.

There were a variety of outcomes reported by the studies. Where the studies did not report outcomes specified in the protocol, we included data from closely related outcomes (for example, self-related health in place of health-related quality of life).

Components of models of care

Studies included in the review evaluated complex interventions, frequently comprising of multiple components. To summarise table 113 below, of the included studies (n=20): 7 featured multidisciplinary care, 11 featured holistic assessment (this component is discussed in more detail in the holistic assessment review section 12.2), 11 featured a care plan, 10 involved care co-ordination, 11 involved telephone follow-up, 11 involved home follow-up, 7 involved ways of promoting self-management and 3 involved some form of medication management.

The following list details models of care (or components of models of care) that were either specified in the review protocol or were aspects of interventions trialled in the included studies of this review. Many of these terms do not have a fixed definition, and the same term may be interpreted differently by different research groups and health care professionals. As a consequence, 2 studies that use the same term may be evaluating interventions that vary in content.

For the purposes of this review, the GDG used consensus to agree the definitions below. These definitions were used to identify the key components of each of the trialled interventions included in the review. This means that descriptions of interventions in this review may vary from the terms used in the published papers.

Multidisciplinary care

Multidisciplinary care is when professionals from a range of disciplines work together to deliver comprehensive care that addresses as many of the patient's needs as possible. This can be delivered by a range of professionals functioning as a team under 1 organisational umbrella or by professionals from a range of organisations, including private practice, brought together as a unique team. As a person's condition changes over time, the composition of the team may change to reflect the changing clinical and psychosocial needs of the person. The size and composition of teams varied considerably between interventions.

Holistic assessment

A trained healthcare professional performs a comprehensive assessment of a person’s physical health, mental health, social situation and functional ability. This assessment is used to generate an individualised treatment plan which may feed into subsequent care through multiple channels (for example, discussion at MDT or list of recommendations for GP).

Self-management

An intervention aimed at increasing a person’s ability to manage their own condition without the need for intensive support from HCPs. This may take many forms including but not limited to education about symptom control, supporting patient or carer identification of exacerbations of chronic conditions and provision of home rescue medication or treatment options that the patient can initiate themselves at appropriate times.

Care plan

A care plan is an agreement between patient and health or social care professional to support management of day to day health and symptoms by the patient and other healthcare professionals or to organise care. It can be a written document or something recorded in patient notes.

Care co-ordination

Either 1 individual (case management by a key worker) or 1 organisation (care management) takes the lead in organising a person's care and support. This may include liaising with other healthcare professionals (for example, GP or specialist services) and, in general, focuses on continuity of care. The key worker or organisation may not necessarily be responsible for delivering any additional intervention.

Telephone follow-up

In this review, telephone-follow up is defined as a pre-arranged telephone call to a patient (or carer) by a healthcare professional. This may be a one-off or multiple telephone calls depending on the aim of follow-up. The aim of telephone support will be specific to the person's needs and context. For example, telephone follow-up may be used to review progress of a patient, to anticipate problems, provide support or advice, or in some cases can be used to deliver an intervention

Home follow-up

As with telephone follow-up, home follow-up is defined here as pre-arranged visits by a healthcare professional to a patient's (or carer's) home, either as a one-off or on multiple occasions. The aim of follow-up will vary according to the person's needs and context.

Medication management

A healthcare professional works collaboratively with the patient, and if necessary other members of the MDT, to optimise safe, effective and appropriate drug therapy. This may involve the healthcare professional checking patient's medicine-taking behaviour, concerns about side effects and reviewing the indications for medicines.

Collaborative care

A complex intervention with 4 key components:

1. A multidisciplinary approach to patient care (including the use of non-medical case-managers)
2. Structured patient care plans
3. Scheduled follow-ups
4. Enhanced inter-professional communication

Integrated care

Integrated care is an integration of medical and social services in a continuum of care with case management programmes. Monitor describes integrated care as person-centred and co-ordinated care within healthcare settings, across mental and physical health and across health and social care. For care to be integrated, organisations and care professionals need to bring together all of the different elements of care that a person needs. Integration can be within a single physical co-location or may be more virtual on an organisational level.

Stepped care

Stepped care provides a framework in which to organise the provision of services supporting patients, carers and healthcare professionals in identifying and accessing the most effective interventions. Stepped care is a system for delivering and monitoring treatment with the explicit aim of providing the most effective, yet least burdensome, treatment to a person first, and which has a self-correcting mechanism built in so if a person does not benefit from an initial intervention they are 'stepped up' to a more complex intervention. Typically, stepped care starts by providing low-intensity interventions. In some stepped-care systems, low-intensity care is received by all individuals, although in other systems patients are stepped up to a higher intensity intervention on immediate contact with the service, (for example, if they are acutely unwell or acutely suicidal; this entry at different levels in relation to risk is also sometimes referred to as 'stratified care').

Table 113: Key features of included Models of Care studies

	Multidisciplinary care	Holistic assessment	Care plan	Care coordination	Telephone follow-up	Home follow-up	Self-management	Medication management
Models of care								
Alkema 20071		✓		✓	✓			
Beck 1997[16]	✓							
Berglund 2015[19]	✓	✓	✓	✓	✓			
Bouman 2008[32]		✓	✓		✓	✓		
Courtney 200926		✓	✓		✓	✓		
Eklund 2013[72]	✓	✓	✓	✓		✓		
Ell 201033				✓	✓			
Hogg 200947	✓				✓	✓		✓
Metzelthin 2013[148]		✓	✓	✓		✓		
Naylor 200478				✓		✓		
Sandberg 2015[209]		✓	✓	✓		✓		
Slaets 199787	✓	✓	✓	✓				
Sommers 2000	✓				✓	✓		
Models of care with a self-management component								
Boult 200811		✓	✓	✓			✓	
Behm 2014[17]b	✓					✓	✓	
Chow 2014[41]a		✓			✓	✓	✓	
Coburn 201223		✓	✓	✓	✓		✓	
Gitlin 200638			✓		✓	✓	✓	
Katon 201051			✓		✓		✓	✓
Legrain201161							✓	✓

(a) Study contained two intervention arms, one arm involved mostly telephone follow-up and one arm involved mostly home follow-up.
(b) Study contained two intervention arms, both arms involved home visits and self-management, only one arm involved multidisciplinary care

Table 114: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
Models of care without a self-management component					
Alkema 2007	Intervention (n=377): The Care Advocate Program (CA Program) bridged medical and social care delivery systems using telephone-based care management to coordinate health and long-term care services for chronically ill older adults. Participants received a call within 1 week of assessment and monthly follow-up calls during the 12 month intervention period to monitor progress. Control (n=404): received usual care from the health plan, which included medical group case management services designed to triage and address members' health- related issues, and facilitate access to insured health plan services (for example, insured durable medical equipment).	Adults (aged 65 years or over; mean intervention 82.98 years (SD 7.12), mean control 83.66 years (SD 7.36). Male to female ratio 35:65 Community Multimorbidity: number of participants with multimorbidity not reported USA	Mortality - during total study length (24 months).	Key features: holistic assessment, care co-ordination, telephone follow-up	Nursing home residents and those enrolled in similar studies were excluded.
Beck 1997	Intervention (n=160): Participants were invited to monthly group visits at the Cooperative Healthcare Clinic. Group visits involved a 30 minute talk by a member of the MDT on a relevant topic, breaks in which nurses took blood pressures and doctors circulated addressing individual concerns of participants and 30 minutes set aside at the end of the talk for participants to get	Adults (aged 65 years or over; mean intervention 72, mean control 75) Male to female ratio 31:69 Community Multimorbidity: number of	Mortality (12 months) Unscheduled care – urgent care visits per participant (12 months) Admission to care facility – proportion of participants hospitalised (12 months)	Key features: multidisciplinary care in (group visits)	None specified

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	one-to-one visits with the physician. Duration 12 months. Control (n=161) Standard care. Nil. Duration 12 months.	participants with multimorbidity not reported USA			
Berglund 2015	Intervention (n=85) Nurse with geriatric expertise made assessment of health/social care need at ED, assessment transferred to ward if participant transferred to ward, also sent to municipal MDT (nurse, social worker, physiotherapist, OT), case manager co-ordinated planning for discharge, case manager contacted relatives to offer support and advice, care-planning meeting after discharge organised in participant's own home with MDT, within 1 week after care-planning meeting older person contacted by case manager and plan for follow-up made, after 6 months a new care-planning meeting could be held if needed. Duration 12 months. Control (n=76) Usual care - some discharge planning in hospital, no meeting or proactive contact after discharge. Duration 12 months.	Adults (aged 65 years or over; mean ages not reported) Male to female ratio 72:89 Inpatients (prior to discharge) Multimorbidity: number of participants with multimorbidity not reported Sweden	Mortality (12 months)	Key features: multidisciplinary care, holistic assessment, care plan, care co-ordination, telephone follow-up	Severe acute illness, dementia, severe cognitive impairment, palliative care
Bouman 2008	Intervention (n=160) Program of eight home visits, with telephone follow-up over 18 month period, visited by trained	Adults (aged 70-84 years; mean 76, SD 3.7)	Mortality (24 months) Length of hospital stay –	Key features: holistic assessment,	Participants who self-rated health status as “moderate or good”, receiving home

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	home nurses, visits included multidimensional geriatric assessment with advice and referral to professional and community services. Differentiated from other CGA studies as each patient had formulaic pattern of follow-up as opposed to individualised treatment plan on back of CGA. Duration 18 months. Control (n=170) Usual care, participants could apply for all available care but no structured follow-up. Duration 18 months.	Male to female ratio 40:60 Community Multimorbidity: number of participants with multimorbidity not reported Netherlands	bed days per patient (24 months) Unscheduled care – hospital admissions (24 months) Admission to care facility – nursing home admissions (24 months)	telephone follow-up, home follow-up	nursing care, on waiting list for care home admission
Courtney 2009	Intervention (n=64): within 72 hours of admission a registered nurse and physiotherapist undertook a comprehensive patient assessment and developed a goal-directed, individualised care plan in consultation with the patient, health professionals, family and caregivers. Plan included: an individually tailored exercise program; nurse home visits; and telephone follow-up. Control (n=64): standard care, discharge planning and rehabilitation advice normally provided.	Adults (aged 65 years or over; mean age 78.8 years, SD 6.9) Male to female ratio 38:62 Inpatient Multimorbidity: number of participants with multimorbidity not reported; median number of conditions 5 (range 0-12). Australia	Health-related quality of life – SF-12 (physical component) (6 months). Health-related quality of life – SF-12 (mental component) (6 months). Unscheduled care – emergency hospital readmissions (6 months). Unscheduled care – emergency GP visits (6 months).	Key features: holistic assessment, care plan, telephone follow-up, home follow-up	Factors that would undermine patients' ability to participate in the intervention: patients requiring home oxygen, patients unable to walk independently for 3 metres (with/without walking aids), patients with neurological or cognitive deficit or disease.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
Eklund 2013	Intervention (n=89): Collaboration between a nurse with geriatric competence at the emergency department, the hospital wards and a multi-professional team in the community. Participants underwent geriatric assessment by nurse with geriatric competence, during admission followed by care co-ordination, care-planning and home follow-up. Focus of intervention was on creating a continuum of care. Control (n=76): Usual care including care planning following a routine assessment by community team following discharge, rehabilitation if needed following assessment.	Adults (aged 80 or older or 65-79 with at least one chronic disease and dependent in at least one ADL, mean not reported) Male to female ratio 45:55 Community (identified when presenting at ED) Multimorbidity: number of participants with multimorbidity not reported Sweden	Functional outcomes – improvement in ADL (12 months) Functional outcomes – worsening in ADL (12 months)	Key features: multidisciplinary care, holistic assessment, care planning, care co-ordination	Acute severe illness, dementia, palliative care
Ell 2010	Intervention (n=193): problem solving therapy and/or antidepressant medication based on a stepped-care algorithm; first-line treatment choice; telephone treatment response; adherence; and relapse prevention follow-up. Control (n=194): standard clinic care plus patient receipt of depression educational	Adults (aged 50 or older, mean intervention group age 75.1, mean control group age 69.1) Male to female ratio 20:80 Community	Health-related quality of life – SF12 mental component (12 and 18 months). Health-related quality of life – SF12 physical component (12 and 18 months).	Key features: care co-ordination, telephone follow-up	Acute suicidal ideation, score of ≥8 on the Alcohol Use Disorders Test alcohol assessment, recent lithium/antipsychotic medication use, inability to speak English or Spanish.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	pamphlets and a community resource list.	Multimorbidity: comorbid depression and diabetes USA	Functional outcomes - Sheehan Disability Scale of functional impairment (12 months and 18 months).		
Hogg 2009	Intervention (n=120): Anticipatory and Preventative Team Care (APTCare) Intervention: home-based multidisciplinary team management with an initial assessment by a nurse practitioner and a medication review by a pharmacist and individualised patient care plan. Control (n=121): patients received usual care from their family physicians.	Adults (mean intervention group age 69.6, mean control group age 72.8) Male to female ratio 103:138 Community Multimorbidity: number of participants with multimorbidity not reported; mean number of chronic conditions: intervention 2.7, control 2.3. Canada	Health-related quality of life - SF36 mental component (15 months). Health-related quality of life - SF36 physical component (15 months). Health-related quality of life - total number of unhealthy days in last 30 days (15 months). Mortality (15 months). Unscheduled care - average number of ED visits (15 months). Unscheduled care - average number of hospital admissions (15 months).	Key features: multidisciplinary care, care plan, telephone follow-up, home follow-up, medication management	Substantial cognitive impairment, language or cultural barriers, life expectancy less than 6 months, and plans to move or to be away for more than 6 weeks during the study period.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
			Caregiver burden (15 months).		
Metzelthin 2013	Intervention (n=193): People received an in home multidimensional assessment by a practice nurse, GP and practice nurse discussed the assessment and the need for other assessments, preliminary treatment plan formulated by GP and practice nurse with or without an MDT meeting, second home visit by practice nurse to formulate final treatment plan with person, practice nurse also acts as case manager to regularly review achievement of goals and need for additional support Control (n=153): Usual care, no further details provided	Age (aged 70 or older, mean intervention group age 77.49 (SD 5.8), mean control group age 76.8 (SD 4.92)) Male to female ratio 42:58 Community Multimorbidity: number of participants with multimorbidity not reported	Functional outcome – (GARS ADL subscale, 24 months) Functional outcome (GARS IADL subscale, 24 months)	Key features: holistic assessment, care plan, home follow-up, care co-ordination	Terminally ill, severe cognitive or psychological impairment, unable to communicate in Dutch
Naylor 2004	Intervention (n=118): collaboration with patients' physicians, 3 advanced practice nurses implemented an intervention extending from index hospital admission through 3 months after the index hospital discharge. Control (n=121): patients received care routine for the admitting hospital, including site-specific heart failure patient management and discharge	Adults (aged 65 or older, mean intervention group age 76.4 (SD 6.9), mean control group age 75.6 (SD 6.5)) Male to female ratio 102:147 Patients identified as inpatients, intervention panned discharge to	Quality of life - Minnesota Living with Heart Failure Questionnaire (total score) (12 months). Mortality (12 months). Functional outcome - Functional Status Score (12 months).	Key features: care co-ordination, home follow-up	Elders with end-stage renal disease were excluded because of their access to unique Medicare services.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	planning critical paths and, if referred, standard home agency care consisting of comprehensive skilled home health services.	community Multimorbidity: number of participants with multimorbidity not reported; mean number of conditions: intervention 6.4 (SD 2.5), control, 6.4 (SD 2.0). USA	Patient and carer satisfaction - patient satisfaction (6 weeks).		
Sandberg 2015	Intervention (n=80) Case management. Patients received traditional case management with assessment, co-ordination, home visits and telephone calls. Patients also received general information about the healthcare system and specific information about their needs. Case managers either had nursing or physiotherapy backgrounds. Monthly visits (over 12 months) took place in the patient's own homes. Each visit lasted ~1 hour and the contents of the visits depended on the individual's care plan. The first visit involved a CGA to inform a care plan to be used for subsequent visits. Duration 12 months. Control (n=73) Usual care. Duration 12 months.	Adults (aged 65 years or over; mean intervention 81.4, mean control 81.6) Male to female ratio 51:102 Community Multimorbidity – all patients had at least 2 "health complaints" Sweden	Mortality (12 months) Length of hospital stay – total length of inpatient stays (12 months) Unscheduled care – hospital admissions per patient (12 months)	Key features: holistic assessment, care co-ordination, home follow-up	Not able to communicate verbally, cognitive impairment, special accommodation

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
Slaets 1997	Intervention (n=140): psychogeriatric intervention, consisting of multidisciplinary joint treatment by a psychogeriatric team (a geriatrician, a specialised geriatric liaison nurse, and a physiotherapist). Weekly multidisciplinary meeting were held, attended by the geriatric team, the nurses, social worker, dietician, and psychiatrist. The geriatrician was present at the weekly ward rounds with the attending physician and the 2 resident physicians. In addition, the geriatric team had their own ward rounds every week. Control (n=97): Usual care consisted of services provided by physicians and nurses in another general medical unit in the same hospital.	Adults (aged 75 years or over; range 75-96, mean 82.8, SD 5) Male to female ratio 30:70 Inpatient Multimorbidity: number of participants with multimorbidity not reported Netherlands	Mortality (unclear time point). Functional outcomes - ADL (unclear time point). Functional outcome - mobility (unclear time point). Length of hospital stay (unclear time point). Admission to care facility (unclear time point).	Key features: multidisciplinary care, holistic assessment, care plan, care co-ordination	Patients admitted for day treatment were excluded.
Sommers 2000	Intervention (n=383): Senior Care Connections (SCC) intervention required collaboration among a primary care physician, nurse with geriatrics training, and a clinical social-worker. Home visit assessment followed by team discussion and development of a risk reduction plan and treatment targets. Throughout the intervention, the team met with trainers to learn team building skills and strategies for coaching patients in chronic disease self-management. The SCC intervention focused on a set of	Adults (mean intervention group age 77 (SD 6.6), mean control group age 78 (SD 6.8)) Male to female ratio 33:67 Living in the community, with difficulties living independently Multimorbidity: 2 or more	Mortality (24 months) Unscheduled care – hospital admissions per year (24 months).	Key features: multidisciplinary care, care plan, telephone follow-up, home follow-up	Not terminally ill, not residing in a nursing home, not under therapy for metastatic disease, Alzheimer disease, or related dementias.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	defined activities for each intervention patient. The nurse or social worker visited the patient in the home. A risk reduction plan was discussed with the patient and his/her family to set target objectives and plan treatment by means of chronic disease self-management strategies. Nurse/social worker monitored the patient's health status between office visits through telephone calls, home visits or office/hospital visits at least once every 6 weeks. PCP/nurse/social worker met at least monthly to review patient's status and revise care plans. Control (n=351): received usual care from their primary care physician. Physicians did not re-review patients as they came in for office visits during enrolment period and no new patients were added.	chronic conditions. USA			
Models of care including a self-management component					
Behm 2014	Intervention1 (n=174): Single home visit. Single home visit made by either a nurse, physiotherapist, social worker or occupational therapist. Participant given verbal and written information on what the urban district provides in terms of meeting places, activities, physical training for seniors, help and support available from professional organisations	Adults (age 80 or over, range of mean ages 85-86) Male to female ratio 46:64 Community Multimorbidity: number of	Quality of life - deterioration in self-rated health by SF-36 (24 months) Quality of life - deterioration in satisfaction with	Key features: multidisciplinary care, home follow-up, self-management	None stated

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	and volunteers. Visitor also identified falls risks and advice given on how to prevent falls. Visit lasted between 1.5 and 2 hours. Intervention 2 (n=171). Senior meetings. Four weekly meetings, no more than six participants in each group, each lasting ~2hrs, focus on information about aging process and consequences and provision of tools/strategies for solving problems that can arise in the home environment. Follow-up home visit two to three weeks after group meetings completed. Group meetings were multi-professional and multi-dimensional, led either by occupational therapist, nurse, physiotherapist or social worker. Control (n=114). Usual care. Access to ordinary range of services in municipality (for example, meals on wheels, help with ADLs)	participants with multimorbidity not reported	physical health (24 months) Quality of life - deterioration in satisfaction with psychological health (24 months)		
Boult 2008 *Boyd 2010 *Boult 2011 Boult 2013	Intervention (n=485): 'Guided Care' programme comprising 8 clinical services including home- based assessment, individual management plan, coaching for self-management with monthly monitoring and coordination of care provision. Delivered by trained guided care nurses.	Adults (intervention group mean age 77.2, range 66-106, control group mean age 78.1, range 66-96) Male to female ratio 409:435 Community	Mortality (32 months) Health-related quality of life – SF-12 (physical component) (32 months). Health-related quality of life – SF-12 (mental	Key features: holistic assessment, care plan, care co-ordination, self-management	Patients who were interviewed in their home for eligibility were considered ineligible if they did not have a telephone, did not speak English, were planning extended travel during the following 2.5 years, or failed a brief cognitive screen and did not

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	Control (n=419): usual care.	Multimorbidity: number of participants with multimorbidity not reported, mean number of self-reported conditions (conditions not specified): intervention: 4.3 (range 0-13); control: 4.3 (range 0-12). USA	component) (32 months) Patient satisfaction – Patient assessment of chronic illness care (PACIC) and 'very satisfied' with regular healthcare (32 months) Unscheduled care – emergency department visits (6-8 months). Continuity of care - management continuity (Primary care assessment survey integration and communication subscales) (32 months) Continuity of care - provider continuity (Access to doctor's appointment 'same day' when sick) (32 months)		have a proxy.
Chow 2014	Intervention 1 (n=96) Case management with home visits. A nurse case manager (NCM) carried out a pre-hospital discharge assessment using the Omaha	Adults (aged 65 years or over; mean 76.5) Male to female ratio	Health-related quality of life – SF-36 mental component (12 weeks)	Key features: holistic assessment, telephone follow-	MMSE <20, discharged to institutional care, unable to communicate, terminally ill

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	system (involves problem classification, interventions and problem rating). Patients received weekly visits for 4 weeks after discharge. Patients were encouraged to make decisions and take action to monitor their condition. Interventions were tailor made for patients. NCM made a home visit in the first week, in the second week the NCM called the patients to monitor and support them, in the third week nursing students visited the patient and in the fourth week the NCM made a final telephone call to remind them about adhering to positive behaviours. Duration 4 weeks. Intervention 2 (n=108) Case management with phone follow-up. A nurse case manager (NCM) carried out a pre-hospital discharge assessment using the Omaha system (involves problem classification, interventions and problem rating). Patients received weekly visits for 4 weeks after discharge. Patients were encouraged to make decisions and take action to monitor their condition. Interventions were tailor made for patients. The NCM made a first telephone call based on the patient's needs identified at assessment, nursing students called the patient in the second and third week post-discharge. Patients	134:147 Inpatient (prior to discharge) Multimorbidity: all patients had at least two co-morbid diseases Hong Kong	Health-related quality of life – SF-36 physical component (12 weeks)	up, home follow-up, self-management	

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	were referred to the goals and interventions developed by the NCM during the assessment. In the fourth week the NCM made a final phone call. Duration 4 weeks. Control (n=108) Placebo phone calls made twice in the 4 weeks, 5 minute calls only about social topics (for example, weather, television programmes, leisure activities). Duration 4 weeks.				
Coburn 2012	Intervention (n=873): HQP programme. Individualised plan developed by nurse case manager, based on: the patient's self-identified primary concerns and unmet needs; findings from their initial and on-going assessments; and the patient's motivational stage of change. The interventions typically incorporated into care plan include: education, symptom monitoring, medication reconciliation, counselling for adherence, help identifying, arranging and monitoring community and social service referrals. Group interventions directly provided by nurse case managers included: structured lifestyle and behaviour change programs for weight loss, weight loss maintenance, exercise classes and a balance and mobility programme for fall prevention.	Older adults (aged ≥ 65 years, mean age 74.8, SD 6.5) Male to female ratio 39:61 Community Multimorbidity: number of participants with multimorbidity not reported; mean number of chronic conditions 3.8 (SD 1.9). USA	Mortality (mean follow-up 4.2 years)	Key features: holistic assessment, care plan, care co-ordination, telephone follow-up, self-management	Dementia; end stage renal disease; schizophrenia; active cancer (except skin) in prior 5 years; life expectancy less than 6 months; current or imminent residence in long term care facility. Assessment of risk classified as low or very low according to a 'disease-specific risk assessment developed by HQP'.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	Control (n=863): usual care				
Gitlin 2006 *Gitlin 2009 *Gitlin 2006	Intervention (n=160): Multicomponent home intervention (the ABLE programme) delivered by occupational therapist (5 contacts, 4x face-to-face for 90 minutes and 1x 20 minute telephone contact) and physical therapist (90 minutes), aimed at reducing functional difficulties; over 6 months, followed by 6 month follow-up and 3 telephone contacts and final home visit. Control (n=159): patients assigned to no-treatment control group did not receive any intervention contact.	Adults (age 70 years or older, mean 79, SD 5.925) Male to female ratio 58:261 Community Multimorbidity: number of participants with multimorbidity not reported; mean number of conditions: intervention 7.1, control, 6.7. USA	Mortality (2, 3, 4 years from study) Functional outcomes – ADL (mean difference across 6 items) (6 months) Functional outcomes – IADL (mean difference across 6 items) (6 months) Functional outcomes – mobility (mean difference across 6 items) (6 months)	Key features: care plan, telephone follow-up, home follow-up, self-management	Acute suicidal ideation, score of ≥8 on the Alcohol Use Disorders Test alcohol assessment, recent lithium/antipsychotic medication use, inability to speak English or Spanish.
Katon 2010 *McGregor 2011 *Ludman 2013 *Von Korff 2012	Intervention (n=106): TEAMcare intervention integrating a treat-to-target programme with structured visits with nurses, individualised care plans and treatment targets, support for self-care combined with pharmacotherapy, provision of self-care materials for patients, weekly meetings to discuss case progression between nurses, primary care physicians, physiatrist and psychologist, electronic registry used to track risk factors and depression scores.	Adults (mean age 56.84, SD 11.35) Male to female ratio 108:112 Community Multimorbidity: patients with comorbid physical and mental health problems (that is, diagnoses of diabetes,	Health-related quality of life - Quality of life score, over the previous month (12 months). Health-related quality of life – Global Quality of Life rating (12 months). Mortality (12 months) Functional outcome – Sheehan Social Role Disability scale (12	Key features: care plan, telephone follow-up, self-management, medication management	Terminal illness, residence in a long-term care facility, severe hearing loss, planned bariatric surgery within 3 months, pregnancy or breast feeding, on-going psychiatric care, bipolar disorder or schizophrenia, use of antipsychotic or mood-stabiliser medication, and observed mental confusion suggesting dementia.

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
	Control (n=108): received "enhanced usual care", that is, after randomisation were advised to consult with their primary care physician to receive care for depression and for diabetes, coronary heart disease, or both.	coronary heart disease, or both and coexisting depression). USA	months). Functional outcome – WHODAS-2 activities of daily living (12 months). Patient and carer satisfaction - satisfaction with care of diabetes, heart disease, or both (12 months). Unscheduled care - proportion hospitalised (had at least 1) (12 months).		
Legrain 2011	Intervention (n=317): intervention led by geriatricians, targeted 3 risk factors for preventable readmissions and consisted of 3 components (comprehensive chronic medication review, education on self-management of disease, and detailed transition-of-care communication with outpatient health professionals). Control (n=348): standard care from the acute geriatric unit; care includes a rehabilitation component in addition to acute care.	Adults (aged 70 years or older, mean 86.4, SD 6.3) Male to female ratio 38:64 Patients identified as inpatients but intervention spans discharge Multimorbidity: number of participants with multimorbidity not reported; mean number of chronic conditions, mean	Mortality (6 months). Unscheduled care - unplanned admission to acute medical care or surgical unit (6 months). Unscheduled care - readmission to acute geriatric unit (6 months).	Key features: self-management, medication management	Expected length of stay less than 5 days; poor chance of survival at 3 months (according to clinical judgement of the senior geriatrician in charge); receiving palliative care; previous participation in OMAGE study; inclusion in another therapeutic trial, not French speaking, impossible to follow up (for example, lived in foreign country), absence of any health insurance (required

Study	Intervention and comparison	Population	Outcomes	Comments & key features	Exclusion criteria
		3.29 (SD 1.64). France			by French law on clinical trials).

10.1.2.1 Models of care

Table 115: Clinical evidence summary: models of care versus usual care – Alkema 2007 (holistic assessment, care co-ordination, telephone follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at 24 months	781 (1 study) 24 months	VERY LOW[a,b] due to risk of bias, indirectness	RR 0.61 (0.44 to 0.83)	-	87 fewer per 1000 (from 38 fewer to 125 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)

Table 116: Clinical evidence summary: models of care versus usual care – Beck 1997 (multidisciplinary care)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at 12 months	321 (1 study) 12 months	LOW[b,c] due to indirectness, imprecision	RR 0.56 (0.19 to 1.63)	56 per 1000	25 fewer per 1000 (from 45 fewer to 35 more)
Unscheduled care (urgent care visits per	321	LOW[a,b]		The mean visits per patient in the control	The mean visits per patient

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
patient) at 12 months	(1 study) 12 months	due to risk of bias, indirectness		group was 0.3	in the intervention group was 0.06 lower (0.23 lower to 0.11 higher)
Unscheduled care (emergency care visits per patient) at 12 months	321 (1 study) 12 months	LOW[a,b] due to risk of bias, indirectness		The mean visits per patient in the control group was 0.67	The mean visits per patient in the intervention group was 0.26 lower (0.54 lower to 0.02 higher)
Unscheduled care (proportion of patients hospitalised) at 12 months	321 (1 study) 12 months	LOW[a,b] due to risk of bias, indirectness		The mean proportion of patients hospitalised in the control group was 0.29	The mean visits per patient in the intervention group was 0.07 lower (0.14 lower to no difference)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 117: Clinical evidence summary: models of care versus usual care – Berglund 2015 (multidisciplinary care, holistic assessment, care plan, care co-ordination, telephone follow-up)

Outcomes	No. of	Quality of the evidence	Relative	Anticipated absolute effects

	Participants (studies) Follow up	(GRADE)	effect (95% CI)	Risk with Control	Risk difference with intervention (95% CI)
Mortality (died during total study) at 12 months	159 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias indirectness, imprecision	RR 1.42 (0.65 to 3.10)	118 per 1000	50 more per 1000 (from 41 fewer to 249 more)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 118: Clinical evidence summary: models of care versus usual care – Bouman 2008 (holistic assessment, telephone follow-up, home follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality (died during total study) at 18 months	330 (1 study) 18 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 1.34 (0.81 to 2.22)	135 per 1000	46 more per 1000 (from 26 fewer to 165 more)
Length of hospital stay (days per patient) at 18 months	330 (1 study) 18 months	LOW[a,b] due to risk of bias, indirectness		The mean days per patient in the control group was 8.54	The mean days per patient in the intervention group was 0.40 lower (4.3 lower to 3.5 higher)
Unscheduled care (hospital admissions) at 18 months	330 (1 study)	VERY LOW[a,b,c]	RR 1.20 (0.95 to	418 per 1000	84 more per 1000 (from 21 fewer to 217 more)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
	18 months	due to risk of bias, indirectness, imprecision	1.52)		
Unscheduled care (nursing home admissions) at 18 months	330 (1 study) 18 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.97 (0.42 to 2.21)	65 per 1000	2 fewer per 1000 (from 38 fewer to 78 more)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 119: Clinical evidence summary: models of care versus usual care –Courtney 2009 (holistic assessment, care plan, telephone follow-up, home follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Unscheduled care (emergency hospital readmission) at 6 months	122 (1 study) 6 months	LOW[b,c] due to risk of bias, indirectness	OR 0.14 (0.04 to 0.45)	-[a]	-[a]
Unscheduled care (emergency GP visits) at 6 months	122 (1 study) 6 months	VERY LOW[b,c] due to risk of bias, indirectness	RR 0.38 (0.24 to 0.61)	672 per 1000	417 fewer per 1000 (from 262 fewer to 511 fewer)

(a) Multivariate analysis with no adjusted raw data
(b) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

(c) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population (downgraded by 1 increment) or by a very indirect population (downgraded by 2 increments)

Table 120: Clinical evidence summary: models of care versus usual care – Eklund 2013 (multidisciplinary care, holistic assessment, care planning, care co-ordination)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality	161 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 1.49 (0.91 to 2.45)	237 per 1000	116 more per 1000 (from 21 fewer to 343 more)
Functional outcomes (any improvement in ADL)	161 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 1.64 (1.01 to 2.66)	237 per 1000	152 more per 1000 (from 2 more to 393 more)
Functional outcomes (any worsening in ADL)	161 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.79 (0.55 to 1.14)	474 per 1000	99 fewer per 1000 (from 213 fewer to 66 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by one/two1 or 2 increments because: the majority of the evidence included an indirect population (downgraded by one 1 increment) or by a very indirect population (downgraded by two 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

Table 121: Clinical evidence summary: models of care versus usual care – Ell 2010 (care co-ordination, telephone follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Health-related quality of life (SF12 mental) at 18 months Health-related quality of life: SF12 mental	387 (1 study) 18 months	VERY LOW[a,b] due to risk of bias,		The mean health related quality of life (sf12 mental) at 18 months in the control groups was	The mean health related quality of life (sf12 mental) at 18 months in the intervention groups was

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
component, high scores = better outcome		imprecision		43.49	1.61 higher (0.77 lower to 3.99 higher)
Health-related quality of life (SF12 physical) at 18 months Health-related quality of life: SF12 physical component, high scores = better outcome	387 (1 study) 18 months	VERY LOW[a,b] due to risk of bias, imprecision		The mean health related quality of life (sf12 physical) at 18 months in the control groups was 41.15	The mean health related quality of life (sf12 physical) at 18 months in the intervention groups was 1.28 lower (3.53 lower to 0.97 higher)
Functional Outcomes (scale of functional impairment) at 18 months Sheehan Disability Scale of functional impairment. Scale from: 1 to 10. Low scores = better outcome	387 (1 study) 18 months	LOW[a] due to risk of bias		The mean functional outcome (scale of functional impairment) at 18 months in the control groups was 3.18	The mean functional outcome (scale of functional impairment) at 18 months in the intervention groups was 0.1 higher (0.5 lower to 0.7 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias

(b) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)

Table 122: Clinical evidence summary: models of care versus usual care – Hogg 2009 (multidisciplinary care, care plan, telephone follow-up, home follow-up, medication management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Health-related quality of life (SF36	223	VERY LOW[a,b,e]		The mean health related quality of life	The mean health related quality of life

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
physical) at 15 months Health-related quality of life: SF36 physical component. Scale from: 0 to 100. High scores = better outcome.	(1 study) 15 months	due to risk of bias, imprecision, indirectness		(sf36 physical) at 15 months in the control groups was 41.5	(sf36 physical) at 15 months in the intervention groups was 1.6 higher (0.85 lower to 4.05 higher)
Health-related quality of life (SF36 mental) at 15 months Health-related quality of life: SF36 mental component. Scale from: 0 to 100. High scores = better outcome.	223 (1 study) 15 months	VERY LOW[a,b,e] due to risk of bias, imprecision, indirectness		The mean health related quality of life (sf36 mental) at 15 months in the control groups was 52.2	The mean health related quality of life (sf36 mental) at 15 months in the intervention groups was 1.1 lower (3.75 lower to 1.55 higher)
Health-related quality of life (total no days unhealthy in last 30 days) at 15 months	228 (1 study) 15 months	LOW[a,e] due to risk of bias, indirectness		The mean number of unhealthy days in the control groups was 9.9	The mean change in the number of unhealthy days in the intervention group was 1.4 lower (4.54 lower to 1.74 higher)
Mortality at 15 months	241 (1 study) 15 months	VERY LOW[a,b,e] due to risk of bias, indirectness	OR 7.58 (0.78 to 73.54)	- [c]	- [c]
Unscheduled care (average no of ED visits) at 15 months	241 (1 study) 15 months	LOW[a,e] due to risk of bias, indirectness		The mean number of ED visits in the control groups was 0.73	The mean change in unscheduled care (average no of ED visits) at 15 months in the intervention groups was 0.1 lower (0.37 lower to 0.17 higher)
Unscheduled care (average no of hospital admissions) at 15 months	241 (1 study)	LOW[a,e] due to risk of bias,		The mean number of ED visits in the control groups was	The mean change in unscheduled care (average no of hospital admissions) at 15 months in the intervention groups

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
	15 months	indirectness		0.73	was 0.06 lower (0.31 lower to 0.19 higher)
Patient/carer treatment burden (caregiver burden) at 15 months Scale (unspecified) from: 0 to 88, high scores = poor outcome.	129 (1 study) 15 months	LOW[a,e] due to risk of bias, indirectness		The mean caregiver burden at 15 months in the control groups was 14.7	The mean change in caregiver burden at 15 months in the intervention groups was 5 higher (1.41 to 8.6 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed on MID or by 2 increments if the confidence interval crossed both MIDs
(c) Absolute effects could not be calculated as control group event rate was 0
(d) Could not be calculated as adjusted raw data was not provided
(e) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population (downgraded by 1 increment) or by a very indirect population (downgraded by 2 increments)

Table 123: Clinical evidence summary: models of care versus usual care – Metzelthin 2013 (holistic assessment, care plan, home follow-up, care co-ordination)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Functional outcome (GARS - ADL	346	VERY LOW[a,b]		-[c]	The mean functional outcome (GARS -

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
subscale, 11-44, higher is worse outcome)	(1 study) 2 years	due to risk of bias, indirectness			ADL subscale, 11-44, higher is worse outcome) in the intervention groups was 0.77 higher (0.05 lower to 1.59 higher)
Functional outcome (GARS - IADL subscale, 7-28, higher is worse outcome) Scale from: 7 to 28.	346 (1 study) 2 years	VERY LOW[a,b] due to risk of bias, indirectness		-[c]	The mean functional outcome (GARS - IADL subscale, 7-28, higher is worse outcome) in the intervention groups was 0.40 higher (0.54 lower to 1.34 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by one/two1 or 2 increments because: the majority of the evidence included an indirect population (downgraded by one 1 increment) or by a very indirect population (downgraded by two 2 increments)
(c) Adjusted control group final scores not provided

Table 124: Clinical evidence summary: models of care versus usual care – Naylor 2004 (care co-ordination, home follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Health-related quality of life (Minnesota Living with Heart Failure Questionnaire) at 12 months Minnesota Living with Heart Failure Questionnaire. Scale from: 0 to 105. High scores = poor outcome.	149 (1 study) 12 months	LOW[a,c] due to risk of bias, indirectness		The mean quality of life (Minnesota living with heart failure questionnaire) at 12 months in the control groups was 2.6	The mean quality of life (Minnesota living with heart failure questionnaire) at 12 months in the intervention groups was 0.2 higher (0.36 lower to 0.76 higher)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at 12 months	239 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias, imprecision, indirectness	RR 0.87 (0.41 to 1.86)	107 per 1000	14 fewer per 1000 (from 63 fewer to 92 more)
Functional Outcomes (functional status score) at 12 months The Enforced Social Dependency Scale. Scale from: 12 to 72. High scores = poor outcome.	147 (1 study) 12 months	LOW[a,b] due to risk of bias, indirectness		The mean functional status (functional status score) at 12 months in the control groups was 2.9	The mean functional status (functional status score) at 12 months in the intervention groups was 0.2 higher (0.3 lower to 0.7 higher)
Patient & Carer Satisfaction (patient satisfaction) at 6 weeks The Patient Satisfaction Score. Scale from: 44 to 100. High scores = better outcome.	183 (1 study) 6 weeks	VERY LOW[a,b,c] due to risk of bias, imprecision, indirectness		The mean patient & carer satisfaction (patient satisfaction) at 6wk in the control groups was 77.8	The mean patient & carer satisfaction (patient satisfaction) at 6wk in the intervention groups was 5.3 higher (2.28 to 8.32 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population (downgraded by 1 increment) or by a very indirect population (downgraded by 2 increments)

Table 125: Clinical evidence summary: models of care versus usual care – Sandberg 2015 (holistic assessment, care co-ordination, home follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality (died during total study) at 12 months	153 (1 study) 12 months	VERY LOW[a,b,c] due to risk of bias indirectness, imprecision	RR 3.04 (0.87 to 10.62)	41 per 1000	84 more per 1000 (from 5 fewer to 395 more)
Length of hospital stay (days per patient) at 12 months	153 (1 study) 12 months	LOW[a,b] due to risk of bias, indirectness		The mean days per patient in the control group was 4.05	The mean days per patient in the intervention group was 0.55 higher (3.77 lower to 4.87 higher)
Unscheduled care (hospital admissions per patient) at 12 months	153 (1 study) 12 months	VERY LOW[a,b] due to risk of bias, indirectness		The mean admissions per patient in the control group was 0.48	The mean admissions per patient in the intervention group was 0.01 lower (0.25 lower to 0.27 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 126: Clinical evidence summary: models of care versus usual care – Slaets 1997 (multidisciplinary care, holistic assessment, care plan, care co-ordination)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at unclear time point	237 (1 study)	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 2.49 (0.96 to 6.49)	52 per 1000	77 more per 1000 (from 2 fewer to 283 more)
Unscheduled care (hospital readmission)	235 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.58 (0.36 to 0.93)	299 per 1000	126 fewer per 1000 (from 21 fewer to 191 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 127: Clinical evidence summary: models of care versus usual care – Sommers 2000 (multidisciplinary care, care plan, telephone follow-up, home follow-up)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at 24 months	543 (1 study) 24 months	VERY LOW[a,b] due to risk of bias, imprecision	RR 0.87 (0.51 to 1.47)	99 per 1000	13 fewer per 1000 (from 48 fewer to 46 more)
Unscheduled care (hospital admission) at 6 months	734 (1 study) 24 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.63 (0.41 to 0.96)	[c]	[c]

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

(c) Could not be calculated as adjusted raw data was not provided

10.1.2.2 Models of care including a self-management component

Table 128: Clinical evidence summary: models of care versus usual care – Behm 2014 (multidisciplinary care, home follow-up, self-management)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Intervention (95% CI)
Quality of life - single visit vs control - deterioration in self-rated health by SF-36	288 (1 study) 24 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	OR 0.64 (0.38 to 1.07)	330 per 1000	90 fewer per 1000 (from 172 fewer to 15 more)
Quality of life - group meetings vs control - deterioration in self-rated health by SF-36	285 (1 study) 24 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	OR 0.95 (0.57 to 1.57)	330 per 1000	11 fewer per 1000 (from 111 fewer to 106 more)
Quality of life - single visit vs control - deterioration in satisfaction with physical health	288 (1 study) 24 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	OR 0.43 (0.22 to 0.84)	210 per 1000	107 fewer per 1000 (from 27 fewer to 155 fewer)
Quality of life - group meetings vs control - deterioration in satisfaction with physical health	285 (1 study) 24 months	LOW[a,b] due to risk of bias, indirectness	OR 0.28 (0.14 to 0.59)	210 per 1000	141 fewer per 1000 (from 74 fewer to 174 fewer)
Quality of life - single visit vs control - deterioration in satisfaction with psychological health	288 (1 study) 24 months	LOW[a,b] due to risk of bias, indirectness	OR 0.30 (0.16 to 0.56)	290 per 1000	181 fewer per 1000 (from 104 fewer to 229 fewer)
Quality of life - group vs control - deterioration in satisfaction with psychological health	285 (1 study) 24 months	LOW[a,b] due to risk of bias, indirectness	OR 0.40 (0.22 to 0.72)	290 per 1000	150 fewer per 1000 (from 63 fewer to 208 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias.
(b) Downgraded by one/two1 or 2 increments because: the majority of the evidence included an indirect population (downgrade by one 1 increment) or a very indirect population (downgrade by two 2 increments).
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 129: Clinical evidence summary: models of care versus usual care – Boult 2008 (holistic assessment, care plan, care co-ordination, self-management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Health-related quality of life (SF-36 physical component). Scale from: 0 to 100. High scores = better outcome.	767 (1 study) 32 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision		-[d]	The mean health related quality of life (sf-36 physical) in the intervention group was 1.31 lower (3.02 lower to 0.4 higher)
Health-related quality of life (SF-36 mental component). Scale from: 0 to 100. High scores = better outcome.	767 (1 study) 32 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision		[d]	The mean health related quality of life (sf-36 mental) in the intervention group was 1.05 higher (1.06 lower to 3.16 higher)
Mortality	904 (1 study) 32 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 0.88 (0.59 to 1.31)	[d]	[d]
Patient and carer satisfaction (patient satisfaction, Patient and assessment of Chronic Illness (PACIC)) Scale not reported.	767 (1 study) 32 months	VERY LOW[a,b] due to risk of bias, indirectness		[d]	The mean patient satisfaction (pacic) in the intervention groups was 0.27 higher (0.08 to 0.46 higher)
Patient and carer satisfaction (patient satisfaction, 'very satisfied' with regular healthcare) Scale not reported.	767 (1 study) 32 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	OR 1.50 (0.77 to 2.90)	[d]	[d]
Unscheduled care (emergency department visits)	767 (1 study) 6-8 months	VERY LOW[a,b,c] due to risk of bias, indirectness,	OR 1.04 (0.81 to 1.34)	[d]	

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
		imprecision			
Continuity of care (integration subscale) Scale not reported.	767 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision		[d]	The mean continuity of care (integration subscale) in the intervention groups was 2.79 higher (0.97 lower to 6.55 higher)
Continuity of care (communication subscale) Scale not reported.	767 (1 study) 6 weeks	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision		[d]	The mean continuity of care (communication subscale) in the intervention groups was 2.97 higher (0.68 lower to 6.62 higher)
Continuity of care (same day access to GP when sick) Scale not reported	767 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	OR 1.20 (0.65 to 2.22)	[d]	

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(d) Could not be calculated as adjusted raw data was not provided

Table 130: Clinical evidence summary: models of care versus usual care – Chow 2014 (holistic assessment, telephone follow-up, home follow-up, self-management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Telephone follow-up vs control					
Health-related quality of life (SF-36 mental component) SF36. Scale from: 0 to 100. High scores = better outcome.	194 (1 study) 12 weeks	LOW[a,c] due to risk of bias, imprecision		The mean final SF-36 mental score in the control group was 53.6	The mean final SF-36 mental score in the intervention group was 1.2 higher (1.5 lower to 3.9 higher)
Health-related quality of life (SF-36 physical component) SF36. Scale from: 0 to 100. High scores = better outcome.	194 (1 study) 12 weeks	LOW[a,c] due to risk of bias, imprecision		The mean final SF-36 physical score in the control group was 39.3	The mean final SF-36 physical score in the intervention group was 3.3 higher (1.2 lower to 5.4 higher)
Home visits vs control					
Health-related quality of life (SF-36 mental component) at 12 weeks SF36. Scale from: 0 to 100. High scores = better outcome.	185 (1 study) 12 weeks	LOW[a,c] due to risk of bias, imprecision		The mean final SF-36 mental score in the control group was 53.6	The mean final SF-36 mental score in the intervention group was 1.9 higher (0.2 lower to 4.0 higher)
Health-related quality of life (SF-36 physical component) at 12 weeks SF36. Scale from: 0 to 100. High scores = better outcome.	185 (1 study) 12 weeks	LOW[a,c] due to risk of bias, imprecision		The mean final SF-36 mental score in the control group was 39.3	The mean final SF-36 physical score in the intervention group was 3.1 higher (1.0 lower to 5.2 higher)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Home visits vs telephone follow-up					
Health-related quality of life (SF-36 mental component) at 12 weeks SF36. Scale from: 0 to 100. High scores = better outcome.	183 (1 study) 12 weeks	LOW[a,c] due to risk of bias, imprecision		The mean final SF-36 mental score in the control (telephone) group was 54.8	The mean final SF-36 mental score in the intervention group was 0.7 higher (1.9 lower to 3.3 higher)
Health-related quality of life (SF-36 physical component) at 12 weeks SF36. Scale from: 0 to 100. High scores = better outcome.	183 (1 study) 12 weeks	MODERATE[a] due to risk of bias		The mean final SF-36 mental score in the control (telephone) group was 42.6	The mean final SF-36 physical score in the intervention group was 0.2 lower (2.4 lower to 2.0 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at a high risk of bias, and downgraded by 2 increments if the majority of the evidence was at a very high risk of bias
(b) Downgraded by 1 or 2 increments because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 131: Clinical evidence summary: models of care versus usual care – Coburn 2012 (holistic assessment, care plan, care co-ordination, telephone follow-up, self-management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality at 4.2 years	197 (1 study) 4.2 years	VERY LOW[b,c,d] due to risk of bias, imprecision, indirectness	HR 0.73 (0.55 to 0.97)	- [a]	- [a]

(a) Multivariate analysis with no adjusted raw data
(b) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(d) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population

Table 132: Clinical evidence summary: models of care versus usual care – Gitlin 2006 (care plan, telephone follow-up, home follow-up, self-management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference intervention (95% CI)
Survival - 2 years	319 (1 study) 24 months	MODERATE[d] due to indirectness	HR 0.39 (0.18 to 0.86)	- [a]	- [a]
Survival - 3 years	319 (1 study) 36 months	LOW[b,d] due to imprecision, indirectness	HR 0.74 (0.45 to 1.23)	- [a]	- [a]
Survival - 4 years	319 (1 study) 48 months	LOW[b,d] due to imprecision, indirectness	HR 0.76 (0.49 to 1.2)	- [a]	- [a]
Function - ADL Scale from: 1 to 5. High scores = poor outcome.	300 (1 study) 12 months	LOW[c,d] due to risk of bias, indirectness		- [a]	The mean function (activities of daily living) in the intervention groups was 0.1 lower (0.21 lower to 0.02 higher)
Function - IADL Scale from: 1 to 5. High scores = poor outcome.	300 (1 study) 12 months	LOW[c,d] due to risk of bias, indirectness		- [a]	The mean function (instrumental activities of daily living) in the intervention groups was 0.12 lower (0.26 lower to 0.03 higher)
Function (Mobility)	300	LOW[c,d]		- [a]	The mean function - mobility in the

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference intervention (95% CI)
Scale from: 1 to 5. High scores = poor outcome.	(1 study) 12 months	due to risk of bias, indirectness			intervention groups was 0.14 lower (0.29 lower to 0.01 higher)

(a) Multivariate analysis with no adjusted raw data
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(d) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population

Table 133: Clinical evidence summary: models of care versus usual care – Katon 2010 (care plan, telephone follow-up, self-management, medication management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Health-related quality of life (Global quality of life rating) Scale from: 0 to 10. High scores = poor outcome.	184 (1 study) 12 months	LOW[a] due to risk of bias		The mean health related quality of life (global quality of life rating) at 12 months in the control groups was 5.2	The mean health related quality of life (global quality of life rating) at 12 months in the intervention groups was 0.8 higher (3.11 lower to 4.71 higher)
Mortality	214 (1 study) 12 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.52 (0.05 to 5.05)	19 per 1000	9 fewer per 1000 (from 18 fewer to 84 more)
Functional outcomes (Sheehan social role disability scale) at 12 months Sheehan social role disability scale. Scale from: 0 to 10. High scores = poor	184 (1 study) 12 months	VERY LOW[a,b] due to risk of bias,		The mean functional outcome (Sheehan social role disability scale) at 12 months in the control groups was	The mean functional outcome (Sheehan social role disability scale) at 12 months in the intervention groups was 0.7 lower

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
outcome.		imprecision		4.5	(1.55 lower to 0.15 higher)
Functional outcomes (WHODAS-2 activities of daily living) at 12 months WHODAS-2 activities of daily living. Scale from: 0 to 4. High scores = better outcome.	184 (1 study) 12 months	LOW[a] due to risk of bias		The mean functional outcome (whodas-2 activities of daily living) at 12 months in the control groups was 12.9	The mean functional outcome (whodas-2 activities of daily living) at 12 months in the intervention groups was 0 higher (3.07 lower to 3.07 higher)
Patient & carer satisfaction (as assessed by the number of patients satisfied with care for diabetes, heart disease or both)	180 (1 study) 12 months	VERY LOW[a,b] due to risk of bias, imprecision	RR 1.22 (1.04 to 1.43)	705 per 1000	154 more per 1000 (from 28 more to 303 more)
Unscheduled care (proportion hospitalised at least once)	214 (1 study) 12 months	VERY LOW[a,b] due to risk of bias, imprecision	RR 1.20 (0.73 to 1.95)	213 per 1000	43 more per 1000 (from 58 fewer to 202 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Unable to calculate absolute affects from odds ratio

Table 134: Clinical evidence summary: models of care versus usual care – Legrain 2011 (self-management, medication management)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Mortality	634 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, imprecision, indirectness	RR 0.86 (0.62 to 1.19)	187 per 1000	26 fewer per 1000 (from 71 fewer to 35 more)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with intervention (95% CI)
Unscheduled care (emergency department visit)	665 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, imprecision, indirectness	RR 0.95 (0.52 to 1.72)	63 per 1000	3 fewer per 1000 (from 30 fewer to 46 more)
Unscheduled care (emergency hospital readmission)	665 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, imprecision, indirectness	RR 0.85 (0.69 to 1.05)	382 per 1000	57 fewer per 1000 (from 118 fewer to 19 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(c) Downgraded by 1 or 2 increments because: the majority of the evidence included an indirect population

10.1.3 Narrative findings

Several studies reported data that could not be analysed as data were incompatible (for example, data from multilevel modelling). These data are now presented in narrative form below.

Models of care (case management and care plan)

One study [26] (Courtney 2009) compared the effect of a comprehensive nursing and physiotherapy assessment and individually tailored program (including exercise strategies and nurse-conducted home visit and telephone follow-up, in addition to standard care) versus standard care alone in people with acute medical admissions, with a mean follow-up of 12 weeks. They reported health-related quality of life (SF12 physical component at 6 months) as a repeated measures analysis of covariance (ANCOVA); findings indicate a significant interaction effect between group*time. F (3,279) =30.4, ηp2 =0.50, p<0.001. These findings can be interpreted as a higher increase in intervention group scores in comparison to control group scores. Similarly, they reported health-related quality of life (SF12 mental component at 6 months) as a repeated measures ANCOVA; findings indicate a significant interaction effect between group*time. F(3,279) = 7.2, ηp2 = 0.19, p<0.001.These findings can be interpreted as a small increase in intervention group scores in comparison to control group scores. This study was of very low quality, due to risk of bias and imprecision.

Models of care (care plan)

A further study [87] (Slaets 1997) compared the effect of psychogeriatric intervention versus usual care, in people 75 years old or older who have been referred to the department of general medicine. They reported on functional outcomes (ADL and mobility), length of stay and admission to care facility. This study was of very low quality, due to risk of bias and imprecision. Evidence from this study is summarised in Table 135. 18. Findings indicate a significant association between group (intervention versus control) and activities of daily living, mobility, length of stay, and discharge to a nursing home, after controlling for gender, age, living condition, and physical functioning on admission.

Table 135: Psychogeriatric intervention versus usual care – efficacy of intervention without adjusting for MMSE (Mini Mental Status Examination) and GDS (Geriatric Depression Scale): results from Slaets 1997[42]

	Ba	SEb	βc	SEd	pe	Intervention n (%)	Usual Care n (%)
Physical functioning on discharge: ADL	-.30	.10	-.13	.04	<.01		
Physical functioning on discharge: mobility	-.27	.10	-.15	.05	<.01		
Length of stay (adjusted for gender, age, and living condition) total	-.35	.23	-.10	.06	.13		
Length of stay (adjusted for gender, age, living condition, and physical functioning on admission) total	-.52	.22	-.15	.06	.02		

(a) B = unstandardised regression coefficient adjusted for age, gender, and living situation
(b) SE = standard error of B
(c) β = standardised regression coefficient
(d) SE = standard error of β
(e) p = significance level

Models of care (Sommers 2000)

A further study[220] (Sommers 2000) reported on emergency department visits, social activities (higher count is more activities, quality of life (SF-36 self-rated health, 0-100, higher score is better function) and functional outcomes (HAQ, higher score is poorer function). This study was of very low quality, due to risk of bias and imprecision. Findings indicate more participants in the intervention group attended the emergency department and that the intervention group achieved better outcomes in terms of social activities and functional outcomes but worse outcomes in terms of self-rated health.

Table 136: Results from Sommers 2000

	Rate (no of patients with event)/adjusted mean score			Difference (1994-1993)	P value
	1992 (baseline)	1993	1994		
More than 1 emergency department visit					
Control group	5.6 (9)	17.4 (26)	16.7 (26)	-0.66	0.77
Intervention group	9.3 (16)	20.2 (33)	21.4 (38)	1.2	
Social activities count (higher is better outcome)					
Control group	9.1	8.9	8.6	-0.3	0.04
Intervention group	8.5	8.5	8.7	0.2	
SF-36 self-rated health (higher is better outcome)					
Control group					
Intervention group	3.1	3.2	3.3	0.1	0.08
	3.1	3.2	3.2	0	
HAQ functional outcome (higher is worse outcome)					
Control group	0.35	0.42	0.50	0.08	0.14
Intervention group	0.40	0.41	0.44	0.03	

(a) p value for experimental condition x year interactions

10.1.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

The study by Katon et al (2012)[123] was included in the clinical review but was excluded from the economic review as it was a US study (a health system that differs markedly from the UK).

One study was identified in the health economic search which relates to nurse-led case management.[86] This study is assessed as partially applicable with very serious limitations and it was selectively excluded. These are listed in Appendix M with reasons for exclusion given.

See also the economic article selection flow chart in Appendix F.

10.1.5 Evidence statements

Clinical

- 20 studies comprising a total of 8832 people evaluated models of care for people with multimorbidity. None of these studies were pooled due to variability in the content of their models of care. Overall the evidence demonstrated limited clinical benefit in critical outcomes compared to usual care. No individual model of care was consistently shown to be more effective than any other. No individual model of care demonstrated a high quality evidence base showing consistent benefit compared with usual care.
- The majority of the evidence was either of low or very low quality. There was moderate quality evidence that there was no clinical difference between home follow-up and telephone follow-up with regards to quality of life in one trial of 312 people. There was moderate quality evidence that there was a clinical benefit in terms of survival from one trial of 319 people. However the latter trial was not supported by any other trials assessing a similar intervention, other evidence from the trial suggested no benefit of the intervention in terms of functional outcomes and, like the majority of trials included in this review, the trial was conducted in an older adult population as a proxy for a population with multimorbidity.

Economic

- No relevant economic evaluations were included.

10.1.6 Recommendations and link to evidence

Recommendations	No recommendations made
Description of current UK services	Current services are broadly divided into specialist care which usually focuses on 1 condition (although sometimes on pairs, as happens in joint antenatal or diabetes care), and generalist care largely delivered by general practice and community nursing, and geriatric medicine for older people. Within these broad service types there is variation in how access, treatment and follow-up are organised, but most regular follow-up is single condition focused even in general practice. The GDG felt that many different models of care are currently being employed across the UK to care for people with multimorbidity. These models include

	various specific components including holistic assessment, multidisciplinary care, care co-ordination, home visits, telephone follow-up and self-management programmes. The GDG considered that the most common model is likely to be Community Matron type model although the function of people in these roles varies.
Relative values of different outcomes	The GDG considered health-related quality of life, mortality, functional outcomes, patient and carer satisfaction, length of hospital stay, unscheduled care and admission to care facility as critical outcomes for evaluating the effectiveness of interventions targeted at improving outcomes and continuity of care. The GDG also considered continuity of care and patient and carer treatment burden as important outcomes.
Trade-off between clinical benefits and harms	The GDG considered evidence for a number of models of care. No model was evaluated in more than one study, although there is an overlap with the interventions described by authors as comprehensive geriatric assessment. The GDG looked across the data and felt that no 1 model emerged with strong consistent evidence of clinical benefit. The GDG noted that many of the studies reported no clinically important difference in many critical outcomes. A small number of models were associated with a clinically important reduction in mortality, unscheduled care and health-related quality of life (physical component). However all trials also demonstrated no clinical difference between the model of care and usual care on other critical and important outcomes. Furthermore there was inconsistency in benefits across trials with similar model components. The complexity of the interventions assessed also prevents an understanding of which component of care is driving the effect. Due to the very low quality of this evidence and lack of a consistent model of intervention, the GDG felt that it was not possible to make any specific recommendations either for or against any one model of intervention.
Economic considerations	The GDG decided not to make any recommendations on any specific model of care as the clinical evidence was deemed to be insufficient to support or advise against any model of care. Modifying the current practice may involve some costs which would not be justified by the evidence found in the clinical review. Therefore, the GDG did not want to make any recommendations which may increase costs without improving outcomes.
Quality of evidence	The overall quality of the evidence varied from moderate to very low. Evidence was downgraded due to risk of bias, imprecision and indirectness. Evidence was downgraded for indirectness if the studies were included because population age was greater than 65 and there was no clear evidence that the population were multimorbid. The GDG also noted that the majority of this evidence was derived from studies assessing an older adult (less than age 65) population as opposed to a definitively multimorbid population. The GDG felt this was a further barrier to any recommendations in the multimorbid population on the basis of this evidence.
Other considerations	The GDG were aware of a body of evidence that has evaluated the effectiveness of collaborative care in people with comorbid depression and a physical health condition. However, the interventions in these trials were principally targeted at improving symptoms in a person's depression. For example, interventions were conducted by mental health professionals with the aim of improving clinical management of depressive symptoms. Where outcomes in physical health were reported, these were assessed as an indirect outcome of improving symptoms of depression. As these interventions were not aimed at improving management of all of the conditions experienced by a person with multimorbidity, these studies were excluded from this review. During the review process and discussion of the evidence, the GDG highlighted a specific model of intervention, a comprehensive geriatric assessment, that they felt may have accrued a sufficient body of evidence to warrant further review. This model is discussed further in the holistic assessment review.

10.2 Holistic assessment

10.2.1 Review question: What is the clinical and cost effectiveness of holistic assessment in patients with multimorbidity?

For full details see review protocol in Appendix C.

Table 137: PICO characteristics of review question

Population	Adults (18 years and over) with multimorbidity Strata: Inpatients; living in the community
Intervention	Holistic assessment
Comparison	Standard care
Outcomes	**Critical** • Health-related quality of life (EQ-5D) • Mortality (dichotomous or time-to-event) • Functional outcomes • Patient and carer satisfaction • Length of hospital stay • Unscheduled care • Admission to care facility **Important** • Continuity of care • Patient/carer treatment burden
Study design	Randomised controlled trials (RTCs) Controlled clinical trials (CCTs) Controlled before and after studies (CBAs) Interrupted time series analyses (ITS)

During the models of care review, the GDG identified that a large literature based on comprehensive geriatric assessment (CGA) that they were aware of had largely been missed in the search. The GDG noted that this intervention is performed in older adults with complex health difficulties, including adults with multimorbidity. However, as the majority of adults who receive a CGA are older adult, they believed that the papers may have not been indexed using multimorbidity terms. The GDG felt that this intervention may not only be relevant to older adults, but may also be relevant to all adults with multimorbidity, as a way of identifying and optimising care. As a consequence, the GDG chose to conduct an additional literature search and systematic review to identify papers evaluating the effectiveness of CGA in adults with multimorbidity.

The term CGA is commonly used in practice as this intervention is mostly conducted with older adults. However, to make this more applicable to a wider population of adults with multimorbidity, the GDG decided to use the term 'holistic assessment'. The GDG defined holistic assessment as a comprehensive assessment of a person that considers their physical health, mental health, social conditions and functional capabilities, which is then followed by the development of a care plan that seeks to address needs identified. As the care received following holistic assessment should be tailored to the needs of the individual person, the GDG expected significant variation within and between studies in the care people will receive. The GDG agreed that papers could be pooled together providing that they conducted a full holistic assessment as defined above, and stated that all care following the assessment was tailored to meet a person's needs identified in the assessment.

The GDG noted that precise nature of the follow-up to the holistic assessment would vary from study to study.

10.2.2 Clinical evidence

One Cochrane review[74] and thirty seven randomised clinical trials reported in 50 publications were included in the review [7,9,11,26,46,48-50,53,54,70,76,82,83,102,103,110,120,121,124,125,129-131,134-136,151,175,179,188,193,197-201,203-206,214-216,227,233,242,245 36,71,147,153]. All of these studies evaluated the effect of holistic assessment with older adults (CGA). Evidence from papers evaluating holistic assessment in an inpatient setting is presented first and evidence from papers evaluating holistic assessment in a community setting is then presented. Please see also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Inpatient holistic assessment

This review includes an updated review of the Cochrane review comparing Comprehensive Geriatric Assessment (CGA) intervention versus usual care for inpatient populations. Twenty of the 22 original papers have been included. In addition 5 papers, identified through an updated search and reference checking, have been included in the review. Studies with people in a care home facility have been pooled with the inpatient studies. This is because the GDG felt that the needs of people living in a care facility may be similar to those of people in hospital, and the method for delivering the intervention may also be similar.

In line with the analysis carried out in the Cochrane review, inpatient studies were split into 2 further strata; ward and team. Care in the 'ward' setting was delivered by a team in a discrete ward, with control over the delivery of the multi-disciplinary team's recommendations. Care in the 'team' setting was delivered by a multidisciplinary team assessing patients and delivering recommendations to the physicians caring for the person. The categorisation of studies into ward or team based holistic assessment was pre-planned and made in the Cochrane review as the 2 models have previously been considered as distinct but related interventions.

Twenty-five studies that evaluated holistic assessment in an inpatient setting were included. As expected, the studies described significant variety in the care received by people following holistic assessment. This included self-management (whereby participants were encouraged to self-care and coached in chronic disease management), multidisciplinary team management (MDT; for example, collaboration between the person's physicians, practice nurse(s), physiotherapist), and medication management (medication reconciliation carried out by pharmacist, medications of potential risk identified and alternative recommended).

Of the 25 inpatient studies: 17 studies were carried out in the USA/Canada; 6 studies were carried out in Europe and 2 studies were carried out in Australia/New Zealand.

Community holistic assessment

Twelve studies that evaluated holistic assessment in a community setting were included. The GDG noted that there was variation between the studies in the format of the assessment and in the number and seniority of clinicians conducting the assessment. As the GDG believed that these factors may impact on the efficacy and cost of the intervention, they decided to stratify the studies into assessments they believed were low in resource intensity (n=5) and those they believed were high in resource intensity (n=7). High intensity studies were those that required highly trained individuals performing interview/examination based assessments over longer periods of time (n=4) or included formal multidisciplinary meetings to formulate care plans (n=3). Low intensity studies typically involved a largely standardised questionnaire based assessment and care plan formulation involving

1 or 2 individuals familiar with the person (for example, the nurse who performed the assessment and a GP).

Studies reporting the risk of mortality and admission to care facility varied considerably in duration of follow-up; from 12 to 74 months. The GDG felt the people in these studies were likely to have high rates of both outcome due to their older age and multiple conditions, and this would mean the rates would vary noticeably based on the length of follow-up. The GDG therefore chose to present this evidence separately at each length of follow-up (follow-up 0-12 months, >12-24 months, >24-36 months).

Of the 12 community studies: 2 studies were carried out in the USA/Canada; 8 studies were carried out in Europe, 1 study was carried out in Australia/New Zealand and 1 study was carried out in Asia.

Table 138: Holistic assessment inpatient (ward)

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Applegate 1990 Miller 1994	Intervention (n=78): CGA and interdisciplinary assessment of medical, social and physiological function was completed within 72 hours of admission by team physicians, rehabilitation nurses, physical therapists, occupational therapists, psychologists, social workers, nutritionists, and specialist in speech therapy and audiology. When patients reached their rehabilitation goals or attained a stable level of function, they were discharged without any subsequent services from the geriatric-assessment-unit team. Control (n=77): received usual care provided by their physicians. The patients in the control group received a wide range of services after discharge from the acute care hospital, including home care in and care in other rehabilitation units. Care would compare favourably with national norms.	Older adult (aged ≥ 65 years, mean age 78.8 years) Male to female ratio 23:77 Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Activities of daily living (functional outcome) (6 months)	Excluded if they had medical problems that were unstable or required continued short-term monitoring, if their survival was estimated to be less than 6 months, if they had serious chronic mental impairment, or if a nursing home placement was considered inevitable.
Asplund 2000	Intervention (n=190): CGA. Acute geriatrics-based ward (AGW). The geriatric approach followed the principles outlined by the Nordic Working Group on Geriatric Assessment and Rehabilitation. Staffing of the ward was designed to optimise the conditions for treatment, nursing, early rehabilitation, and planning of care for older, acutely ill patients. Staff were recruited from the geriatric, medical and surgical departments. Consultants from	Older adults (aged ≥ 70 years) Male to female ratio 162:25 Multimorbidity: number of patients with multimorbidity not reported	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Length of stay (3 months) Hospital readmissions (3	Patients who required treatment in specialised units, such as the intensive care unit, coronary care unit, or acute stroke unit, or required treatment in 1 of the designated subspecialties, such as in a renal unit, were excluded.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	both the geriatric and medical departments had joint responsibility for medical care on the ward, with the internist having the main responsibility for acute diagnosis and medical treatment and the geriatrician taking over as soon as the medical condition had stabilised. The AGW had 11 beds and shared facilities with a surgical ward. Control (n=223): general medical wards each with 30 beds. Both were mixed wards in which acutely ill patients from the local hospital catchment area constituted the majority of patients.	Sweden	months)	
Cohen 2002 Phibbs 2006	Intervention (n=694): CGA. The inpatient and outpatients intervention teams, each consisting of a geriatrician, a social worker, and a nurse, followed their standard protocols for geriatric evaluation and management, with specific instructions to complete the history taking and physical examination, including screening for geriatric syndromes such as incontinence or falls; develop a list of problems; assess the patient's functional, cognitive, affective, and nutritional status; evaluate the caregiver's capabilities; and assess the patient's social situation. A plan of care was developed, and the team on the geriatric evaluation and management unit met at least twice a week to discuss the plan. Preventative and management services were coordinated to address the problems identified.	Older adults (aged ≥ 65 years, mean 74.2) Male to female ratio 1355:33 Multimorbidity: number of patients with multimorbidity not reported USA	Health-related Quality of Life (SF-36, 12 months) GEMC/UCOP Mortality (end of follow-up, time point unclear) GEMC/UCOP Admission to care facility (end of follow-up, time point unclear) GEMC/UCOP Length of stay (12 months) GEMC/UCOP	Admitted from a nursing home, were already receiving care at an outpatient clinic for geriatric evaluation and management, had previously been hospitalised in an inpatient unit for geriatric evaluation and management, were currently enrolled in another clinical trial, had a severe disabling disease or terminal condition or severe dementia, did not speak English, lacked access to a telephone for follow-up, or were unwilling or unable to return for follow-up clinic visits.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	Control (n=694): received all appropriate hospital services except for those provided by the team on the geriatric evaluation and management unit. Outpatients assigned to receive usual care were provided with at least 1 follow-up appointment in an appropriate clinic.			
Collard 1985	Intervention (n=218): CGA. Patients on the Geriatric Special Care Unit (GSCU) are cared for by registered nurses and nursing assistants. The 2 GSCUs share a full-time social worker, and each has a medical director. Within a short time of admission to the GSCU, a detailed assessment of each patient is performed by the primary nurse who coordinated the patient's hospital care. On the basis of the assessment, an individualised nursing care plan is developed for each patient. The care plan emphasises maximum patient independence. Discharge planning begins at admission. All members of the patient care team attend interdisciplinary conferences twice a week as they work. Shortly after discharge, the primary nurse calls the patient at home to see how well they are adjusting. Approximately 3 weeks after discharge, the primary nurse visits the patient at home to ascertain their progress and to identify problems that might have arisen since the patient left the hospital. Control (n=477): Usual care patients received care on 1 of the traditional medical/surgical units.	Older adult (aged ≥ 65 years, mean intervention group age 77.7 years, mean control group age 77.4 years) Male to female ratio 205:327 Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear)	None reported.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Counsell 2000	Intervention (n=767): CGA. Acute Care for Elders (ACE) intervention. Daily interdisciplinary team rounds were conducted by the geriatrician medical director and a geriatric clinical nurse specialist. Suggestions by the interdisciplinary team were recorded and communicated to the attending physician. Nursing care plans for fall risk assessment, mobility, self-care, skin integrity, nutrition, continence, confusion, depression, and anxiety, which had been modified for the intervention from those used routinely on usual care units, were implemented when appropriate. Medications of potential risk to older patients were identified by the medical director, who recommended alternatives. Hospital records were reviewed. Control (n=764): Usual care, no other information provided.	Older adults (aged ≥70 years) Male to female ratio 605:926 Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Readmission (unscheduled care, at 12 months) Patient & carer satisfaction (carer, at discharge)	Transferred from a nursing facility or another hospital, required speciality unit admission (for example, intensive care, coronary care, telemetry, or oncology), were admitted electively, had a length of stay less than 2 days, or had been previously enrolled in the study.
Fretwell 1990 Silliman 1990	Intervention (n=221): patients admitted to the Senior Care Unit, a regular 18-bed medical ward, were evaluated by the geriatric assessment team; which included a physician specialising in geriatric medicine, the nurse coordinator, a physical therapist, a clinical pharmacist, a dietician, and a social worker. The screening functional assessment was administered by the patient's primary nurse and reviewed within 24 hours of admission by the nurse coordinator. During the next 48 hours, each patient was evaluated by all	Older adults (aged ≥75 years) Male to female ratio 28:72 Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Length of stay (at discharge)	None reported.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	members of the team who, approximately 72 hours after randomisation, participated in an interdisciplinary team conference facilitated by the nurse coordinator. The team systematically reviewed medical diagnosis, medications, and problems in 6 areas of concern (nutrition, continence, cognition, emotion, mobility, and social support). Individualised care plan was developed and updated before discharge. The nurse coordinator provided telephone follow-up for a 2-month interval. Control (n=215): patients were housed on traditional medical and surgical floors and received the standard medical care of the hospital. A small number of control patients had consultation assessments by geriatricians but they did not receive the organised team intervention or follow-up that was provided for the treatment patients.		Carer treatment burden (self-reported health, at 3 months) Carer treatment burden (emotional health, at 3 months)	
Harris 1991	Intervention (n=97): CGA. The Geriatric Assessment Units (GAUs), 14-bed centre is 1 of 8 medical units, each of which practices general medicine together with a speciality interest. The GAU has a higher level of nursing staff and dedicated physiotherapy, occupational therapy and social work time. All 8 medical units participate in a roster which involved each unit being responsible for all medical admissions though the Emergency Department for a 24-hour period. Each unit has access to allied health professionals and all units undertake discharge planning. Follow-up	Older adults (aged ≥ 70 years) Male to female ratio 38:62 Multimorbidity: number of patients with multimorbidity not reported Australia	Mortality (end of follow-up, time point unclear) Activities of daily living (Functional outcome, at 12 months) Length of stay (at discharge)	None reported.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	interviews with patients at their place of residence were arranged at 3, 6, 9, and 12 months after discharge. Control (n=170): usual care patients admitted to 1 of 2 general medical units (GMUs).No other information provided.			
Harvey 2014	Intervention (n=57): geriatrician-led outreach service. The team comprised 2 part-time geriatricians and an aged care nurse consultant. Patients were recruited during their acute hospital stay and followed up at the residential care facility (RCF) for 6 months. The intervention group received a post-discharge home visit within 96 hours, at which a comprehensive geriatric assessment was performed and a care plan developed. Patients and their families were also offered further meetings to discuss Advanced Care Planning and document Advanced Directives. Control (n=59): usual care group was managed by the treating medical unit according to standard hospital protocols and received standard discharge planning, with follow-up at the RCF by their primary care physician service.	Older adult (aged ≥ 65 years, mean intervention group age 83.8, SD 7, mean control group age 86.7, SD 2.5) Male to female ratio 43:73 Multimorbidity: number of patients with multimorbidity not reported; mean number of conditions: intervention 7.7 (SD 2.7), control 5.7 (SD 2.5). Australia	Mortality (6 months). Patient & carer satisfaction – family/resident satisfaction (6 months) Unscheduled care – emergency department presentations (6 months). Unscheduled care – readmission rate (6 months).	Less than 65 years of age, were not living permanently in residential care facilities, had already been enrolled, had non-medical primary diagnoses, were expected to die during their index admission, lived outside the health service catchment area, exhibited severe behaviour disturbance, or consent was not obtained for study participation.
Kay 1992	Intervention (n=30): CGA. Weekly multidisciplinary (unit-based professional staff from the disciplines of nursing, medicine, social work, occupational therapy, physiotherapy, pharmacy and nutrition) team meetings, to	Older adults (aged ≥70 years, mean intervention group age 81.9, mean control group age 81.4)	Mortality (end of follow-up, time point unclear) Admission to care facility	None reported.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	evaluate client progress towards set goals and to formulate discharge plans, were facilitated by the primary nurse. Assessments of physical, cognitive and ADL functioning, as well as monitoring of medications, morale and discharge positions. All patients referred to the GAU were assessed by the consulting physician to the GAU project. Control (n=29): control patients were evaluated according to the research instrument; however, they did not move to the GAU and their care remained the same.	Male to female ratio 26:33 Multimorbidity: number of patients with multimorbidity not reported; mean number of problem areas: intervention 2.8 (range 1-6), control 2.5 (range 1-6). Canada	(end of follow-up, time point unclear)	
Landefeld 1995 Covinsky 1998 Covinsky 1997	Intervention (n=105): CGA. Each patient was assigned a primary nurse, 2 resident physicians, and an attending physician. Special unit designed to help older persons maintain or achieve independence in self-care activities. Under the leadership of the medical and nursing directors, the primary nurse assigned to each patient in the intervention group was responsible for assessing the patient's specific needs daily and implementing protocols for the prevention of disability and for rehabilitation. Control (n=324): usual care consisted of services provided by physicians and nurses in other acute care medical units. Staff of the intervention unit was not involved in the care of the patients receiving usual care, and none of the 4 elements of the program were implemented in usual care units.	Older adults (aged ≥ 70 years, intervention group mean age 80.2, SD 6.9, control group mean age 80.1, SD 6.6) Male to female ratio 216:435 Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Unscheduled care (readmission, at unclear, discharge from hospital) Length of stay (until discharge) Functional outcomes (improvement in ADL, at	Patients who were admitted to a speciality unit (for example, intensive care, cardiology-telemetry, or oncology) were ineligible.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
			discharge) Functional outcomes (worsening in ADL, at discharge)	
Nikolaus 1999	Intervention (n=179): CGA. Care plan (duration: assessment at discharge) - Comprehensive Geriatric Assessment (CGA) with recommendations, followed by usual care at home. The CGA was carried out once patients were in a stable medical condition. Intervention (n=181): CGA. Care plan (duration: mean = 7.6 days [range = 1 - 41 days]) - Comprehensive Geriatric Assessment (CGA) and additional in-hospital and post-discharge follow-up treatment by an interdisciplinary home intervention team. The CGA was carried out once patients were in a stable medical condition. The home intervention team (nurse, physiotherapist, OT) worked closely with hospital staff and the primary care physician. While the patient was in hospital the team gave them additional treatment (such as additional training in washing, eating dressing, and/or walking). One home visit was carried out during the hospital stay to evaluate the patient's home. After discharge, the team provided treatment which home services could not or could not immediately provide for as	Older adults (aged ≥ 65 years, mean age 81.4) Male to female ratio 145:400 Multimorbidity: number of patients with multimorbidity not reported Germany	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Functional outcome (activities of daily living) Length of stay Unscheduled care (hospital readmissions)	Patients with a terminal illness or severe dementia, patients who lived too far away (>15 km) for the home intervention team to make visits.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	long as necessary (twice a week, up to twice a day, for a minimum of 30 minutes). Control (n=185): usual care (duration: assessment at discharge) – assessment of activities of daily living and cognition, followed by usual care at home.			
Rubenstein 1984 Rubenstein 1988 Rubenstein 1995	Intervention (n=63): CGA. Innovative geriatric evaluation unit intended to provide diagnostic assessment, therapy, rehabilitation and placement. Patients were assigned to the geriatric evaluation unit, usually within 48 hours. Consultative and other hospital services available to patients in the control group were identical to services on the unit. Patients discharged from the unit usually received follow-up care in the geriatric medical outpatient clinic. Control (n=60): patients followed a natural course though the acute-care services of the hospital and were discharged to their homes or placed in long-term care facilities in the usual manner by acute-service personnel.	Older adults (aged ≥ 65 years, mean intervention group 78.8, mean control group 77.1) Male to female ratio 96:4 Inpatient Multimorbidity: number of patients with multimorbidity not reported; number of conditions: intervention 4.48 (SEM 0.27), control 4.45 (SEM 0.26). USA	Mortality (end of follow-up, time point unclear) Admission to care facility (up to 6 months) Admission to care facility (end of follow-up, time point unclear) Unscheduled care (readmission, unclear time frame) Functional outcome (independent in at least 2 ADL, at 24 months)	The following groups of patients were excluded: those with well-diagnosed severe dementia or another disabling disease (for example, multiple sclerosis or end-stage cirrhosis) resistant to further medical management who could perform no more than 3 activities of daily living, and who had no social support system that might be capable of preventing a nursing-home placement, those in the terminal phases of severe medical disorders (for example, malignant conditions or end-stage heart failure resistant to medical management), and those on the verge of discharge who were functioning well and would definitely return to the community without the need of support services or extended care.
Saltvedt 2002	Intervention (n=127): CGA. Patients allocated to the geriatric evaluation and management	Older adults (aged ≥ 75 years, mean intervention group 81.8,	Mortality (end of follow-up,	Patients with acute stroke were only included if the Stroke Unit

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Saltvedt 2004 Saltvedt 2005 Saltvedt 2006	unit (GEMU) were transferred on the day of inclusion. In the GEMU, the treatment strategy emphasised interdisciplinary assessment of all relevant disorders, prevention of complications and iatrogenic conditions, early mobilisation/rehabilitation, and comprehensive discharge planning. The staff of the GEMU consisted of 1 geriatrician and 1 (occasionally 2) resident. The number of nurses was comparable with that of other medical wards (MWs), although some of these nurses also had formal training in geriatric nursing. In addition, the GEMU had 2 occupational therapists and 1 physiotherapist. During the study period, a nurse was assigned to organise the study, recruit patients, and perform assessments during the index stay and follow-up. A social worker, a dentist, and other medical specialists were consulted when necessary. An interdisciplinary approach was employed, with close collaboration between all disciplines involved. Meetings were arranged twice a week to report assessments, set goals, discuss problems, and plan discharge. When necessary, relevant rehabilitation measures were initiated. In the GEMU, meetings were arranged to discuss necessary arrangements after discharge; patients, their family members, and representatives from the home services and the staff of the GEMU were invited. If necessary, an occupational therapist visited the patients at home to assess the need for adjustments. After patients were discharged from hospital, the GPs were responsible for the medical treatment of patients in both groups.	SD 4.8, mean control group 82.4, SD 5.2) Male to female ratio 89:165 Inpatient Multimorbidity: number of patients with multimorbidity not reported Norway	time point unclear) Admission to care facility (end of follow-up, time point unclear) Length of stay (unclear, 12 months) Unscheduled care (readmission, unclear 12 months) Functional outcomes (dependence in ADLs, at 12 months) Functional outcomes (dependence in IADLs, at 12 months)	was full. Nursing home patients and those previously fully independent and who seemed to recover quickly from the acute illness were not included, nor were patients for whom discharge was planned within 3 days. Other exclusion criteria were cancer with metastasis, other disease with expected survival less than 6 months, and known severe dementia before admission to hospital.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	Control (n=127): patients received treatment as usual from the Department of Internal Medicine. Residents and specialists in internal medicine and different subspecialties were responsible for the care provided. In the MW, home care nurses were telephoned to discuss arrangements after discharge if the hospital staff found it necessary. After patients were discharged from hospital, the GPs were responsible for the medical treatment of patients in both groups.			
Shamian 1984	Intervention (n=20): CGA. Patients were relocated for 9 weeks, following which they were moved back to the nursing units of origin. During the period of relocation, the patients fell under the care of a different health care team. In each case the patients was relocated from a unit with an acute medical or surgical focus to a unit where the focus was geriatric medicine. All experimental and control patients underwent 4 evaluations within 30-day interval and were observed for 90 days. All 4 evaluations included data on: mortality and morbidity; activities of daily living; and, medication management. At zero time, all experimental and control patients were evaluated on their original units. Following the initial evaluation, the experimental patients were transferred to the temporary unit, which was staffed by a geriatrician, a head nurse who was a geriatrics specialist, and a nursing staff which included both experienced geriatrics	Older adults (aged ≥ 65 years) Male to female ratio 14:12 Inpatient Multimorbidity: number of patients with multimorbidity not reported Canada	Mortality (end of follow-up, time point unclear)	Not reported.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	nurses and newly hired nursing staff. There was no occupational therapist or physiotherapist assigned to the unit, although these professionals were available as consultants from the regular geriatrics unit, and all subjects retained their previous social workers. Care was based on the multidisciplinary team approach used on the established geriatrics unit. Control (n=16): patients remained in their original units. Patients in the control remained on their units and received the same care as they had received prior to entrance in the study.			
White 1994	Intervention (n=20): CGA. An interdisciplinary geriatric team was developed consisting of a medical director/geriatrician, a gerontological nurse practitioner, a social worker, a dietician, a pharmacist, and an occupational therapist. The service was nurse-managed, with the philosophy of care encompassing a shift in focus from acute illness-driven care to restorative, functional-based care. The geriatric service performed consultations initiated by attending or resident physicians, social workers, and rehabilitation and nursing staff. The service comprised of 6 beds. Patients in the study group experienced a change in attending physician, transfer from a teaching, resident-managed service to a non-teaching, nurse-managed service.	Older adults (aged ≥ 65 years, mean intervention group age 77, SD 54, meant control group age 76, SD 54) Male to female ratio not reported Inpatient Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Unscheduled care (readmission, unclear 30 days post discharge)	Patients with a do not resuscitate order who are deemed to be “imminently terminal”.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	Control (n=20): patients received only a formal consultation with recommendations from the geriatric service. These patients remained with their original attending and resident physicians and received their care in the usual manner.			

Table 139: Holistic assessment inpatient (team)

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Altfeld 2013	Intervention (n=360): Intervention participants received the telephone-based Enhanced Discharge Planning (EDPP) assessment and an individualised plan following program protocols to address identified transitional care needs. The model involved the creation of a personalised intervention plan addressing both psychosocial and health issues, including connecting older adults to community resources, and collaborating with health care professional such as the discharge planning team, home health providers, and the physicians. The EDPP intervention began with a review of a referred patient for relevant medical and psychosocial information. The EDPP worker confirmed the post-discharge plan of care and identified potential problem area that required additional assessment. The EDPP social worker contacted patients or caregivers by telephone within 2 working days of discharge to assess the patient's post-discharge adjustment and needs.	Older adult (aged ≥ 65 years, mean 74.5, SD 6.9) Male to female ratio not reported Inpatient Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (30 days). Unscheduled care – readmission (30 days).	Patients unable to effectively communicate in English, discharged to a skilled nursing or home institutional care facility, or those involved in another transitional care intervention were excluded.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	Control (n=360): The Usual care group received conventional care given all patients discharged from the medical centre which did not include any post-discharge contact between hospital staff and patients or caregivers.			
Edmans 2013	Intervention (n=216): Intervention: usual care plus interface geriatrician. Usual care on the acute medical units before recruitment for both the control and intervention groups comprised assessment and treatment by a consultant physician and attending medical team. Patients in the intervention group were assessed before discharge from the acute medical unit by 1 of 12 geriatricians, who aimed to coordinate the delivery of whatever additional immediate care or aftercare they deemed necessary. Such care could include a review of diagnosis; a drug review; further assessment at home or in a clinic or by recommending admission rather than discharge; advanced care planning; or liaison with primary care, intermediate care, and specialist community services. The interface geriatricians from both centres met monthly to discuss their experiences and cases. Control (n=217): patients received no additional intervention over and above usual care.	Older adult (aged ≥70 years, mean 83, SD 6.8) Male to female ratio 159:274 Inpatient Multimorbidity: number of patients with multimorbidity not reported UK	Health-related quality of life (EQ-5D, 90 days) Mortality (90 days) Functional outcome (activities of daily living, ADL ≥17, 90 days) Admission to care facility (90 days)	Not being resident in the hospital catchment area, lacking mental capacity to give informed consent and without a family member to provide consent if lacking capacity, any exceptional reason cited by acute medical unit staff why patients should not be recruited, and participation in other related studies.
Hogan 1987	Intervention (n=57): CGA. Initial stage of	Older adults (aged ≥ 75 years,	Mortality (end of follow-up,	Patients were excluded if they

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	intervention was a medical consultation performed by the geriatrician, who made specific recommendations to the attending staff. After this the other service members (nurse and a physiotherapist) became involved, and recommendations and care came from any of them. Patients were seen daily on weekdays by at least 1 of the service members; full-team rounds were held once per week. At the time of discharge the assessor reviewed the discharge medications. Control (n=56): No CGA. No further information provided.	mean intervention group age 82.2, SD 6.2, mean control group age 83.3, SD 6) Male to female ratio 40:60 Inpatient Multimorbidity: number of patients with multimorbidity not reported; mean number of active conditions 6.1 (SD 3.2). Canada	time point unclear) Admission to care facility (end of follow-up, time point unclear) Length of stay (at unclear)	were in an intensive care unit, had suffered an acute cerebrovascular accident or if permission was refused by the patient or the attending staff physician
Kircher 2007	Intervention (n=105): CGA. The consultation service teams comprised a social worker and physician. The geriatrician summarised problems and recommendations in a structured treatment note. Team conferences were held at least weekly. Treatment was evaluated, and the implementation of recommendations was appraised. When necessary, the nurse or social worker visited the patient's home to appraise living conditions. The GP was contacted about the recommendations by the consultation service physician shortly before discharge. Community services received a detailed and structured recommendation plan and were contacted by telephone before discharge. Control (n=129): patients in the control group received all appropriate hospital services	Older adults (aged ≥ 65 years, mean intervention group age 79, SD 6.9, mean control group age 78.4, SD 6.9, mean external comparison group age 76.9, SD 7.5) Male to female ratio 106:254 Inpatient Multimorbidity: number of patients with multimorbidity not reported Germany	Health-related quality of life (end of follow-up, time point unclear) Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Unscheduled care (readmission, 12 months)	Admitted from a nursing home, had previously been hospitalised in a geriatric evaluation and management inpatient unit, had a terminal condition or severe dementia, did not speak German, were living beyond a 60km radius of the coordinating centre, would not need help at home or could not give informed consent.

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	except those provided by the consultation team.			
Naughton 1994	Intervention (n=51): CGA. Patients were admitted to the direct care of a team consisting of medical house staff, a social worker, and an attending geriatrician. The geriatrician and social worker comprised the core geriatric evaluation and management (GEM) team. A nurse clinical specialist and a physical therapist joined the core team as needed. The team regularly evaluated the patients' mental status, psychosocial condition, functional status, and medical condition to determine the medical, rehabilitative, and social needs of the patients. Information about the patients was discussed at team conferences 2 or 3 times per week. Responsibility for implementing the care plan was apportioned among team members. Control (n=60): patients were given 'usual care' by medical house staff and an attending physician. The care of these patients was assigned during each attending physician's clinical teaching rotation. The services of social workers and discharge planners were available upon request.	Older adults (aged ≥ 70 years, mean intervention group age 80.1, SD 6.6, mean control group age 80.1, SD 6.4) Male to female ratio 55:45 Inpatient Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Length of stay (at unclear, assume discharge from hospital)	Patients admitted to an intensive care unit or transferred from the medical service to a surgical service (for example, general surgery, urology, gynaecology).
Reuben 1995	Intervention (n=1261): CGA. Patients were interviewed and examined by a team comprising a social worker, a nurse practitioner, and a geriatric. Using standardised, multidimensional assessment instrument, the nurse practitioner recorded each patient's medical history and performed a	Older adults (aged ≥65 years, mean intervention group age 77.6, mean control group age 76.7) Male to female ratio 48:52	Mortality (end of follow-up, time point unclear)	Patients were excluded from the study if they had been admitted to a hospice or for terminal care, were not members of the HMO's health plan, lived outside the HMO's medical-service area or were usually cared for at a

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	limited physical examination, focusing on geriatric issues; the social worker assessed functional status and cognitive and emotional health, noted stressful or otherwise important events in the patient's life, and reviewed the patient's social support system, use of community services, and advance directives. After these evaluations, the nurse, social worker and geriatrician discussed the case. The geriatrician summarised the geriatric problems and the team's recommendations in a structured consultation note that was sent to both the attending physician and the patient's primary care physician. Team conferences were held daily. Control (n=1016): Patients assigned to the control group received usual care. No other information.	Inpatient Multimorbidity: number of patients with multimorbidity not reported USA		medical centre in the HMO that was not in the study, were discharged or died before randomisation, did not speak English, or were admitted from a nursing home.
Rubin 1993; Rubin 1992	Intervention (n=97): CGA. Comprehensive geriatric evaluation and development of a long term care plan conducted by geriatric assessment team (GAT). GAT consisted of geriatric-internist, geriatric psychiatrist, geriatric clinical nurse specialist and geriatric social worker. Control (n=97): Usual inpatient care - care for my medical team consisting of attending physician, resident intern and medical students. No access to geriatric consultation; could not have been referred to geriatric clinic after discharge.	Older adults (aged ≥70 years, mean intervention group age 76.8, SD 5.8, mean control group age 76.7, SD 5.3) Male to female ratio 39:61 Inpatient Multimorbidity: number of patients with multimorbidity not reported; mean number of conditions: intervention 4.2	Mortality (1 year) Katz ADL (1 year) Five-Item OARS IADL (1 year) Patient treatment burden ('health troubles stand in the way of doing things a great deal') (1 year)	Unable to give informed consent for example, medical instability or severe cognitive impairment; admitted to non-medicine service; known to be terminally ill upon admission; under care of private physician; judged too socially and medically stable and independent

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
		(SD 3.2), control 3.6 (SD 1.4). USA		
Thomas 1993	Intervention (n=62): CGA. Multi-dimensional geriatric team assessment, leading to formal recommendations to the attending physician. A standard proprietary instrument, the Functional Assessment Inventory, was used to evaluate each patient, then the experimental group received individual assessments from each team member consisting of a physician, geriatric nurse specialist, home health nurse, medical social worker, dietician, pharmacist, and physical therapist. Team discussion of each patient led to formal recommendations place in the patient's chart. An additional copy of the consultation was mailed to the attending physician's office. The team continued to monitor progress of the experimental group. Control (n=58): patients received same standard proprietary instrument assessment as intervention group but did not receive the individual assessments by each member of the team, their recommendations or subsequent visits.	Older adults (aged ≥70 years, mean intervention group age 77, SD 5.4, mean control group age 76, SD 5.4) Male to female ratio 46:74 Inpatient Multimorbidity: older adult, multimorbidity data not reported. USA	Mortality (end of follow-up, time point unclear) Functional outcome (activities of daily living, unclear, 12 months) Length of stay (12 months)	Patients were excluded for the refusal of consent, admission to the intensive care unit or coronary care unit, an obvious terminal illness, renal haemodialysis, or place of residence greater than 50 miles from the hospital.
Trentini 2001	Intervention (n=79): CGA (performed at end of the hospitalisation period before discharge) and CGA-based interventions (conducted after discharge). Participants received a complete and personalised treatment plan based on	Older adults (aged ≥65 years, mean intervention group age 78.7, SD 0.8, mean control group age 80.0, SD 0.7)	Mortality (1 year) Admission to care facility (1 year)	Age <65; terminal disease; completely bed-ridden; living in a nursing home; good health (defined as no need for home care); severe disabling

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	results of CGA and performed by the same geriatric team in the outpatient clinic or day hospital. Planned evaluations at 3, 6 and 12 months. Control (n=73): Intervention 2: CGA (performed at end of the hospitalisation period before discharge). No personalised care plan. Entrusted to GP with standard discharge letter. Planned evaluations at 3, 6 and 12 months.	Male to female ratio 40:60 Inpatient Multimorbidity: number of patients with multimorbidity not reported; mean number of conditions: intervention 4.2 (SD 0.2), control 3.9 (SD 0.2). Italy		irreversible conditions; likely non-compliance
Winograd 1993	Intervention (n=99): CGA. Consisted of a comprehensive functional, mental, medical, and social evaluation and recommendations by an interdisciplinary team consisting of an attending faculty geriatrician, a geriatric fellow, and internal medicine house officer, a social worker, and a clinical nurse specialist. After initial evaluation, the team met as a group to discuss the patient and formulate recommendations, Recommendations were directed primarily at 5 areas: medical issues, referral for rehabilitation, evaluation and management of geriatric syndromes, discharge planning, and psychological issues. A formal consultation note outlining recommendation was place in the patients' charts and discussed with the primary care team. Patients were seen a minimum of 3 times per week throughout the hospital stay and follow-up notes were written on at least a weekly basis. Control (n=98): patients receive usual care and	Older males (aged ≥65 years, mean intervention group age 75.7, SD 9.0, mean control group age 76.6, SD 9.7) Male to female ratio 100:0 Inpatient Multimorbidity: number of patients with multimorbidity not reported; mean number of conditions: intervention 4.4 (SD 2.0), control 4.3 (SD 1.7). USA	Mortality (end of follow-up, time point unclear) Admission to care facility (end of follow-up, time point unclear) Functional outcome (activities of daily living, 12 months) Length of stay (12 months)	Patients were excluded if they were independent in all activities of daily living prior to hospital admission ('too independent'), were a permanent nursing home resident, and had a terminal illness with life expectancy of less than 6 months by report of primary physicians ('too impaired').

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	were not evaluated by the consultation team.			

Table 140: Community holistic assessment (low intensity)

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Boorsma 2011	Intervention (n=201): CGA every 3 months until end of follow-up. The interview consisted of a computerised assessment of functional health, activities of daily living, depression, cognition, satisfaction with care, and use of medications. CGA done using web tool, no information on personnel carrying out CGA. Control (n=139): family physician responsible for medical care and offered it on request.	Older adult (mean intervention group age 85.8, SD 6.2, mean control group age 85.5, SD 8.0) Male to female ratio 84:256 Community dwelling - care facility Multimorbidity: number of patients with multimorbidity not reported Netherlands	Mortality (6 months) Hospitalisation (6 months) SF-12 (6 months)	People with terminal illness
Frese 2012	Intervention (n=630): in-home CGA, using the STEP-tool (standardised assessment of elderly people in primary care in Europe; a combination of a structured questionnaire and a structured physical examination) and (Barthel-Index, Lambeth questionnaire, Tinetti-gait score, Hamilton Depression scale, MMSE, Hierarchic Dementia scale, clock drawing test and COOP-Charts), followed by recommendations for the general practitioner. CGA performed by trained medical students, took up to 1 hour. Recommendations made by	Older adults (aged ≥70 years, intervention group age range 79.65-84.04, control group age range 79.74-87.94 years) Male to female ratio 460:1137 Community dwelling - approximately 20% nursing home	Mortality (6.2 years) Time to mortality (6.2 years) Admission to care facility (6.2 years) Time to admission to care facility or death (6.2 years)	Participants who moved out of screening area; refusal to participate before or at the appointment time

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	geriatrician trainees under supervision from experienced geriatrician. 100 patients had 2 CGAs (3 years apart), 236 patients had 1 CGA and the remainder did not receive a CGA. Control (n=990): received usual general practitioner care, including home visits by their GP when necessary. GPs were asked to rate every patient's state of health. In the context of the German health care system, usual care means that the patient should consult their GP at first, but they can also directly consult specialists. All the patients have in principle equal access to the necessary health care resources.	Multimorbidity: number of patients with multimorbidity not reported Germany		
Li 2010	Intervention (n=152): comprehensive geriatric assessment and care plan. Recommendations for care included. medication adjustment, exercise instruction, nutrition support, physical rehabilitation, social worker consultation, and specialty referral. Skilled nurses trained in assessment undertook CGA; geriatricians formulated care plans on basis of CGA. Control (n=158): received screening evaluation only.	Older adults (mean intervention group age 78.4, SD 8.2, mean control group age 79.3, SD 8.5) Male to female ratio 162:142 Community dwelling Multimorbidity: number of patients with multimorbidity not reported Taiwan	Mortality (6 months) Activities of daily living – Barthel Index (6 months)	Conditions such as being bedridden, receiving home care by visiting nurses, less than 6 months' life expectancy (such as terminal cancer patients), and difficulty in verbal communication (such as severe cognitive or hearing impairments).

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Monteserin 2010	Intervention (n=151) CGA by trained nurses including sociodemographic data, health status, sensory evaluation, falls, urinary incontinence, Charlson Index (co-morbidity), Barthel index (functional status), Lawton index (ADLs), 5-Y Depression scale, Pfeiffer's test (mental state), nutritional assessment, Gijon social scale. If patients were deemed at risk of frailty, they had an individual educational session by a geriatrician including an extended 30 minute visit informing patient about lifestyle changes, making shared plans re: drug therapy, sensory impairment, falls, incontinence aids and dietary modifications. Duration 18 months. Control (n=134) The control group also received a comprehensive assessment by trained nurses including sociodemographic data, health status, sensory evaluation, falls, urinary incontinence, Charlson Index (co-morbidity), Barthel index (functional status), Lawton index (ADLs), 5-Y Depression scale, Pfeiffer's test (mental state), nutritional assessment, Gijon social scale. The control then received usual standard care from their GP, no care plan was formulated and care was not anticipated to change. Duration 18 months.	Older adults (Mean age 79.9 years (range 75-94), Male to female ratio 40:60 Community dwelling Multimorbidity: number of patients with multimorbidity not reported Spain	Mortality (18 months) Admission to nursing home (18 months)	Concurrent inclusion in another study, diagnosis of terminal disease, institutionalisation, severe cognitive impairment, difficulties in accessing primary health care centre, inability/unwillingness to give consent
Senior 2014	Intervention (n=52): CGA and care plan (the restorative care service) delivered in short-stay residential care facilities and at participants' residences with the aim of reducing the requirement for permanent residential care.	Older adults (aged ≥65 years, mean intervention group age 83.6, SD 6.9, mean control group age 81.9, SD 6.8)	Mortality (24 months) Admission to care facility – residential care placements	Excluded if in order to maintain person's safety they required immediate permanent residential care placement; inability to communicate in

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	Case manager met with patient and conducted CGA, CGA + patient's own goals used to develop care plan by team (nurse, occupational therapist and physiotherapist). Control (n=53): content of control intervention unclear, included community services or permanent placement in residential care. Older people were assessed and service coordinated by a centrally based needs co-ordinator	Male to female ratio 46:54 Community dwelling Multimorbidity: number of patients with multimorbidity not reported New Zealand	(24 months)	English.

Table 141: Community holistic assessment high intensity

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Brettschneider 2015	Intervention (n=150): On first visit CGA performed by trained personnel (nursing scientist, psychologist or sociologist – allocation and number present at each CGA not specified) in assessing nutrition status, sight and hearing, incontinence, loss of functional muscle mass. Social activities, housing conditions, economic conditions, polypharmacy and cognitive status determined. Visit followed by a case conference with nursing scientist, psychologist, gerontopsychiatrist, nutritionist and social worker within 3 weeks of assessment, which provided individualised recommendations based on analysis of identified self-care deficits and risk factors for institutionalisation. Second	Older adults (mean 84, SD 3.5) Male to female ratio (31:69) Community dwelling Multimorbidity: number of patients with multimorbidity not reported Germany	Mortality (18 months) Admission to nursing home at 18 months EQ-5D at 18 months	Insufficient German language skills, cognitive impairment, not able to give consent, care level >1 (if needed assistance with more than 2 activities of basic nursing more than once a day, maximum amount of care must not exceed 3 hours a day).

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	visit performed by same personnel who performed first visit, reported to patient on outcome of case conference and presented recommendations. A third visit 4 weeks later, evaluated adherence to recommendations and identified obstacles and facilitators, recommendations were reviewed and further support offered. Duration was 4 weeks. Control (n=155) Usual care (every service offered by statutory health insurance system and utilised at patient's own initiative). Duration 4 weeks.			
Counsell 2007	Intervention (n=474): CGA (GRACE intervention). The GRACE support team consisted of an advanced practise nurse and social worker, who care for low-income older adults, in collaboration with the patient's primary care physician and a geriatrics interdisciplinary team led by a geriatrician. The support team met with the patient in the home to conduct an initial CGA. The support team then presented their findings to the larger GRACE interdisciplinary team to develop an individualised care plan. Then the support team met face-to-face with the patient's primary care physician to discuss the care plan and make any modifications. The support team then implemented the plan through face-to-facer and telephone contact with patients, family members, caregivers and healthcare professionals. Each patient received a minimum of 1 home follow-up to review care	Older adults (aged ≥65 years, mean intervention group age 71.8, SD 5.6, mean control group age 71.6, SD 5.8) Male to female ratio 24:76 Living in the community Multimorbidity: number of patients with multimorbidity not reported USA	Basic ADL (2 years) IADL (2 years) SF-36 (physical component) (2 years) SF-36 (mental component) (2 years)	Residence in a nursing home; living with a study participant already enrolled in the trial; enrolled in another research study; receiving dialysis; severe hearing loss; English language barrier; no access to a telephone; severe cognitive impairment (defined by Short Portable Mental Status Questionnaire score ≤5); without an available caregiver to consent to participate

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
	plan, 1 telephone or face-to-face contact per month and a face-to-face home visit after any ED visit or hospitalisation. Duration 2 years Comparison (n=477): usual care			
Ekdahl 2015	Intervention (n=208): Initial CGA was performed based on a standardised procedure at the ambulatory geriatric unit. Thereafter all care was personalized according to patients' situations and preferences, best-known evidence and practice and team members' competences. Nurses reassessed patients after 1 year and began home visits if needed. The team of professionals at the ambulatory geriatric unit (nurse, geriatrician, care manager, occupational therapist, physiotherapist, dietician) planned care during team meetings; the common goal was increasing quality of life. Care managers contacted patients and informed them of available forms of support from municipality service. Intensity of follow-up ranged from few contacts per year to daily/weekly visits. Many activities had preventive goals (for example, physiotherapy training programmes). Nurses also ensured patients understood new prescriptions and visited patients who were admitted to hospital to provide further information to staff caring for them. Control (n=174): Usual care.	Older adults (aged 75 years or older, mean age 82.5, SD 4.9) Male to female ratio 51:49 Community dwelling Multimorbidity: all patients had a minimum of 3 concomitant medical diagnoses Sweden	Health-related Quality of Life (24 months) Mortality (24 months) Inpatient days per patient (24 months) Hospitalisations per patient (24 months) Nursing home admissions (24 months)	None reported

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Epstein 1990	Intervention (n=185): Two hour CGA conducted by geriatrician, geriatric nurse practitioner and a geriatric social worker - reviewed medical records, performed a comprehensive physical examination that focused on drugs, nutrition, new diagnoses and the function impact of illness, measured functioning (ADLs and IADLs), reviewed social support, social activities, coping style, psychological function, and economic and environmental issues. Care plan - team met for approximately 15 minutes after seeing the patient to generate a care plan and consult as a group with the patient and family. 3 follow-up telephone contacts with the patient or family during the first 2 months after the examination. Control: (n=205): Standard care using traditional health maintenance organisation services	Older adults (aged ≥70 years, mean intervention group age 76.7, SD 4.9, mean control group age 76.9, SD 4.6) Male to female ratio 49:51 Living in the community Multimorbidity: number of patients with multimorbidity not reported USA	Mortality (1 year) Sickness Impact Profile (4 physical function scales, 51 items) (1 year) Patient satisfaction (developed from literature, based primarily on scale of Di Matteo and Hays, 12 item)(1 year) Length of hospital stay (1 year) Hospitalisation (1 year)	Not reported
Karppi 1995; Karppi 1995A	Intervention (n=104): Patients moved to inpatient geriatric rehabilitation unit for comprehensive multidisciplinary assessment (mean length of stay 16.5 days). In the ward there were 1 geriatrician, 5 nurses, 7 auxiliary nurses, 3 assistants, 2 physiotherapists, 1 psychologist, 1 occupational therapist and 1 part-time social worker. 1 psychiatrist visited once a week. Specialists were consulted when needed. Given a rehabilitation plan to be followed at home. Control (n=208): Usual supervised home care.	Older adults (aged ≥65 years, mean 78.5, SD 4.3) Male to female ratio 22:78 Living in the community Multimorbidity: number of patients with multimorbidity not reported Finland	Mortality (1 year) Katz ADLs (3 months) Lawton & Brody IADL (3 months) Admission to care facility (1 year)	Terminal phase of illness; only a single acute disease or injury; psychosis; care in the geriatric unit in the last year

Study	Intervention and comparison	Population	Outcomes	Exclusion criteria
Lampela 2013 Lampela 2010 Lihavainen 2012 Lihavainen 2012a	Intervention (n=500): in addition to health status monitoring by a trained nurse, participants underwent CGA. This included an overall health status assessment (2 weeks after patient's visit at the study nurse) including medication assessment by a physician. The CGA also included nutritional status assessment and mobility, balance and muscle strength assessment. Persons in the intervention group also had counselling and case manager services by a trained nurse Control (n=158): received no interventions, standard health care services in public and private sector were available for them. The health status was monitored annually.	Older adults (aged ≥75 years, mean 81.1, SD 5) Male to female ratio 30:70 Community dwelling Multimorbidity: number of patients with multimorbidity not reported Finland	Mortality (3 years)	Not reported
Melis 2008	Intervention (n = 88): Geriatric specialist nurse visited the patient at home. Up to 6 visits for additional geriatric evaluation and management were planned within the next 3 months. Starting with a wide multidimensional assessment, the team developed an individualised, integrated treatment plan for each patient. The nurse conducted the main part of the intervention. The nurse and geriatrician made recommendations to the primary care physicians. Control (n = 67). Usual care as per primary care physician.	Older adults (age 70 or older, mean intervention group age 82.8, SD 6.6, mean control group age 81.7, SD 5.9) Male to female ratio 48:113 Community dwelling Multimorbidity: number of patients with multimorbidity not reported Holland	Health related quality of life – mental (6 months) Health related quality of life – physical (3 months) Health related quality of life – function (3 months) Mortality (2 years) Function (6 months)	Request for help has an acute nature or purely medical diagnostic issue, MMSE <20 or proven moderate to severe dementia, patient already receiving form of intermediate care from social care/geriatrician, patient on waiting list for nursing home, life expectancy <6 months.

Holistic assessment inpatient – ward

Table 142: Clinical evidence summary: Holistic assessment inpatient (ward) versus usual care

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
Health-related quality of life (SF36) at 12 months - Physical functioning 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	LOW[c,d] due to indirectness, imprecision		The mean health related quality of life (sf36) at 12 months - physical functioning in the control groups was 4.5	The mean health related quality of life (sf36) at 12 months - physical functioning in the intervention groups was 2.3 higher (1.4 to 3.2 higher)
Health-related quality of life (SF36) at 12 months - Physical limitations 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	LOW[c,d] due to indirectness, imprecision		The mean health related quality of life (sf36) at 12 months - physical limitations in the control groups was 32.5	The mean health related quality of life (sf36) at 12 months - physical limitations in the intervention groups was 1.2 lower (4.02 lower to 1.62 higher)
Health-related quality of life (SF36) at 12 months - Emotional limitations 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	MODERATE[c] due to indirectness		The mean health related quality of life (sf36) at 12 months - emotional limitations in the control groups was 20.2	The mean health related quality of life (sf36) at 12 months - emotional limitations in the intervention groups was 1.9 higher (0.99 to 2.81 higher)
Health-related quality of life (SF36) at 12 months - Bodily pain 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	MODERATE[c] due to indirectness		The mean health related quality of life (sf36) at 12 months - bodily pain in the control groups was 22.9	The mean health related quality of life (sf36) at 12 months - bodily pain in the intervention groups was 1 lower (2.04 lower to 0.04 higher)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
Health-related quality of life (SF36) at 12 months - Energy 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	MODERATE[c] due to indirectness		The mean health related quality of life (sf36) at 12 months - energy in the control groups was 1	The mean health related quality of life (sf36) at 12 months - energy in the intervention groups was 4.4 higher (4.04 to 4.76 higher)
Health-related quality of life (SF36) at 12 months - Mental health 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	MODERATE[c] due to indirectness		The mean health related quality of life (sf36) at 12 months - mental health in the control groups was 0.8	The mean health related quality of life (sf36) at 12 months - mental health in the intervention groups was 5.5 higher (5.06 to 5.94 higher)
Health-related quality of life (SF36) at 12 months - Social activity 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	LOW[c,d] due to indirectness, imprecision		The mean health related quality of life (sf36) at 12 months - social activity in the control groups was 16.4	The mean health related quality of life (sf36) at 12 months - social activity in the intervention groups was 1.9 higher (0.33 to 3.47 higher)
Health-related quality of life (SF36) at 12 months - General health 36-item Short-Form General Health Survey (SF-36). Scale from: 0 to 100. Higher is better	1388 (1 study) 12 months	MODERATE[c] due to indirectness		The mean health related quality of life (sf36) at 12 months - general health in the control groups was -8.2	The mean health related quality of life (sf36) at 12 months - general health in the intervention groups was 3.8 higher (3.13 to 4.47 higher)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
Mortality	6709 (17 studies) 1-24 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	RR 0.99 (0.9 to 1.08)	208 per 1000	2 fewer per 1000 (from 21 fewer to 17 more)
Functional Outcomes (activities of daily living)	967 (4 studies) 6-12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision		The mean functional outcomes (activities of daily living) in the control groups was 48.46	The mean activities of daily living - ward in the intervention groups was 0.11 standard deviations higher (0.03 lower to 0.24 higher)
Functional outcomes (participants with improving ADL)	650 (1 study) at discharge	LOW[c,d] due to indirectness, imprecision	RR 1.41 (1.11 to 1.81)	241 per 1000	99 more per 1000 (from 26 more to 195 more)
Functional outcomes (participants with worsening ADL)	650 (1 study) at discharge	LOW[c,d] due to indirectness, imprecision	RR 0.76 (0.55 to 1.05)	210 per 1000	50 fewer per 1000 (from 94 fewer to 10 more)
Functional outcomes (participants independent in at least 2 ADL)	123 (1 study) 24 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	RR 1.33 (0.85 to 2.10)	333 per 1000	110 more per 1000 (from 50 fewer to 367 more)
Functional outcomes (participants dependent for ADL – Barthel index <12)	133 (1 study) 12 months	VERY LOW[c,d] due to indirectness, imprecision	1.09 (0.59 to 2.00)	230 per 1000	21 more per 1000 (from 94 fewer to 230 more)
Functional outcomes (participants dependent for IADL – Lawton score <4)	131 (1 study)	VERY LOW[c,d] due to indirectness,	1.01 (0.69 to	441 per 1000	4 more per 1000 (from 137 fewer to 212 more)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
	12 months	imprecision	1.48)		
Patient & carer satisfaction (family/resident satisfaction, tool not specified, number of people satisfied at end of follow-up)	44 (1 study) 6 months	VERY LOW[a,c,d] due to risk of bias, imprecision, indirectness	RR 1.63 (1.14 to 2.23)	583 per 1000	367 more per 1000 (from 82 more to 770 more)
Patient & carer satisfaction (caregiver satisfaction) at discharge, 12 months Scale from: 0 to 100.	333 (1 study) 12 months	VERY LOW[a,c,d] due to risk of bias, imprecision, indirectness	-	The mean patient & carer satisfaction (caregiver satisfaction) at discharge in the control groups was 59	The mean patient & carer satisfaction (caregiver satisfaction) at discharge in the intervention groups was 3 higher (0.96 to 5.04 higher)
Patient & carer satisfaction (patient satisfaction), at 1 month Scale from: 0 to 100. Higher is better.	958 (1 study) 12 months	LOW[a,c] due to risk of bias, indirectness	-	The mean patient & carer satisfaction (patient) at 1 months in the control groups was 72	The mean patient & carer satisfaction (patient) at 1 months in the intervention groups was 3 higher (0.91 to 5.09 higher)
Length of stay	3303 (9 studies) 3-12 months	VERY LOW[a,b,c] due to risk of bias, indirectness, inconsistency, imprecision		The mean length of stay - ward in the control groups was 14.8	The mean length of stay - ward in the intervention groups was 1.41 higher (1.14 lower to 3.95 higher)
Unscheduled care (emergency department presentations)	116 (1 study) 6 months	VERY LOW[a,c,d] due to risk of bias, imprecision, indirectness	RR 0.70 (0.45 to 1.11)	475 per 1000	142 fewer per 1000 (from 261 fewer to 52 more)
Unscheduled care (hospital readmissions)	3543	LOW[a,c]	RR 1.00	280 per 1000	0 more per 1000 (from 28 fewer to

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
	(8 studies)	due to risk of bias, indirectness	(0.90 to 1.11)		31 more)
Admission to care facility	6252 (14 studies) 1-24 months	MODERATE[d] indirectness	RR 0.79 (0.71 to 0.87)	222 per 1000	47 fewer per 1000 (from 29 fewer to 64 fewer)
Carer burden (poor self-reported health)	105 (1 study) 3 months	VERY LOW[a,c,d] due to risk of bias, imprecision, indirectness	OR 0.51 (0.29 to 0.90)	-[e]	-[e]
Carer burden (poor emotional health)	105 (1 study) 3 months	VERY LOW[a,c,d] due to risk of bias, imprecision, indirectness	OR 0.77 (0.49 to 1.21)	-[e]	-[e]

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1/2 increments because: heterogeneity, I2>50%, unexplained by subgroup analysis (risk of bias and date of publication), analysis conducted using random effects
(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(d) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(e) Could not be calculated as adjusted raw data was not provided

Holistic assessment inpatient - team

Table 143: Clinical evidence summary: Holistic assessment inpatient (team) versus usual care

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
Health-related quality of life (Quality of Life Philadelphia Geriatric Centre Morale Scale) Scale not reported.	279 (1 study) 12 months	LOW[a,c] due to risk of bias, indirectness		8 (7-9) median and interquartile	8 (7-10) median and interquartile
Health-related quality of life (EQ-5D) Scale 0 to 10. Higher is better.	285 (1 study) 90 days	LOW[a,c] due to risk of bias, indirectness		The mean activities of daily living - team in the control groups was 0.45	The mean health related quality of life (eq-5d) in the intervention groups was 0 lower (0.07 lower to 0.07 higher)
Mortality	4418 (9 studies) 1-12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	RR 0.99 (0.88 to 1.11)	208 per 1000	0 fewer per 1000 (from 17 fewer to 19 more)
Mortality - time to event	433 (1 study) 90 days	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	HR 1.22 (0.57 to 2.61)	- [e]	- [e]
Functional outcomes (activities of daily living)	3329 (2 studies) 12 months	LOW[a,c] due to risk of bias, indirectness		The mean activities of daily living - team in the control groups was 9	The mean activities of daily living - team in the intervention groups was 0.25 lower (0.76 lower to 0.26 higher)
Functional outcomes (activities of daily living:	433	VERY LOW[a,c,d]	OR 1.25	- [e]	- [e]

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with holistic assessment versus usual care (95% CI)
Barthel ADL), at 90 days Barthel ADL ≥17. Scale not reported.	(1 study) 90 days	due to risk of bias, indirectness, imprecision	(0.72 to 2.17)		
Functional outcomes (activities of daily living: Katz ADL – improved) Scale not reported.	194 (1 study) 12 months	VERY LOW[a,c,d] due to imprecision, indirectness	RR 0.86 (0.49 to 1.51)	216 per 1000	30 fewer per 1000 (from 110 fewer to 110 more)
Functional outcomes (activities of daily living: five-items OARS IADL – improved) Scale not reported.	194 (1 study) 12 months	LOW[a,c] due to indirectness, imprecision	RR 2.0 (0.95 to 4.23)	93 per 1000	93 more per 1000 (from 5 fewer to 300 more)
Length of hospital stay	563 (4 studies) 12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision		The mean length of stay in the control groups was 12.2	The mean length of stay in the intervention groups was 1.07 lower (2.66 lower to 0.52 higher)
Unscheduled care (hospital readmissions)	999 (2 studies) 1-12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 1.16 (0.87 to 1.56)	- [e]	- [e]
Admission to care facility	1441 (7 studies) 1-12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 0.87 (0.64 to 1.19)	- [e]	- [e]
Patient/carer treatment burden (patient treatment burden, 'health troubles stand in the way of doing things a great deal')	120 (1 study) 12 months	LOW[a,c] due to indirectness, imprecision	RR 0.56 (0.36 to 0.86)	567 per 1000	249 fewer per 1000 (from 79 fewer to 363 fewer)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1/2 increments because: the heterogeneity I2>50%

(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(d) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(e) Could not be calculated as adjusted raw data was not provided

Community holistic assessment - low intensity

Table 144: Clinical evidence summary: Community holistic assessment (low intensity) versus usual care

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with low intensity Community holistic assessment (95% CI)
Health-related Quality of Life (SF-12) Better indicated by higher values, scale from 0 to 100	234 (1 study) 6 months	LOW[a,b] due to risk of bias, indirectness		The mean health related quality of life (sf-12) in the control groups was 42.56	The mean health related quality of life (sf-12) in the intervention groups was 0.25 lower (1.9 lower to 1.4 higher)
Mortality - 0-12 months follow-up	650 (2 studies) 6 months	VERY LOW[a,b,c,d] due to risk of bias, inconsistency, indirectness, imprecision	OR 1.1 (0.88 to 1.37)	84 per 1000	8 more per 1000 (from 9 fewer to 28 more)
Mortality - 12-24 months follow-up	390 (2 studies) 18-24 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 0.65 (0.35 to 1.23)	144 per 1000	46 fewer per 1000 (from 89 fewer to 28 more)
Mortality - 74 month follow-up	1620 (1 study) 74 months	LOW[a,c] due to risk of bias, indirectness	OR 0.78 (0.67 to 0.91)	- [e]	- [e]
Mortality - time to event	1725 (2 studies) 24-74 months	LOW[a,c] due to risk of bias, indirectness	HR 0.79 (0.69 to 0.9)	- [e]	- [e]
Functional outcome: Barthel Index Scale from: 0 to 100. Better indicated	269 (1 study)	VERY LOW[a,c,d] due to risk of bias,		The mean score on the Barthel index in the control group was	The mean score on the Barthel index in the intervention group was

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with low intensity Community holistic assessment (95% CI)
by higher values.	6 months	indirectness, imprecision		91.6	4 higher (0.27 lower to 8.27 higher)
Admission to care facility >12-24 month follow-up	390 (2 studies) 18-24 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 0.67 (0.32 to 1.38)	144 per 1000	43 fewer per 1000 (from 93 fewer to 45 more)
Admission to care facility - 74 month follow-up	1620 (1 study) 74 months	LOW[a,c] due to risk of bias, indirectness	OR 0.8 (0.68 to 0.95)	- [e]	- [e]
Admission to care facility - time to event	1725 (2 studies) 24-74 months	LOW[a,c] due to risk of bias, indirectness	HR 0.80 (0.69 to 0.92)	- [e]	- [e]
Unscheduled care (hospitalisation)	227 (1 study) 6 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 1.32 (0.94 to 1.85)	141 per 1000	37 more per 1000 (from 7 fewer to 92 more)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 or 2 increments because the point estimate varies widely across studies
(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(d) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(e) Could not be calculated as adjusted raw data was not provided

Community holistic assessment – high intensity

Table 145: Clinical evidence summary: Community holistic assessment (high intensity) versus usual care

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
Health-related quality of life (EQ-5D) Better indicated by higher values. Scale from: 0 to 1.	525 (2 studies) 18-24 months	LOW[a,c] due to risk of bias, indirectness	-	The mean final EQ-5D score in the control groups was 0.59	The mean final EQ-5D score in the intervention groups was 0 higher from 0.06 lower to 0.05 higher
Health-related quality of life (MOS-20, mental health) Better indicated by higher values. Scale from: 0 to 100.	155 (1 studies) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	The mean final MOS-20 score in the control groups was 53.2	The mean final MOS-20 mental health score in the intervention groups was 9.1 higher from 2.4 higher to 15.6 higher
Health-related quality of life (MOS-20, physical performance) Better indicated by higher values. Scale from: 0 to 100.	155 (1 studies) 3 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	-[d]	The mean final MOS-20 physical performance score in the intervention groups was 4.3 higher from 2.9 lower to 11.2 higher
Health-related quality of life (MOS-20, role functioning) Better indicated by higher values. Scale from: 0 to 100.	155 (1 studies) 3 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	-[d]	The mean final MOS-20 role functioning score in the intervention groups was 4.7 higher from -9.8 lower to 19.3 higher
Health-related quality of life (SF-36,	951	MODERATE[c]	-	The mean change in sf-36	The mean sf-36 (physical

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
physical component) Better indicated by higher values. Scale from: 0 to 100.	(1 study) 2 years	due to indirectness		(physical component) in the control groups was -1.6	component) in the intervention groups was 0.5 higher (0.62 lower to 1.62 higher)
Health-related quality of life (SF-36, mental component) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 (mental component) in the control groups was -0.3	The mean sf-36 (mental component) in the intervention groups was 2.4 higher (1.06 to 3.74 higher)
Health-related quality of life (SF-36 scales - physical functioning) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - physical functioning in the control groups was -6.8	The mean sf-36 scales - physical functioning in the intervention groups was 1.5 higher (1.4 lower to 4.4 higher)
Health-related quality of life (SF-36 scales - role-physical) Better indicated by higher values. Scale from: 0 to 100. Higher is better.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - role-physical in the control groups was -2.7	The mean sf-36 scales - role-physical in the intervention groups was 4.6 higher (0.35 lower to 9.55 higher)
Health-related quality of life (SF-36 scales - bodily pain) Better indicated by higher values. Scale from: 0 to 100. Higher is better.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - bodily pain in the control groups was 0.8	The mean sf-36 scales - bodily pain in the intervention groups was 0.7 lower (3.91 lower to 2.51 higher)

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
Health-related quality of life (SF-36 scales - general health) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - general health in the control groups was -2.3	The mean sf-36 scales - general health in the intervention groups was 2.5 higher (0.06 to 4.94 higher)
Health-related quality of life (SF-36 scales – vitality) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	MODERATE[c] due to indirectness	-	The mean change in sf-36 scales - vitality in the control groups was -2.6	The mean sf-36 scales - vitality in the intervention groups was 5.2 higher (2.55 to 7.85 higher)
Health-related quality of life (SF-36 scales - social functioning) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - social functioning in the control groups was -2.3	The mean sf-36 scales - social functioning in the intervention groups was 5.3 higher (1.43 to 9.17 higher)
Health-related quality of life (SF-36 scales - role-emotional) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - role-emotional in the control groups was -2.6	The mean sf-36 scales - role-emotional in the intervention groups was 2.1 higher (3.42 lower to 7.62 higher)
Health-related quality of life (SF-36 scales - mental health) Better indicated by higher values. Scale from: 0 to 100.	951 (1 study) 2 years	LOW[c,d] due to indirectness, imprecision	-	The mean change in sf-36 scales - mental health in the control groups was -0.3	The mean sf-36 scales - mental health in the intervention groups was 3.9 higher

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
					(1.57 to 6.23 higher)
Mortality - 0-12 month follow-up	702 (2 studies) 12 months	VERY LOW[a,b,c] due to inconsistency, indirectness, imprecision	OR 0.99 (0.58 to 1.7)	92 per 1000	1 fewer per 1000 (from 36 fewer to 55 more)
Mortality - >12-24 month follow-up	815 (3 studies) 18-24 months	LOW[a,c] due to risk of bias, indirectness	OR 0.56 (0.39 to 0.80)	236 per 1000	88 fewer per 1000 (from 38 fewer to 128 fewer)
Mortality - >24-36 month follow-up	1000 (1 study) 36 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	OR 1.15 (0.81 to 1.62)	144 per 1000	18 more per 1000 (from 24 fewer to 70 more)
Mortality – time to event	382 (1 study) 24 months	LOW[a,d] due to risk of bias, imprecision	HR 0.66 (0.43 to 1.01)	270 per 1000	82 fewer per 1000 (from 143 fewer to 2 more)
Functional outcomes (activities of daily living: Katz ADL) Scale from: 0 to 6. Better indicated by higher values.	1252 (2 studies) 0.4-2 years	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	-	The mean Katz ADL ranged across control groups from -1.6 to 4.9 change score	The mean Katz ADL in the intervention groups was 0.06 lower (0.3 lower to 0.19 higher)
Functional outcomes (activities of daily living: Lawton & Brody IADL) Scale from: 0 to 8. Better indicated by higher values.	1252 (2 studies) 0.4-2 years	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	-	The mean Lawton & Brody IADL ranged across control groups from 0.4 to 4.9 change score	The mean Lawton & Brody IADL in the intervention groups was 0.12 lower (0.45 lower to 0.22 higher)
Functional outcomes (Sickness Impact Profile)	382 (1 study)	VERY LOW[a,c,d] due to risk of bias,	-	The mean Sickness Impact Profile in the control groups was	The mean Sickness Impact Profile in the intervention

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
Scale from: 0 to 100. Better indicated by lower values.	1 year	indirectness, imprecision		89	groups was 2 higher (0.41 lower to 4.41 higher)
Functional outcomes (GARS-3) Scale from: 0 to 100. Better indicated by lower values.	155 (1 study) 6 months	LOW[a,c] due to risk of bias, indirectness	-	The mean final GARS-3 score in the control group was 37.0	The mean final GARS-3 score in the intervention group was 1.6 lower (3.9 lower to 0.7 higher)
Patient & carer satisfaction (patient satisfaction) Better indicated by higher values. Unvalidated scale. Scale from: 0 to 5.	382 (1 study) 12 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	-	The mean patient satisfaction in the control groups was 4.28	The mean patient satisfaction in the intervention groups was 0.11 higher (0.06 lower to 0.28 higher)
Length of hospital stay (inpatient days per patient)	382 (1 study) 12 months	MODERATE[a] due to risk of bias	-	The mean inpatient days per patient in the control group was 15.2	The mean inpatient days per patient in the intervention group was 4.1 lower (7.8 lower to 0.4 higher)
Unscheduled care (hospitalisation)	390 (1 study) 12 months	LOW[c,d] due to indirectness, imprecision	RR 0.94 (0.67 to 1.33)	263 per 1000	16 fewer per 1000 (from 87 fewer to 87 more)
Unscheduled care (hospitalisation per patient)	252 (1 study) 24 months	MODERATE[a] due to risk of bias	-	The mean number of hospitalisations per patient in the control group was 2.4	The mean number of hospitalisations per patient in the intervention group was 0.30 lower (0.81 lower to 0.21 higher)
Admission to care facility	694	VERY LOW[a,b,d]	OR 0.78	134 per 1000	26 fewer more per 1000

Outcomes	No. of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with Community holistic assessment (95% CI)
	(2 studies) 12-24 months	due to risk of bias, inconsistency, imprecision	(0.49 to 1.22)		(from 63 fewer to 25 more) Clinical benefit of intervention
Admission to care facility (time to admission to care facility)	972 (3 studies) 12-24 months	VERY LOW[a,c,d] due to risk of bias, indirectness, imprecision	HR 0.71 (0.48 to 1.05)	- [e]	- [e]

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 or 2 increments because the point estimate varies widely across studies
(c) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(d) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
(e) Could not be calculated as adjusted raw data was not provided

10.2.3 Economic evidence

Published literature

Four economic evaluations were identified comparing holistic assessment to usual care[36,71,139,226]. These are summarised in the economic evidence profile below (Table 146) and the economic evidence tables in Appendix I.

One economic evaluation relating to this review question was identified but was excluded due to limited applicability[179]. This is listed in Appendix M, with reasons for exclusion given.

See also the economic article selection flow chart in Appendix F.

Table 146: Economic evidence profile: outpatient holistic assessment versus usual care

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
MacNeil Vroomen 2012[139] (Netherlands)	Partially applicable [a]	Potentially serious limitations [b]	• Within-trial analysis (RCT, Boorsma 2011)[26] • Population: Residential care facility residents, with physical or cognitive disabilities. • Comparators: 1. Usual care 2. Multidisciplinary Integrated Care (comprehensive geriatric assessment • Follow-up: 6 months	£305[c]	0.00 QALYs [d]	Intervention 1 dominates intervention 2	Bootstrapping undertaken to estimate uncertainty surrounding ICER. Probability Intervention 2 cost-effective (£20K): ~5% (from graph) Sensitivity analyses were undertaken:-Including only the complete cases in the analysis. -Including only the licensing and subscription costs associated with InterRAI. - Including people who provided no baseline data or died, with missing cost and effect data imputed (intention-to treat). None of these analyses resulted in a change in the conclusion regarding cost-effectiveness
Brettschneider 2015[36] (Germany)	Partially applicable [e]	Potentially serious limitations [f]	• Within-trial analysis (same associated RCT) • Population: Community dwelling adults 80 years or older. • Comparators: 1. Usual care 2. Holistic assessment • Follow-up: 18 months	£648[g]	0.0061 QALYs[h]	£106,229 per QALY gained	Probability of holistic assessment cost effective at a threshold of 50,000 euros per QALY using a societal perspective was 15%. When only patients with complete data were used holistic assessment was more costly and less effective than usual care.

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Ekdahl 2015[71] (Sweden)	Partially applicable [i]	Potentially serious limitations [j]	• Within-trial analysis (same associated RCT) • Population: Community dwelling adults aged 75 years or older who had received inpatient care 3 or more times in the past 12 months and had 3 or more concomitant medical diagnoses. • Comparators: 1. Usual care 2. Outpatient high-intensity holistic assessment • Follow-up: 24 months	£1,781 [k]	See associated clinical paper	NR (mortality and EQ5D reported but not combined into QALYs)	Using alternative methods for missing data replacement did not lead to any change in the conclusion on the EQ5D data.
NCGC model	Partially applicable [l]	Potentially serious limitations [m]	• Markov model based on the study by Frese et al (2012)[82] included in the clinical review • Population reflects the population in the clinical study • Comparators: 1. Usual care 2. Holistic Assessment Lifetime horizon See section 10.2.4 and Appendix N for more details.	£781	0.3239 QALYs	£2411 per QALY	Probability that holistic assessment is most cost-effective at a £20k per QALY threshold is 99%. Results did not change in a series of sensitivity analysis:

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; RCT: randomised controlled trial

(a) Dutch resource use data (2006-7) and unit costs (2007) may not reflect current NHS context. Residential care facility residents aged >65 years, may not reflect all people with multimorbidity

(b) Time horizon may not be sufficient to capture all benefits and costs if benefits persist beyond 6 months. QALYs calculated from SF-12/SF-6D rather than EQ-5D. Within-trial analysis and so does not reflect full body of available evidence for this comparison; Boorsma 2011 is 1 of 28 studies included in the clinical review for comprehensive geriatric assessments.

(c) 2007 Netherlands euros converted to UK pounds.[184] Cost components incorporated: Informal care, primary and secondary care, medication use and costs associated with the interventions

(d) SF-12 collected at baseline and 6 months follow-up, QALYs were calculated by converting SF-12 into SF-6D utility values.

(e) German resource use data (2007-08) and unit costs (2008) may not reflect current NHS context. Community dwelling adults aged >80 years may not represent all people with multimorbidity.

(f) Time horizon may not be sufficient to capture all benefits and costs if benefits persist beyond 18 months. Within-trial analysis and so does not reflect full body of available evidence for this comparison.

(g) 2008 German euros converted to UK pounds.[184] Cost components incorporated: intervention costs (assessment, case conference, home visit), inpatient services, outpatient services (including GP), medication, medical devices, ambulatory care. Cost of nursing home care, informal care, modification of buildings, transportation not included in this analysis.

(h) QALYs adjusted for study region, age, gender, baseline HRQoL, cost at baseline, by means of OLS regression. EQ5D data used to estimate QALYs at 18 months using linear interpolation between measurement points.

(i) Swedish resource use data may not reflect current NHS context; conversion rate used to GBP not reported.

(j) Within-trial analysis and so does not reflect full body of available evidence for this comparison. No QALYs reported.

(k) 2014 UK pounds. Cost components incorporated: Intervention costs, other ambulatory care in hospital, primary health care, inpatient care. Cost of home help services and institutional living not included here.

(l) Indirect population in the clinical study informing the effectiveness data.

(m) Limitations in the clinical evidence informing the clinical data; transition between home or care home settings was not incorporated into the model.

Table 147: Economic evidence profile: inpatient holistic assessment versus usual care

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Tanajewski 2015[226] (UK)	Partially applicable [a]	Potentially serious limitations [b]	• Within-trial analysis (RCT, Edmans 2013[70]) • Population: Patients discharged from an acute medical unit within 72 hours of attending hospital, aged 70 or over, and identified as being at heightened risk of future health problems. • Comparators:	£302 [c]	-0.001 [d]	Intervention 1 dominates intervention 2	Probability Intervention 2 cost-effective (£20K): 0% Using only the complete (adjusted) case data, the incremental cost and QALYs of Intervention 2 vs 1 are respectively £235 and 0.002, with a resulting ICER of £116,326 per QALY gained. The probability of Intervention 2 being cost effective at £20k threshold is 1%.

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			1. Usual care 2. Inpatient holistic assessment • Follow-up: 90 days				

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; RCT: randomised controlled trial

(a) Patients may not represent all people with multimorbidity

(b) Time horizon may not be sufficient to capture all benefits and costs if benefits persist beyond 90 days. Within-trial analysis and so does not reflect full body of available evidence for this comparison. Unclear if social care costs include only the assessment and care plan formulation or also other modifications.

(c) 2012 UK pounds. Cost components incorporated: intervention cost (geriatrician time), primary care services, ambulance services, hospital care, social care (assessments and care plans including home, day, residential and telephone care, housing and meals-on-wheels). The cost of delivering the intervention was £208. Missing data imputed using multiple imputations by chained equations. Adjusted costs estimated using ordinary least squares (OLS) method controlling for age, sex, hospital location and baseline utility.

(d) Adjusted EQ5D data using ordinary least squares (OLS) method controlling for age, sex, hospital location and baseline utility. Missing data imputed using multiple imputations by chained equations.

10.2.4 Economic modelling

The full economic write-up which details all assumptions and model inputs can be found in Appendix N. A summary of the model and its results is provided in the sections below.

10.2.4.1 Model overview

The model compared community low intensity holistic assessment (HA) to usual care. The details of the intervention (HA) were obtained from the clinical study which contributed the most to the clinical outcomes, that is, the study which had the highest weight in the meta-analysis on mortality. In this study by Frese et al (2012)[82] people in the HA arm received an assessment from a nurse, followed by the formulation and agreement of a care plan which is jointly done by a GP and a nurse. The usual care arm received no assessment or care plan. Few patients in the clinical study had a repeated HA, therefore in a sensitivity analysis we assumed the HA was repeated every year for the first three years.
The analysis follows the standard assumptions of the NICE reference case including discounting at 3.5% for costs and health effects, and the NHS and personal and social services perspective; a lifetime horizon was chosen to take into account the mortality outcome.
The model is a Markov model where people start either at home, in a residential care home, or in a nursing care home. They will then move to the 'Death' state according to the intervention-specific probability. There is no other possible transition between 'at home', 'residential care home', and 'nursing care home' states because no evidence was found to inform these transition probabilities. Mortality specific to the residential status of individuals could not be incorporated into the model and this was considered independent from the setting. Costs and QALYs are accrued in each cycle based on the proportion of individuals in each health state, and according to the cost of the intervention in cycle 0.

Key data and assumptions

Initial population settings were obtained from Richardson et al (2011)[194]: 92.3% of the population would start in the community and the remaining 7.7% would start in a care home (equally split between a residential and nursing care home), and the male-female ratio was 36%/ 64%. The initial age of the population was 80 years, based on the main source of effectiveness data.[82]
The population in the model was defined in terms of number of conditions or number of medications, these two characteristics being linked. The link was created using the data from Barnett et al. (2012),[14] a cross-sectional study where a database of 1,751,841 people registered with 314 medical practices in Scotland[115] was used to extract data on multimorbidity. In the subgroup of people aged 65 and over, those with multimorbidity (defined as two or more conditions) had an average of 5 prescriptions, while those without multimorbidity had an average of 3 prescriptions. Baseline mortality was taken from the UK National Life Tables provided by the Office of National Statistics (ONS)[183] and then adjusted to account for this being a population with multimorbidity using the HR of 1.18 for each additional medication reported in the study by Ahmad et al (2005).[4]

The mortality rate (and associated transition probability) for the HA arm was calculated by applying the HR of 0.78 from the Frese 2012 study[82] to the baseline mortality in the usual care arm. This study was selected to inform the effectiveness data in the model as the intervention described in this study reflected the type of HA that the GDG considered for recommendation as it is a low intensity intervention with a potential impact on effectiveness; furthermore this study had the longest of follow-up and was the main contributor to the measure of clinical effectiveness in the clinical review. When this study was combined with the other long-term follow up (>24 months) study by Senior et al (2014)[214] the resulting HR was very similar (0.79).

Utility values were obtained by Heyworth et al (2009)[106] which reported EQ5D values specific to the number of chronic conditions (see Table 148 below). In the base case this is linked to the average number of conditions in the MM population as defined in the population setting (4 conditions on average).

Table 148: Weighted EQ5D scores for people with MM

# chronic conditions	# people	EQ5D	Weighted EQ5D for general MM population
0	2934	0.83	-
1	1209	0.69	-
2	510	0.61	0.4489
3	164	0.5	0.1183
4	15	0.39	0.0084
5	4	0.25	0.0014
Average EQ5D score for MM population			0.5771

The intervention cost was based on its definition and the health care professionals involved at every stage, the costing for each component and the resulting total cost are reported in Table 149 below.

Table 149: Cost of holistic assessment

Component	Health care professional involved	Cost per hour (a)	Time required	Total cost
Assessment	Community nurse – band 6	£57[b]	1 hour	£57
Formulation of care plan	Community nurse – band 6	£57[b]	0.5 hour	£28.5
	GP	£109[c]	0.5 hours	£54.5
TOTAL				**£140**

(a) Source: PSSRU 2014[57]
(b) Cost of one hour of patient-related work
(c) Cost of general medical service contract activity, excluding travel time costs and direct care staff costs.

The estimated cost of £140 for each HA was applied to each individual in the HA arm in the model, while no cost was applied to the usual care arm since the intervention is assumed to be implemented in addition to usual care.

Carrying out the intervention may as well lead to additional cost if a need for further care or a change in management is identified as a consequence of the holistic assessment. The quantification of these costs was discussed extensively with the GDG and it was agreed that there was no point estimate or range that could be used with some degree of certainty. This is because each individual patient may require expensive further care or none at all and it was difficult to decide on the cost of different levels of care and the proportion of people receiving it. Due to these difficulties, we decided to assess the impact of this cost in a threshold analysis where we varied the additional cost from £0 to any positive value.

The cost of the 'own home' health state was based on the average number of consultations with GP, practice nurse, and community nurse for people living in their own home from a Scottish national statistics dataset[115] and was a total of £37 per month. The costs of people living in a nursing or residential care were based on PSSRU publications and are reported in Table 150 below.

Table 150: Resource use and cost for the 'nursing care home' and residential care home' health states

	Cost– Nursing care home	Cost– Residential care home
GP services (per week)	£31[a]	£20[a]
Community nurse (per week)	£0.81[a]	£9.5[a]
NHS contribution to nursing care (per week)	£110[b]	-
Total per week	£141.81	£29.5
Total per month	**£615**	**£128**

(a) Source: PSSRU 2010[56] uplifted using inflation index from PSSRU 2014[57]
(b) Source: PSSRU 2014[57]

10.2.4.2 Results

The base case analysis was run both deterministically and probabilistically. The probabilistic results are reported in Table 151 below.

Table 151: Probabilistic base case analysis results (mean per patient)

	Mean cost	Mean QALYs	ICER (£/QALY)	Probability that strategy is most cost-effective [£20k per QALY]
Usual care	£4,704	2.3764		1%
Holistic Assessment	£5,484	2.7003		99%
Incremental	*£781*	*0.3239*	**2,411**	

The results show that HA is more costly but also more effective than usual care in the base case and the increase in cost is acceptable at the NICE threshold of £20,000 per QALY gained as the ICER is £2,411, well below this threshold. The QALY gain produced by HA is a consequence of a decrease in mortality associated with this intervention, while no increase in quality of life was observed. Mean undiscounted life years in the intervention arm were 8.61 while this value was 7.43 in the usual care arm.

Many assumptions and parameters were tested in a series of sensitivity analyses; throughout these analyses HA remained cost effective under reasonable parameter values. The sensitivity analysis on the relative effectiveness of HA compared to usual care was deemed to be the most important SA conducted for the model as the main driver of the cost effectiveness results was the effectiveness of the intervention at reducing mortality. If the mortality HR is at least 0.994, which corresponds to a very small improvement in survival, HA is cost effective. Although this value is much higher than the base case value (0.78) this was considered by the GDG to be probably an overestimate as what it was expected from this intervention from a clinical point of you was possible a change in quality of life but no significant improvement in survival.

The model had potentially serious limitations which were mainly due to a lack of data or a poor quality of data. We also had no data on the potential cost of a change in management as a consequence of holistic assessment. A sensitivity analysis highlighted this would have to be a considerable ongoing cost for HA not to be cost effective, therefore this was not considered a major limitation per se in the model.

The major limitation that made the GDG less confident in the model conclusions was the source of the effectiveness data. The only clinical outcome incorporated into the model was mortality and this was based mainly on one study[82] as other studies did not have a long enough follow up time. This study had some important limitations: although the study was randomised, the authors employed a stratified randomisation procedure. The intervention group was composed of a stratified sample of

patients randomly selected from 6 “health states” specified by the authors. The control group was composed of the remaining patients who had been recruited to the study. The result was that the control group was predominantly composed of patients from the “less healthy” health states compared to the intervention group. Although the authors adjusted their analyses for some potential confounding factors, including health states, the GDG had concerns that this would not completely address the risk of bias from the population differences at baseline.

Due to the low credibility on the effectiveness estimate and on any reduction in mortality generated by the intervention, the GDG were cautious on the main findings of the model.

10.2.5 Evidence statements

Clinical

- 25 studies comprising a total of 11038 people evaluated inpatient holistic assessment compared to usual care. The evidence demonstrated some limited benefit of an inpatient holistic assessment compared to usual care for some critical outcomes such as mortality (very low quality evidence from 17 studies comprising 6709 people), however there was no benefit for other critical outcomes such as quality of life (low quality evidence from 1 study comprising 285 people). The majority of the evidence was either low or very low quality.
- 12 studies comprising a total of 5813 people evaluated community holistic assessment compared to usual care. Again the evidence demonstrated some limited benefit of community holistic assessment compared to usual care for some critical outcomes. The GDG noted that community holistic assessment – low intensity improved outcomes such as mortality at 74 month follow-up (low quality evidence from 1 study comprising 1620 people) but generally had less impact on quality of life (low quality evidence from 1 study comprising 234 people).

Economic

Inpatient holistic assessment

- One cost–utility analysis found that usual care was dominant compared to holistic assessment for people with multimorbidity. This analysis was assessed as partially applicable with potentially serious limitations.

Community holistic assessment

- One cost–utility analysis found that usual care was dominant compared to holistic assessment for people with multimorbidity. This analysis was assessed as partially applicable with potentially serious limitations.
- Another cost-utility analysis found that holistic assessment was not cost effective compared to usual care in people with multimorbidity (ICER: £106,229 per QALY gained). This analysis was assessed as partially applicable with potentially serious limitations.
- One cost–consequence analysis found that holistic assessment was more costly and more effective than usual care for patients with multimorbidity (£1781 more per patient) and had 0.082 fewer deaths per patient, but has a lower EQ5D score. This analysis was assessed as partially applicable with potentially serious limitations.
- One original cost–utility analysis found that holistic assessment was cost effective compared to usual care for people with multimorbidity (ICER: £2411 per QALY gained). This analysis was assessed as partially applicable with potentially serious limitations.

10.2.6 Recommendations and link to evidence

Recommendations	**35. Start a comprehensive assessment of older people with complex needs at the point of admission and preferably in a specialist unit for older people. [This recommendation is from the NICE guideline on transition between inpatient hospital settings and community or care home settings for adults with social care needs.]**
Research Recommendation	**3. What is the clinical and cost effectiveness of a community holistic assessment and intervention for people living with high levels of multimorbidity?**
Description of current UK services	There are a number of existing ways in which holistic assessment and care planning currently happen, including various forms of case management, but these vary considerably in their scope and triggers, with many being condition specific. There is also a significant variability in how accessible and utilised these assessments are, and no specific recommended method of assessment or treatment planning to guide practice. Examples include case management of people at high risk of emergency hospital admission by community matrons, and various forms of multidisciplinary Comprehensive Geriatric Assessment (CGA) usually carried out in hospital settings. The GDG was not aware of any commonly used intervention similar to CGA for people aged less than 65 years outside of single condition case management (for example, in mental health care) and programmes targeting people at high risk of emergency hospital admission.
Relative values of different outcomes	The GDG considered health-related quality of life, mortality, functional outcomes, patient and carer satisfaction, length of hospital stay, unscheduled care and admission to care facility as critical outcomes for evaluating the effectiveness of holistic assessment. The GDG also considered continuity of care and patient and carer treatment burden as important outcomes.
Trade-off between clinical benefits and harms	**Holistic assessment inpatient ward** The evidence demonstrated a clinical benefit of ward based inpatient holistic assessment compared to usual care for quality of life – physical functioning, quality of life – energy, quality of life – mental health, quality of life – general health, mortality, functional outcomes – proportion of people improving/worsening in ADL score or achieving independence, family/resident satisfaction, unscheduled care – emergency department presentations, admission to care facility and carer burden – self-rated health. The evidence demonstrated no clinical difference of ward based inpatient holistic assessment compared to usual care for quality of life – physical limitations, quality of life – emotional limitations, quality of life – bodily pain, quality of life – social activity, functional outcomes – by mean scores on validated scales or dependent on ADL/IADL, caregiver satisfaction, patient satisfaction, length of stay, unscheduled care – hospital readmissions and carer burden – emotional health **Holistic assessment inpatient team** The evidence demonstrated a clinical benefit of team based inpatient holistic assessment compared to usual care for mortality, mortality – time to event, functional outcomes – Barthel score ≥17, functional outcomes – improvement on OARS iADL and patient treatment burden and admission to care facility. The evidence demonstrated no clinical difference of team based inpatient holistic assessment compared to usual care for quality of life – Philadelphia geriatric centre morale scale, quality of life – EQ-5D, functional outcomes – ADL, functional outcomes – improvement on Katz ADL, length of hospital stay and hospital readmissions. **Holistic assessment inpatient summary**

The GDG agreed that reductions in mortality and admission to care facility were important to people with multimorbidity. Therefore the GDG agreed that, despite some inconsistency and a lack of findings for other critical outcomes, overall inpatient holistic assessment may provide a clinical benefit to older adults with multimorbidity.

Community holistic assessment – low intensity

The evidence demonstrated a clinical benefit of low intensity community holistic assessment compared to usual care for mortality – 12-24 month follow-up, mortality – 74 month follow-up, mortality – time to event, admission to care facility - 12-24 month follow-up, admission to care facility – 74 month follow-up and admission to care facility – time to event. The GDG had particular concerns over the quality of the long term mortality evidence as summarised in the quality of evidence section below.

The evidence demonstrated no clinical difference of low intensity community holistic assessment compared to usual care for quality of life and functional outcome – Barthel index.

The evidence demonstrated a clinical harm of low intensity community holistic assessment compared to usual care for mortality - 0-12 month follow-up and hospitalisation.

Community holistic assessment – high intensity

The evidence demonstrated a clinical benefit of high intensity community holistic assessment compared to usual care for quality of life – mental health, quality of life – SF-36 general health, role, vitality, mental health and social functioning, mortality – 0-12 month follow-up, mortality – 12-24 month follow-up, mortality – time to event, admission to care facility and time to admission to care facility.

The evidence demonstrated no clinical difference of high intensity community holistic assessment compared to usual care for quality of life – EQ-5D, quality of life – physical performance, quality of life – role functioning, quality of life – SF-36 physical, mental, bodily pain and emotional components, functional outcomes – Katz ADL, Lawton & Brody IADL and sickness impact profile, patient and carer satisfaction, length of hospital stay, hospitalisation.

The evidence demonstrated a clinical harm of high intensity community holistic assessment compared to usual care for mortality 24-36 month follow-up.

Community holistic assessment - summary

While the GDG noted that the evidence indicates that holistic assessment has no impact on use of unscheduled care, patients' functional outcomes and limited impact on quality of life, they noted that the intervention has a clinically important impact on mortality and admission to care facility. The evidence indicated that holistic assessment may reduce the risk of mortality and may reduce the risk of care home admissions. The GDG felt that, based on their clinical experience, there may be other benefits of the intervention, such as a reduction in adverse events by proactively identifying risk factors rather than reactively caring for their consequences.

The GDG noted that the group of older adults with multimorbidity living in the community are likely to be healthier than the inpatient population and therefore have a lower risk of negative outcomes, such as mortality and admission to care facility. However, the GDG thought that a holistic assessment could still potentially be of use in the group living in the community, as undertaking the assessment could identify serious issues likely to lead to hospital admission and thereby act preventatively.

The GDG discussed whether they could generalise the evidence in the review to a wider multimorbid population as the majority of evidence came from an older adult population. The GDG felt that the evidence for care home admission could not be

	generalised, as very few younger adults with multimorbidity will be admitted to long-term care. As a consequence the GDG felt that there was insufficient evidence in this review to recommend community holistic assessment to a wider population of adults with multimorbidity. However, the GDG felt that within the younger adults with multimorbidity population, the higher risk individuals may still benefit.
Economic considerations	**Holistic assessment inpatient:** One relevant economic evaluation was included that compared inpatient holistic assessment to usual care; this was a UK study based on an RCT[70] included in our clinical review. The study showed that inpatient holistic assessment was more costly (£302 per patient) and not more effective than usual care, where effectiveness was measured as EQ5D adjusted by different baseline characteristics. This analysis was limited for a series of factors: the population may not reflect the guideline population as it included people discharged from an acute medical unit within 72 hours of attending hospital, aged 70 or over, and identified as being at heightened risk of future health problems. Also the time horizon was very short (90 days) and would not capture any benefits that may persist beyond 90 days. **Community holistic assessment:** Three relevant economic evaluations were included that compared outpatient comprehensive geriatric assessment to usual care.[36,71,139] All the studies were based on RCTs included in the clinical review and only assessed the cost effectiveness of holistic assessment based on the quality of life outcome, while mortality was not considered in these studies. Their conclusion was that the improvement in quality of life was not significant and it did not justify the increase in costs. The GDG noted that the studies had limited applicability due to the settings or definition of the included population which may not reflect the guideline population A major limitations of these studies was the short time horizon which may not be sufficient to capture all benefits and costs if benefits persist beyond the follow up time. An original economic analysis was conducted on this question. This was based on the clinical review conducted for this guideline, the effectiveness estimate derived by Frese et al (2012), the main included study, which also determined the definition of the intervention. This consisted of an assessment from a nurse, followed by the formulation and agreement of a care plan which is jointly done by a GP and a nurse. This was considered a low intensity intervention with a cost of £140 per patient. The outcomes of the model were driven by the reduction of mortality observed in the holistic assessment arm. In the base case holistic assessment was more costly but also more effective than usual care; the probability of the intervention being cost effective was 99%. These conclusions were also stable to a series of sensitivity analyses which were conducted on the main parameters and assumptions. The only change in conclusion was observed when the Hazard Ratio for mortality was increased from 0.78 in the base case to 0.994 in a threshold analysis; at this value of HR holistic assessment would not be cost effective anymore. Despite the stable results of the model, the GDG expressed their scepticism especially around the effectiveness estimate which was considered to be an overestimate. In fact, this intervention was expected to have a potential change in quality of life but no significant improvement in survival. Other RCTs and associated economic evaluations showed no improvement in quality of life and a cost ineffectiveness of holistic assessment when quality of life was the outcome considered. For this reason the GDG did not believe the clinical evidence informing the model was robust enough to make a recommendation in favour of holistic assessment for every patient with multimorbidity. Although the intervention was costed at £140, if this was implemented for every patient with multimorbidity in England, this would lead to a high implementation cost and running cost overall.
Quality of evidence	**General:** All outcomes were downgraded for indirectness as it was not clear if all people in the

	studies included were multimorbid. However, the GDG believed that the majority would be within the trials, given their age and the descriptions of their health status. The GDG noted that much of the evidence came from the USA and from studies conducted more than 10 years ago. **Holistic assessment - inpatient:** For holistic assessment inpatient ward, the quality of the evidence varied from moderate to very low. The evidence showing a clinical benefit in terms of some of the SF-36 subscales (physical functioning) was low due to indirectness and imprecision, for others (energy, mental health, general health) it was moderate due to indirectness. The evidence showing a clinical benefit for mortality, family/resident satisfaction and unscheduled care was very low quality due to risk of bias, indirectness and imprecision. The evidence showing a clinical benefit for admission to care facility was moderate quality due to indirectness. For holistic assessment inpatient team, the quality of the evidence varied from low to very low. The evidence showing a clinical benefit in terms of mortality, functional outcomes – Barthel score and admission to care facility was very low quality due to risk of bias, indirectness and imprecision. The evidence showing a clinical benefit in terms of functional outcomes – OARS IADL and patient treatment burden was low quality due to indirectness and imprecision. **Community holistic assessment low intensity:** The quality of the evidence varied from low to very low. The evidence showing a clinical benefit in terms of mortality 12-24 months follow-up and admission to care facility 12-24 months follow-up was very low quality due to risk of bias, indirectness, imprecision. The evidence showing a clinical benefit in terms of mortality 74 month follow-up, mortality time to event, admission to care facility 74 month follow-up and admission to care facility time to event was low quality due to risk of bias and indirectness. The GDG was concerned about the quality of evidence in Frese's study due to the unconventional randomisation strategy that left the intervention and control arms notably different at baseline. Although Frese performed an adjusted analysis, without the details of this analysis and more complete information on how the groups varied at baseline, the GDG did not feel the evidence was high quality enough to make a strong recommendation. **Community holistic assessment high intensity** The quality of the evidence varied from moderate to very low. The evidence showing a clinical benefit in terms of quality of life – mental health, mortality 0-12 month follow up, admission to care facility and time to admission to care facility was very low quality due to risk of bias, indirectness and imprecision. The evidence showing a clinical benefit in terms of quality of life - general health, role, mental health, social functioning, mortality 12-24 month follow-up and time to mortality was low quality due to indirectness and imprecision. The evidence showing a clinical benefit in terms of quality of life - vitality was moderate quality due to indirectness.
Barriers to implementation	The GDG discussed possible barriers to implementation of a holistic assessment for people with multimorbidity. Since the population targeted will have very heterogeneous needs, the GDG noted that the type of care plan developed and implemented following holistic assessment will vary widely, as would the need for repeated assessment and care plan revision. This means that services will need to be flexible to allow for varying length of care and need for follow-up. Furthermore, where care plans specify support across a number of domains of a person's life, a professional or service will need to have clear responsibility for ensuring implementation by collaborating closely with multiple services and health care professionals involved in a person's care to share information and ensure continuity of care. The GDG also noted that services should ensure that healthcare professionals who perform holistic assessment should be adequately trained. The GDG also noted that some structured assessments that services may choose to use within holistic assessment (for example, assessments of physical and psychological function) may

	have an access cost.
Other considerations	**Inpatient holistic assessment** The GDG noted that holistic assessment performed in a dedicated ward would likely have benefits over a mobile team. On a dedicated ward people are likely to have greater contact with their healthcare professionals which would help generate a better relationship and in turn greater patient satisfaction and other outcomes. The GDG decided to make explicit cross reference to the recommendation from NICE guideline on transition between inpatient hospital settings and community or care home settings for adults with social care needs which recommends such care. **Community holistic assessment** On the balance of limited evidence of clinical benefit of community holistic assessment and in the context of concerns regarding the quality of the long term evidence, the GDG chose not to make a recommendation on the use of holistic assessment. However the GDG discussed the principles and partialities of potential holistic assessment interventions as outlined below. The GDG noted that in the community the first point of contact of older adults with multimorbidity may be a GP, carer, district nurse, home help or other community-based healthcare service. The GDG thought that any of these people who come in contact with the outpatient may be able to identify older adults with multimorbidity who may benefit from a community holistic assessment. The GDG felt that a trigger assessment would be helpful to identify people who may benefit from a community holistic assessment, rather than administering the assessment to everyone. The GDG thought that this was important to ensure that only those people who are most likely to benefit from the intervention receive it, therefore minimising unnecessary resource use. The GDG discussed how a short assessment, taking 1-2 minutes, that could be asked by a GP or district nurse in the patient's home or GP practice or community outpatient centre, may trigger a community holistic assessment. The GDG thought that the community holistic assessment should be less resource intensive than the inpatient assessment. This is partly due to the availability of multiple members of staff in an inpatient setting as opposed to in the community but also due to the likely lower risk nature of patients in the community compared with inpatients. The GDG agreed that the community holistic assessment could be undertaken by a suitably skilled health professional, who is competent in all domains of the assessment, (for example, a trained nurse) and could be conducted at the patient's home or GP practice or community outpatient centre. The GDG also noted that some of the outpatient studies repeated the assessment after a certain period of time. Frese (2012) provided a follow-up holistic assessment 3 years after the initial holistic assessment for approximately one third of their intervention population. The GDG felt that it would be appropriate to for holistic assessments to be repeated annually if a person continued to meet triggers for the assessment, but also noted that follow-up holistic assessments may be shorter as much background information would already have been gathered. The GDG agreed to develop a research recommendation to evaluate low intensity holistic assessment in the community. They recognised that similar interventions occur in schemes that target unplanned admissions and such schemes appear to make intuitive sense. There is however no convincing clinical evidence of benefit. The GDG considered that such research should be a priority for the NHS so that robust evidence is available to inform whether such programmes should be implemented. Further details on this proposed research can be found in Appendix O.

11 Self-Management

11.1 Introduction

People with long-term conditions are the most frequent users of healthcare services. Empowering the patient which chronic disease to manage their disease and its effect on their life has been seen as a cornerstone of chronic disease management and it is envisaged that the majority of people with ling-term conditions can be supported to manage their conditions. There is some evidence that self-management programmes can improve health outcomes in individual conditions. This chapter reports on an evidence review that sought to answer the question whether such programmes might work specifically for people with multimorbidity.

11.2 Review question: What is the clinical- and cost-effectiveness of self-management and expert patient programmes for people with multimorbidity?

For full details see review protocol in Appendix C.

Table 152: PICO characteristics of review question

Population	Adults (aged 18 years and over) with multimorbidity Strata: Adults with physical health conditions only; adults with both physical and mental health conditions
Interventions	Interventions delivered either directly to patients with multimorbidity or to healthcare professionals including: • Expert patient programmes • Self-management programmes
Comparisons	A combination of the above Inactive control intervention Usual care
Outcomes	**Critical:** • Health-related quality of life (for example EQ-5D) • Mortality • Functional outcomes (for example mobility, activities of daily living, FIM, or Barthel index, performance status) • Patient and carer satisfaction • Length of hospital stay (including annualised stay) • Unscheduled care (for example crisis appointments, hospital admission, readmission) **Important:** • Continuity of care metrics • Admission to care facility • Patient/carer treatment burden • Patient self-efficacy
Study design	RCTs; cohort studies if no RCTs retrieved

We sought studies evaluating the clinical- and cost-effectiveness of self-management and expert patient programmes for people with multimorbidity. Studies were included if the intervention was delivered to people with multimorbidity and the intervention targeted all of patients' health conditions. As a consequence, interventions targeted at improving patient self-management and outcomes for a single condition amongst people with multimorbidity were excluded. The GDG was

also interested in self-management interventions that aim to improve a person's management of their conditions and ongoing treatment, and chose not to include studies that aimed to improve self-management of acute events (for example, discharge from hospital).

11.3 Clinical evidence

Thirteen RCTs were included in the review [15,24,66-69,85,87,95,109,137,141,186]; these are summarised in Table 153 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 154, Table 157 and Table 158). See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Four studies[24,67,68,137] included participants with only physical health conditions, 6 studies[15,66,69,87,95,109] included participants with comorbid physical and mental health conditions. One study[186] included participants with dementia and the GDG felt that this population was distinct and should not be pooled with either strata. Two studies[85,141] were conducted with an older adult population and no information was provided on the proportion of participants with multimorbidity, and whether participants were diagnosed with only physical, or with physical and mental, health conditions.

The interventions included in the review varied in terms of their duration (range: 6 weeks to 22 months), method of delivery (individual sessions or group sessions or a combination; face to face or telephone or a combination), trainer (peer or lay person or healthcare professional), key components (combinations of problem-solving therapy, goal setting, psychological or emotional techniques, skills for liaising with HCPs, exercise and health eating, medication management, managing fatigue, peer support, coordinating services, disease-specific education and coaching), and study populations.

One study[109] aimed to support participants to better manage their treatment; for example, to prepare for appointments and communicate better with health professional, while all other studies aimed to support participants to better manage their health conditions. Due to the different aims of the interventions, the GDG decided to analyse the data from this single study separately.

There were a variety of outcomes reported by the studies. Where the studies did not report outcomes specified in the protocol, we included data from closely related outcomes (for example, self-rated health in place of health-related quality of life).

Table 153: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments
Battersby 2013[15]	**Intervention (n=46):** Flinders Program + alcohol use self-help materials + optional Stanford Chronic Disease Self-Management group programme (SCDSMP). **Method of delivery:** Patient works with a healthcare professional to achieve medical and psychosocial goals. Intervention aims to increase patients' knowledge and understanding of their conditions, and improve their ability to self-manage. Patients also provided with self-help materials to support	Adult male Vietnam veterans with multimorbidity (mean age (years): intervention = 60.55, comparison = 60.18). Multimorbidity not clearly reported. High presence of cardiovascular	Assessment of Quality of Life (AQOL). Self-efficacy, indicated by PIH score.	Unclear how many patients opted to participate in SCDSMP. Comorbid physical and mental health conditions.

Study	Intervention and comparison	Population	Outcomes	Comments
	management of alcohol use. Patients given the option of participating in the SCDSMP; group sessions conducted weekly for 2.5 hours led by peers or health professionals to improve self-efficacy in role, disease and emotional management. **Duration:** 12-months, with optional 6 week SCDSMP course **Key components:** Motivation, knowledge of conditions, problems solving, decision making, resource utilisation, managing the patient-provider partnership, action planning, emotional management. **Comparison (n=31):** Standard care	conditions (intervention = 72%; control = 80.6%) and depression (intervention = 76%; control = 84%). Other conditions also reported across the sample.		
Blakeman 2014[24]	**Intervention (n=215):** BRIGHT intervention to support self-management in patients with chronic kidney disease (CKD) and other comorbid conditions. **Method of delivery:** Provision of information booklet on CKD, access to PLANS booklet and website, and telephone guided support from a lay health worker. **Duration:** 12-week follow-up **Key components:** Provision of information about CKD to increase awareness of CKD and encourage patients to consider changes to maintain general vascular health; access to PLANS booklet and website to support patients to self-identify their own health and social needs (not condition-specific), and provide links to community resources and local support, as well as advice on lifestyle changes; telephone support at 1- and 5-weeks to guide participants through PLANS and support access to community resources.	UK Patients with a clinical diagnosis of chronic kidney disease. Age: 72.1 (9.1) 99% of sample had 1 or more additional comorbid conditions (42% cardiovascular; 23% diabetes; none other reported).	Health related quality of life – EQ-5D (6 months). Functional outcomes – positive and active engagement (6 months). Functional outcomes – MOS social/role activities (6 months).	Comorbid physical health conditions

Study	Intervention and comparison	Population	Outcomes	Comments
	Control (n=221): Usual care. Patients were sent the kidney information guidebook and the PLANS booklet with links to the website at the end of the trial period.			
	Intervention (telephone; N = 108): Pre- and post-discharge self-management intervention delivered by a nurse care manager and nursing students. **Method of delivery:** Four telephone calls **Duration:** 4-weeks **Key components:** Patients were supported to make decisions and take action to monitor their conditions. The care manager supported participants to identify health maintenance goals (for example nutrition, monitoring of symptoms, medication adherence). The care manager identified and supported participants to overcome barriers to goal attainment. **Control (n=108):** Standard, inpatient care, which included inpatient nursing care, basic health advice, information on medication and adherence, and arrangements for outpatient follow-up.			
Druss 2010[66]	**Intervention (n=41):** HARP, an adaption of the Chronic Disease Self-Management Program (CDSMP). **Method of delivery:** 6 group sessions led by trained mental health peer specialists **Duration:** 6 weeks **Key components:** Overview of self-management; exercise and physical activity; pain and fatigue management; healthy eating on a limited budget; medication	Adults with comorbid mental and physical health conditions (mean age intervention =47.8, comparison = 48.4). Mental health conditions included schizophrenia, bipolar disorder, major depression,	Health related quality of life (physical and mental component) Minutes/week spent in moderate or vigorous exercise. Patient Activation Measure (PAM).	Comorbid physical and mental health conditions

Study	Intervention and comparison	Population	Outcomes	Comments
	management; finding and working with a regular doctor. Development of action plans for choosing domains of health behaviour change. This involves identifying a problem that is of particular concern, listing ideas for solving the problem, developing a plan outlining specific short-term goals for improvement.	and PTSD. Physical health problems included hypertension, arthritis, asthma, and heart disease.		
	Comparison (n=39): Standard care			
Dunbar 2014[67]	**Intervention (n=46):** Integrated heart failure and diabetes education and self-management support. **Method of delivery:** Delivered by trained research nurses **Duration:** Clinic visits 4 weeks, follow-up 90 days **Key components:** Intervention nurse uses flip chart and script for educational sessions with purpose of increasing knowledge and skills related to diet, medication taking, symptom monitoring, physical activity. Patients given an intervention resource notebook which presented all information in written form and additional materials to which they could refer to in the home setting. Follow-up education and counselling for integrated self-care was provided with a 15 minute phone call 48-72 hours after discharge.	Adults (mean age = 59.7 years) diagnosed with comorbid type II diabetes and heart failure.	Audit of Diabetes Dependent Quality of Life (ADDQOL). Minnesota Living with Heart Failure (MLWHF). Perceived Diabetes Self-Management Scale (PDSMS). Self-Care in Heart Failure Index (SCHFI)	Comorbid physical health conditions
	Comparison (n=19): Standard care			

Study	Intervention and comparison	Population	Outcomes	Comments
Dunbar 2015[68]	**Intervention (n=54):** Integrated heart failure and diabetes intervention **Method of delivery:** 7 sessions. Educational strategies included: Individual teaching and discussion with illustrated content; Coordinated written materials; Health literacy: 6th-grade reading level and multiple illustrations; Demonstration, return demonstration (for example, label reading for portion, sodium, carbohydrates, symptom and self-monitoring interpretation); Questions and answers; Repetition of content, recheck of learning (follow-up home and clinic visits, telephone calls). Behavioural strategies included: Goal setting and evaluation; Symptom and self-monitoring; Problem solving; Seeking support; Motivational messages. **Duration:** 4.5 months **Key components:** education on the following content: HF and diabetes, and how these interact; how to care for HF and diabetes; dietary principles; medication goals; potential medication conflicts; behaviour to promote medication adherence; symptom monitoring; physical activity; oral and foot care **Comparison (n=54):** Provided with informational brochures on "Taking Control of Your Heart Failure" (developed by the Heart Failure Society of America) and "Four Steps to Control Your Diabetes for Life" (developed by the National Diabetes Education Program).	USA Adults (mean age 57.4±10.6; range 21-80) with heart failure and type II diabetes	EQ-5D at 6 months Six minute walk test at 6 months Community Healthy Activities Model Program for Seniors (CHAMPS) score >6 at 6 months	
Eakin 2007[69]	**Intervention (n=101):** Diet and physical activity intervention with self-management support **Method of delivery:** 2 face-to-face (60-90 minutes) meetings with health educator 3 months apart, 3 follow-up calls,	Adults with multimorbidity (mean age = 49.51 years; ≥80% with more than one chronic condition including	Change in minutes walking per week	Comorbid physical and mental health conditions (proportion of patients with mental health conditions

Study	Intervention and comparison	Population	Outcomes	Comments
	and 3 newsletters tailored to the behavioural goals of each participant. **Duration:** 6 months **Key components:** Identification of self-management goal, action planning, assessment of goal attainment, assessment of goal attainment, problem solving of barriers. **Comparison (n=99):** Standard care plus mailed a local area community resources guide and 3 newsletters on basic financial management (careers and employment, budgeting skills, and establishing credit).	hypertension, chronic pain, hypercholesterolemia, depression, type 2 diabetes, osteoarthritis, obesity, chronic lung disease, heart disease, osteoporosis, hepatitis, history of cancer, previous stroke, multiple sclerosis).		unclear).
Friedman 2014[85]	**Intervention (n=382)** Home visiting nurse patient empowerment in chronic disease self-management. **Method of delivery** Home visiting nurse visited patients for an hour in their home once a month for 24 months or until study withdrawal. HVNs empowered patients and educated them on using behaviour change models to facilitate chronic disease self-management. HVNs reviewed medication at each visit. HVNs used the PRECEDE-PROCEED health education planning model to organise disease prevention, health promotion, chronic disease self-management and health behaviour change. There was "often" telephone follow-up after the home visit. Hands-on nursing care (for example, dressing changes) was minimal unless the patient was high risk. HVNs had prior special training in geriatrics and exercise education. **Duration** 24 months or until death/withdrawal. **Key components**	Older adults generally although not exclusively with chronic conditions (Mean age 77, mean number of conditions 4.4, SD 2.2).	Mortality at 24 months Some difficulty with ADLs Great difficulty with ADLs	Not exclusively co-morbid patients but mean number of conditions 4.4, no mention of nature of chronic conditions (including whether mental or physical).

Study	Intervention and comparison	Population	Outcomes	Comments
	Education on behaviour change models, review of medication **Comparison (n=384)** Usual care of all types (hospital, nursing home, home care and ambulatory) as reimbursed by third parties or self-pay. These included home visits as usually provided by Medicare, other third party payers and self-pay.			
Garvey 2015[87]	**Intervention (N = 22):** Occupational therapy led self-management support programme for people with multimorbidity (OPTIMAL) **Method of delivery:** led and facilitated by occupational therapists delivered in primary care. Includes weekly group period for 6 weeks **Duration:** 6 weeks **Key components:** occupational therapy focus; peer support; goal setting and prioritization based on patient preferences. Occupational therapy interventions to support patient self-management used in the groups include: self-management; fatigue and energy management; managing stress and anxiety and maintaining mental health and well-being; keeping physically active; healthy eating; managing medications; effective communication strategies; goal setting. One of the weekly sessions incorporated education on physical activity delivered by a physiotherapist and another incorporated medicines management, delivered by a pharmacist **Comparison (N = 22):** Usual care, placed on a waiting list	Ireland Adults (aged over 18 years) with multimorbidity (two or more chronic conditions) and a minimum 4 repeat prescriptions	Hospital admissions at 2 weeks EQ-VAS at 2 weeks Canadian Occupational Performance Measure (COPM): satisfaction at 2 weeks Occupational Performance Measure (COPM): performance at 2 weeks Nottingham Extended Activities of Daily Living (NEADL) at 2 weeks Stanford Chronic Disease Self-Efficacy 6-item Scale at 2 weeks	
Goldberg 2013[95]	**Intervention (N = 32):** Living well (modified chronic disease self-management programme) **Method of delivery:** 13 sessions of	Adults (mean age = 49.5 years) with a diagnosis of schizophrenia or bipolar disorder	SF-12 general health functioning; physical activity;	Comorbid physical and mental health conditions

Study	Intervention and comparison	Population	Outcomes	Comments
	60-75 minutes, delivered by 2 mental health peers or mental health provider and a peer co-leader. Between sessions peer facilitators telephoned group participants to review progress on their weekly action plan. Two booster sessions following the intervention. **Duration:** 13 weeks **Key components:** Sessions 1-3 focus on basic strategies of self-management including action planning, peer feedback and support, modelling and problem solving. Remaining sessions focus on education and training in specific disease management techniques and the application of these skills to the topics of nutrition, exercise, sleep, medication management, addictive behaviours, and coordination of medical service. Materials including a tool to track action plans and self-management goals and a personal health workbook.	with psychotic features and at least one other general medical condition (for example diabetes, asthma, COPD, cardiovascular disease, arthritis).	unplanned hospital admission; self-management self-efficacy; patient activation.	
	Comparison (N = 31): Usual care			
Hochhalter 2010[109]	**Intervention (N = 26):** 'Making the most of your healthcare' **Method of delivery:** One 2-hour workshop and two follow-up telephone calls delivered by 'coaches'. **Duration:** 6-months **Key components:** The intervention offered tools and taught skills to (a) prepare for healthcare appointments, (b) communicate effectively during healthcare appointments, and (c) follow through on plans of care. Coaches and participants took part in a brief coaching phone call within a week before a scheduled appointment and another call within a week after that appointment. Participants received print copies	Older adults with multimorbidity (aged >65 years; treated for at least two of seven chronic illnesses including arthritis, lung disease, heart disease, diabetes, hypertension, depression, osteoporosis).	CDC healthy days measure (HRQOL-14); self-efficacy.	Comorbid physical and mental health conditions (proportion of patients with mental health conditions unclear).

Study	Intervention and comparison	Population	Outcomes	Comments
	of 'A guide for older people: Talking with your doctor, bound for your good health' and a list of local community resources. **Comparison 1 (N = 27):** Control intervention consisting of the same type and number of contacts as the self-management intervention, except with a focus on general safety for older adults. This included arranging the home environment to avoid falls risks and fire risks, identity theft and caregiver stress. **Comparison 2 (N = 26):** Usual care			
Lorig 1999[137]	**Intervention (N = 311):** Chronic Disease Self-Management Program (CDSMP). **Method of delivery:** Weekly 2.5 hour group sessions led by a pair of trained, volunteer lay leaders in a variety of settings. **Duration:** 7 weeks **Key components:** Exercise; use of cognitive symptom management techniques; nutrition; fatigue and sleep management; use of community resources; use of medications; dealing with the emotions of fear, anger, and depression; communication with others including health professionals; problem-solving; and decision-making. Strategies aimed at increasing self-efficacy, including; weekly action planning and feedback, modelling of behaviours and problem-solving by participants for each other, reinterpretation of symptoms as well as several different management techniques, group problem solving, and individual decision-making.	Within trial subgroup of adults with multimorbidity (mean age = 65 years). Unclear how many conditions participants had; conditions included chronic lung disease, heart disease, stroke, or chronic arthritis, and other unspecified conditions.	Self-reported health status; disability; psychological wellbeing (MHI-5, as taken from the SF-36); energy/ fatigue; health distress.	Comorbid physical health conditions

Study	Intervention and comparison	Population	Outcomes	Comments
	Comparison (N = 225): Waiting list control			
Marek 2013[141]	**Intervention (n=137):** Self-management and medication support **Method of delivery:** Nurse care coordinator visited patients at least every 2 weeks and communicated with a pharmacist and patient's physician. Patients received a med planner. **Duration:** 12-months **Key components:** Pharmacy screen and medication review. A nurse care coordinator supported patients to identify health goals and provided education and tools for patients to self-manage their conditions including monitoring of specific signs and symptoms and medication; support for patients to communicate with healthcare professionals and social services. Participants received a med planner device (simple box with separate components for individual medication times over the course of 2-weeks). **Control (n=125):** Pharmacy screen and medication review alongside standard care.	USA Mean age = 79 years No multimorbidity reported	Health related quality of life - SF-36 physical component (12 months). Health related quality of life - SF-36 mental (12 months). Functional outcome - physical performance test (12 months).	The proportion of patients with multimorbidity and whether participants are diagnosed with mental health conditions is unclear. Intervention includes a medication management component.
Park 2014[186]	**Intervention (n=25):** Health coaching self-management programme **Method of delivery:** Twice weekly group-level activities (including a weekly exercise session) and an approach to increasing individual self-management. Each session lasted approximately 1 hour; delivered by pairs of research team members, who were geriatric nurse specialists and trained to provide health coaching strategies. **Duration:** 8 weeks	Older adults in a nursing home with multimorbidity (aged >65 years and a diagnosis of two or more chronic conditions including stroke, dementia, Parkinson's disease).	Self-rated health; health assessment questionnaire; social role/activities limitations; health distress; chronic disease self-efficacy.	Comorbid physical health conditions, includes participants with dementia

Study	Intervention and comparison	Population	Outcomes	Comments
	Key components: Group health education and group exercise in the group-level approach and individual counselling prior to each group session for goal setting in the individual-level approach. Components included: individual health assessment; goal setting and counselling; group discussion; enhancing cognition activities; exercise sessions; and an activity to encourage the facility's cooperation. The exercise sessions consisted of stretching, hands and feet exercise, and joint movement training.			
	Comparison (n=25): Usual care			

11.3.1 Self-management interventions aimed at improving individuals' management of their health conditions

Table 154: Clinical evidence summary: Self-management programmes versus usual care (participants with comorbid physical health conditions)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions (95% CI)
Health related quality of life EQ-5D/EQ-VAS	589 (3 studies) 0.5-6 months	⨁⊖⊖⊖ VERY LOW[a,b,c] due to risk of bias, inconsistency			The mean health related quality of life in the intervention groups was 0.05 standard deviations higher (0.02 to 0.09 higher)
Self-rated health CDSMP questionnaire & National health interview survey. Scale from: 1 to 5 (lower is better)	536 (1 study) 6 months	MODERATE[b] due to risk of bias		The mean self-rated health in the control groups was 0.04	The mean self-rated health in the intervention groups was 0.12 lower (0.24 lower to 0 higher)
Disability Modification of the Health Assessment Questionnaire - disability scale. Scale from: 0 to 3 (lower is better)	536 (1 study) 6 months	MODERATE[b] due to risk of bias		The mean disability in the control group was 0.02	The mean disability in the intervention group was 0.03 lower (0.09 lower to 0.03 higher)
Psychological wellbeing MHI-5, as taken from the SF-36. Scale from: 0 to 5 (higher is better)	536 (1 study) 6 months	MODERATE[b] due to risk of bias		The mean psychological wellbeing in the control group was 0.03	The mean psychological wellbeing in the intervention group was 0.04 higher (0.08 lower to 0.16 higher)
Positive & active engagement in life Scale from: 0 to 100.	374 (1 study) 6 months	LOW[c] due to risk of bias		The mean positive & active engagement in life in the control group was 66.5	The mean positive & active engagement in life in the intervention group was 0 higher (3.2 lower to 3.2 higher)
Role activities limitations Scale from: 0 to 4 (lower is better)	536 (1 study) 6 months	MODERATE[b] due to risk of bias		The mean role activities limitations in the control group was	The mean role activities limitations in the intervention group was 0.07 lower

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions (95% CI)
				0.1	(0.22 lower to 0.08 higher)
Social/role activities Scale from: 0 to 100 (higher is better)	371 (1 study) 6 months	LOW[c] due to risk of bias		The mean social/role activities in the control group was 68.7	The mean social/role activities in the intervention group was 1.85 higher (3.68 lower to 7.38 higher)
Mortality	499 (1 study) 22 months	VERY LOW [a,b] due to risk of bias, imprecision	RR 1.01 (0.7 to 1.47)	Moderate 179 per 1000 No clinical benefit	2 more per 1000 (from 54 fewer to 84 more)
Some difficulty with ADL - Bathing Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.58 (0.37 to 0.91)	- [d] Clinical benefit	
Some difficulty with ADL - Dressing Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.75 (0.48 to 1.17)	- [d] Clinical benefit	
Some difficulty with ADL - Eating Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.84 (0.5 to 1.41)	- [d] No clinical benefit	
Some difficulty with ADL - Toileting Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.7 (0.44 to 1.11)	- [d] Clinical benefit	
Some difficulty with ADL - Transferring Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 1.14 (0.72 to 1.81)	- [d] No clinical benefit	
Some difficulty with ADL - Walking Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.9 (0.53 to 1.53)	- [d] No clinical benefit	

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions (95% CI)
Great difficulty with ADL - Bathing Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.4 (0.2 to 0.8)	- [d] Clinical benefit	
Great difficulty with ADL - Dressing Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.39 (0.18 to 0.85)	- [d] Clinical benefit	
Great difficulty with ADL - Eating Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.36 (0.1 to 1.3)	- [d] Clinical benefit	
Great difficulty with ADL - Toileting Patient interview	232 (1 study) 22 months	VERY LOW[a,b]due to risk of bias, imprecision	OR 0.76 (0.26 to 2.22)	- [d] No clinical benefit	
Great difficulty with ADL - Transferring Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.82 (0.35 to 1.92)	- [d] No clinical benefit	
Great difficulty with ADL - Walking Patient interview	232 (1 study) 22 months	VERY LOW[a,b] due to risk of bias, imprecision	OR 0.76 (0.34 to 1.7)	- [d] No clinical benefit	
Health distress Scale from: 0 to 20 (lower is better)	536 (1 study) 2 months	LOW[c] due to risk of bias		The mean health distress in the control group was 0.8	The mean health distress in the intervention group was 0.16 lower (0.34 lower to 0.02 higher)
Nottingham Extended Activities of Daily Living (NEADL)	44 (1 study) 2 weeks	⨁⊖⊖⊖ VERY LOW[a,b] due to risk of bias, imprecision		The mean nottingham extended activities of daily living (neadl) in the control groups was 40.73	The mean nottingham extended activities of daily living (neadl) in the intervention groups was 6.45 higher (0.23 lower to 13.13 higher)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions (95% CI)
Canadian Occupational Performance Measure (COPM): satisfaction Scale from: 0 to 10.	44 (1 study) 2 weeks	⨁⊖⊖⊖ VERY LOW[a,b] due to risk of bias, imprecision		The mean Canadian occupational performance measure (copm): satisfaction in the control groups was 3.42	The mean Canadian occupational performance measure (copm): satisfaction in the intervention groups was 2.15 higher (1.01 to 3.29 higher)
Canadian Occupational Performance Measure (COPM): performance Scale from: 0 to 10.	44 (1 study) 2 weeks	⨁⊖⊖⊖ VERY LOW[a,b] due to risk of bias, imprecision		The mean Canadian occupational performance measure (copm): performance in the control groups was 4.1	The mean Canadian occupational performance measure (copm): performance in the intervention groups was 1.67 higher (0.72 to 2.62 higher)
Community Healthy Activities Model Program for Seniors (CHAMPS) score >6	108 (1 study) 6 months	⨁⨁⊖⊖ LOW[a,b] due to risk of bias, imprecision	RR 1.33 (1 to 1.77)	556 per 1000	183 more per 1000 (from 0 more to 428 more)
Hospital admissions	44 (1 study) 2 weeks	⨁⨁⊖⊖ LOW[a,b] due to risk of bias, imprecision		The mean hospital admissions in the control groups was 0.15	The mean hospital admissions in the intervention groups was 0.06 higher (0.17 lower to 0.29 higher)
Stanford Chronic Disease Self-Efficacy 6-item Scale Scale from: 0 to 10.	44 (1 study) 2 weeks	⨁⊖⊖⊖ VERY LOW[a,b] due to risk of bias, imprecision		The mean stanford chronic disease self-efficacy 6-item scale in the control groups was 5.32	The mean stanford chronic disease self-efficacy 6-item scale in the intervention groups was 1.47 higher (0.45 to 2.49 higher)

(a) Downgraded by 1 increment as the CI crossed 1 MID
(b) Downgraded by 1 increment as the majority of the evidence was at high risk of bias
(c) Downgraded by 2 increments as the majority of the evidence was at very high risk of bias
(d) Adjusted odds ratios were provided in study but no information on event rates was provided

Table 155: Clinical evidence summary: Self-management programmes versus usual care (participants with comorbid physical and mental health conditions)

Outcomes	Number of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with self-management (conditions) (95% CI)
Health-related quality of life - Physical component HRQOL/SF-36. Scale from: 0 to 100 (higher is better)	137 (2 studies) 2-6 months	LOW[a,b] due to risk of bias, imprecision		The mean health-related quality of life - physical component in the control groups was 35	The mean health-related quality of life - physical component in the intervention groups was 2.95 higher (1.26 lower to 7.17 higher)
Health-related quality of life - Mental component HRQOL/SF-36. Scale from: 0 to 100 (higher is better)	137 (2 studies) 2-6 months	HIGH		The mean health-related quality of life - mental component in the control groups was 40	The mean health-related quality of life - mental component in the intervention groups was 1.11 higher (2.58 lower to 4.8 higher)
Health-related quality of life Assessment of Quality of Life (AQoL)c. Scale from: 0 to 45 (higher is better)	57 (1 study) 18 months	HIGH		Not reported	The mean health-related quality of life in the intervention groups was 0.35 higher (0.14 lower to 0.84 higher)
Physical activity Scale from: 0 to 5 (higher is better)	57 (1 study) 2 months	VERY LOW[b,e] due to risk of bias, indirectness, imprecision		The mean physical activity in the control groups was 2.2	The mean physical activity in the intervention groups was 1 higher (0.32 to 1.68 higher)
Walking change in minutes per week	162 (1 study)	MODERATE[a] due to risk of bias		The mean walking in the control groups was -11	The mean walking in the intervention groups was 27 higher (20.34 to 33.66 higher)
Moderate/vigorous activity minutes per week	162 (1 study)	MODERATE[b] due to		The mean moderate/vigorous activity in the control groups was	The mean moderate/vigorous activity in the intervention groups was 39 higher

Outcomes	Number of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with self-management (conditions) (95% CI)
	6 months	imprecision		152 minutes	(42.17 lower to 120.17 higher)
Use of emergency department	57 (1 study) 2 months	VERY LOW[a,f] due to risk of bias, imprecision	RR 0.39 (0.11 to 1.32)	Moderate	
				276 per 1000	168 fewer per 1000 (from 246 fewer to 88 more)
Self-efficacy Self-management self-efficacy scale. Scale from: 0 to 10 (higher is better)	57 (1 study) 2 months	LOW[a,b] due to risk of bias, imprecision		The mean self-efficacy in the control groups was 6.9	The mean self-efficacy in the intervention groups was 0.3 higher (0.79 lower to 1.39 higher)
Patient activation Patient activation scale/measure. Scale from: 0 to 100 (higher is better)	137 (2 studies) 2-6 months	MODERATE[b] due to imprecision		The mean patient activation in the control groups was 50	The mean patient activation in the intervention groups was 6.71 higher (2.92 to 10.5 higher)

(a) Downgraded by 1 increment as the majority of the evidence was at high risk of bias
(b) Downgraded by 1 increment as the CI crossed 1 MID
(c) Range of the scale is not reported in the paper and the entry here is based on uses of the scale in other studies
(d) Final values not reported
(e) Downgraded by 1 increment as study sample included some participants who did not have multimorbidity
(f) Downgraded by 2 increments as the CI crossed 2 MIDs

Table 156: Clinical evidence summary: Self-management programmes versus usual care (participants with physical health conditions, including participants diagnosed with dementia)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions 95% CI)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (conditions (95% CI)
Self-rated health CDSMP questionnaire & National health interview survey. Scale from: 1 to 5 (lower is better)	43 (1 study) 8 weeks	VERY LOW[a,b] due to risk of bias, imprecision		The mean self-rated health in the control groups was 1	The mean self-rated health in the intervention groups was 0.5 lower (1 lower to 0 higher)
Self-efficacy Chronic disease self-efficacy scale. Scale from: 6 to 60 (higher is better)	43 (1 study) 8 weeks	VERY LOW[a,b] due to risk of bias, imprecision		The mean self-efficacy in the control groups was 28.3	The mean self-efficacy in the intervention groups was 2.3 higher (5.28 lower to 9.88 higher)

(a) Downgraded by 1 increment as the CI crossed 1 MID.
(b) Downgraded by 2 increments as the majority of the evidence was at very high risk of bias

11.3.2 Self-management interventions aimed at improving individuals' management of their treatment

Table 157: Clinical evidence summary: Self-management programmes versus usual care (participants with comorbid physical and mental health conditions)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with self-management (treatment) (95% CI)
Self-efficacy Self-efficacy for managing chronic disease. Scale from: 1 to 10 (higher is better)	41 (1 study) 6 months	MODERATE[a] due to imprecision		The mean self-efficacy in the control groups was 6.9	The mean self-efficacy in the intervention groups was 0.2 higher (0.84 lower to 1.24 higher)

(a) Downgraded by 1 increment as the CI crossed 1 MID

Table 158: Clinical evidence summary: Self-management programmes versus control intervention (participants with comorbid physical and mental health conditions)

Outcomes	Number of	Quality of the	Relative	Anticipated absolute effects

	participants (studies) Follow up	evidence (GRADE)	effect (95% CI)	Risk with control	Risk difference with self-management (95% CI)
Self-efficacy Self-efficacy for managing chronic disease. Scale from: 1 to 10 (higher is better)	43 (1 study) 6 months	LOW[a,b] due to risk of bias, imprecision		The mean self-efficacy in the control groups was 8	The mean self-efficacy in the intervention groups was 0.6 lower (1.53 lower to 0.33 higher)

(a) Downgraded by 1 increment as the majority of the evidence was at high risk of bias.
(b) Downgraded by 1 increment as the CI crossed 1 MID.

Narrative findings

Several studies reported data that could not be analysed as they were presented in an alternative format (for example, data from multilevel modelling) or outcomes were incompletely reported. These data are now presented in narrative form, below.

Self-management versus usual care (patients with comorbid physical conditions)

Health related quality of life

One study[67] used repeated measures multilevel modelling to evaluate the difference between the intervention and control group on the Audit of Diabetes Dependent Quality of Life scale (ADDQOL) and the Minnesota Living with Heart Failure scale (MLWHF) at 90 days. The results indicated no significant difference between the intervention and the control group on either scale (ADDQOL group*time F(1,41) = 0.14, p = 0.71; MLWHF group*time F(1,41) = 0.03, p = 0.86). This study is at high risk of bias.

One study[141] assessed health-related quality of life (physical and mental components; SF-36) at 12-month follow-up in the intervention and control group. After controlling for covariates, mixed model parameter estimates provided for the time*group interaction indicated reduced quality of life (physical component) in the control group compared to the intervention group (b = -0.99, SE = 0.25, t = -3.97, p<.001). The annualised effect size for this (cohen's d) = 0.39, indicating a small effect of group allocation on health-related quality of life. Furthermore, the same study reported reduced health-related quality of life (mental component) for the control group as compared to the intervention group (b = -1.63, SE = 0.30, t = 5.35, p<.001). The annualised effect size for this (cohen's d) = 0.58, indicating a moderate effect of group allocation on health related quality of life. This study was of very low quality, due to risk of bias and indirectness.

Function

One study[141] reported patient function on the physical performance test at 12-months. After controlling for covariates, mixed model parameter estimates provided for the time*group interaction indicated reduced physical functioning in the control group compared to the intervention group (b = -1.01, SE = 0.12, t = -8.21, p<.001). The annualised effect size for this (cohen's d) = 0.66, indicating a moderate effect of group allocation on function. This study was of very low quality, due to risk of bias and indirectness.

Self-efficacy

One study [67] used multilevel modelling to evaluate the difference between the intervention and control group on the Perceived Diabetes Self-management Scale (PDSMS) and the Self-care in Heart Failure Index confidence (SCHFI) at 90 days. The results indicated no significant difference between the intervention and the control group on either scale (PDSMS group*time F(2,94) = 0.11, p = 0.90; SCHFI group*time F(2,95) = 0.04, p = 0.96). This study is at very high risk of bias.

Self-management aimed at improving individuals' management of their health conditions versus placebo (patients with comorbid physical health conditions, including participants diagnosed with dementia)

Health related quality of life

One study[186] reported the mean change in scores on the health assessment questionnaire at 8-weeks for the intervention and control group (scale range = 0-24, high is poor outcome). Findings indicated that there was no significant change in either group, and no significant difference between the intervention and the control group (intervention MD = 0.5 p = 0.383; control MD = 0.5 p = 0.383; effect size = 0.00, p = 0.98). This study is at very high risk of bias.

One study[186] reported the mean change in scores on the social role/activities limitations scale at 8-weeks for the intervention and control group (scale range = 0-16, high is poor outcome). Findings indicated that there was a significant improvement in scores for the intervention group (MD = 5.6, p <.001), and no significant change in scores for the control group (MD = 0.2, p = 0.81). The difference between groups was significant (effect size = 0.16, indicating a small effect of group; p value = 0.008). This study is at very high risk of bias.

Treatment burden

One study[186] reported the mean change in health distress at 8-weeks for the intervention and control group (scale range = 0-20, high is poor outcome). Findings indicated that there was no significant change in either group, and no significant difference between the intervention and the control group (intervention MD = 0.2 p = 0.757; control MD = 0.8 p = 0.585; effect size = 0.00, p = 0.53). This study is at very high risk of bias.

Self-management aimed at improving individuals' management of their treatment versus placebo (patients with comorbid physical and mental health conditions)

One study[109] used mixed linear modelling to evaluate the impact of the intervention on the number of unhealthy days reported by patients (as assessed using the HRQOL) at 6-months compared to usual care. Data indicated no significant difference between the self-management and usual care condition (follow-up*group coefficient = -0.45, SE = 0.49, p = .36). This study is at low risk of bias.

Self-management aimed at improving individuals' management of their treatment versus control intervention (patients with comorbid physical and mental health conditions)

One study[109] used mixed linear modelling to evaluate the impact of the intervention on the number of unhealthy days reported by patients (as assessed using the HRQOL) compared to a control intervention (general safety in older adults). Data indicated no significant difference between the self-management and usual care condition (follow-up*group coefficient = 0.39, SE = 0.51, p = .44). This study is at high risk of bias.

11.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

Two economic evaluations relating to this review question were identified but were excluded due to a combination of applicability and methodological limitations.[15,24] These are listed in Appendix M, with reasons for exclusion given.

See also the economic article selection flow chart in Appendix F.

11.5 Evidence statements

Clinical

Self-management interventions to improve people's management of their health conditions:

Five studies comprising a total of 1922 people evaluated self-management programmes to improve people's management of their physical health conditions.

Very low quality evidence demonstrated no clinical difference between self-management and usual care for mortality, hospital admissions, health-related quality of life, self-rated health, psychological wellbeing, health distress, self-efficacy, positive and active engagement in life, disability, role activities limitations, and social role activities.

Very low quality evidence from 1 RCT with 232 participants demonstrated a clinical benefit of self-management interventions compared to usual care for some activities of daily living (reduction in some difficulty with bathing, dressing and toileting, and a great difficulty with eating), but no clinical difference for others (some difficulty eating, transferring and walking, and a great difficulty with toileting, transferring and walking).

Very low quality evidence from 1 RCT with 108 participants showed a clinical benefit with regards to functioning measured using the Community Healthy Activities Model Program for Seniors (CHAMPS). However, very low quality evidence from 1 RCT with 44 participants demonstrated no clinical benefit when functioning was measured using the Nottingham Extended Activities of Daily Living and the Canadian Occupational Performance Measure (satisfaction and performance).

Very low quality evidence from 1 RCT comprising of 43 people with comorbid physical conditions, including people with dementia, demonstrated a clinical benefit of self-management interventions to improve management of their conditions with regards to self-rated health but no clinical difference in self-efficacy.

Management of co-morbid physical and mental health conditions:

Four studies comprising a total of 420 people evaluated self-management interventions to improve people's management of their comorbid physical and mental health conditions.

Two RCTs with 137 participants demonstrated a clinical benefit of self-management interventions compared to usual care with regards to physical component (low quality evidence), but no clinical difference with regards to the mental component (high quality evidence).

High quality evidence from 1 RCT with 57 participants showed no clinical difference with regards to overall health-related quality of life measured using the Assessment of Quality of Life scale.

Very low quality evidence from 1 RCT with 57 participants demonstrated a clinical benefit with regards to physical activity.

Moderate quality evidence from 1 RCT with 162 participants showed a clinical benefit with regards to walking, but no clinical difference with regards to moderate/vigorous activity.

Moderate quality evidence from 2 RCTs with 137 participants showed a clinical benefit with regards to patient activation.

Self-management interventions aimed at improving peoples' management of their treatment

Very low quality evidence from 1 RCT with 43 participants demonstrated a clinical harm of self-management interventions aimed at improving peoples' management of their treatment compared to a control intervention, and no clinical difference when compared to usual care with regards to self-efficacy.

Economic

- No relevant economic evaluations were identified.

11.5.1 Recommendations and link to evidence

Recommendations	**No recommendations.**
Description of current UK services	Currently, many adults with multimorbidity in the UK are provided with information and advice to promote self-management behaviours. This includes advice on diet and exercise, strategies to improve treatment adherence (including medication management and hospital and appointment attendance), improving knowledge and understanding of individual conditions and potential interactions between conditions and treatments, and strategies for coping with symptoms. A minority of adults with multimorbidity will also attend teaching sessions, led by a healthcare professional or expert patient, to learn self-management strategies. These classes are available to people with certain chronic conditions and are specific to learning self-management strategies for a single condition. Only people with these conditions who also express an interest in learning self-management are referred for these classes. Currently, self-management teaching sessions that are targeted specifically to help people with multimorbidity manage their multiple conditions are not routinely available.
Relative values of different outcomes	The GDG identified health related quality of life, mortality, functional outcomes, patient and carer satisfaction, length of hospital stay, and unscheduled care as critical outcomes for evaluating the efficacy of self-management and expert patient programmes. Continuity of care, admission to care facility, patient and carer treatment burden, and patient self-efficacy were also identified as important outcomes.
Trade-off between clinical benefits and harms	The GDG noted that some of the studies identified evaluated self-management interventions that were aimed at improving people's ability to manage their health conditions, while others were targeted specifically at improving management of treatment. The GDG decided to stratify these interventions. Self-management programmes aiming to improve people's management of their health conditions In people with comorbid physical health conditions, the evidence demonstrated a benefit of self-management for physical functioning, and no clinical difference for

	self-rated health, disability, psychological wellbeing, positive and active engagement in life, role activities limitations, social role activities, health distress and self-efficacy, when compared to usual care. The evidence demonstrated inconsistent findings for health related quality of life; 1 study demonstrated a clinical benefit of self-management on health related quality of life, 1 study demonstrated a small effect of self-management on health related quality of life physical component and a moderate effect of self-management on health related quality of life mental component, and a further study demonstrated no clinical difference between self-management and usual care for health related quality of life (both the diabetes and living with heart failure scales). In people with comorbid physical and mental health conditions, the evidence demonstrated a clinical benefit of self-management for health related quality of life physical component, physical activity, change in walking per week, use of emergency department, and patient activation when compared to usual care. The evidence demonstrated no clinical difference between self-management and usual care for health related quality of life mental component, overall health related quality of life, moderate or vigorous activity, and self-efficacy. In people with comorbid physical conditions, including people with dementia, the evidence demonstrated a clinical benefit for self-management for self-rated health and social role activities, and no clinical difference between self-management and usual care for the health assessment questionnaire, self-efficacy and health distress. <u>Self-management programmes aiming to improve patients' management of their treatment</u> In people with comorbid physical and mental health conditions, the evidence demonstrated no clinical difference between self-management and usual care for number of unhealthy days and self-efficacy. In people with comorbid physical and mental health conditions, the evidence demonstrated a clinical harm of self-management for self-efficacy compared to a control intervention (general safety for older adults), and no clinical difference for the number of unhealthy days. <u>Summary</u> The GDG felt that there was not sufficient evidence in this review to recommend the use of self-management programmes for people with multimorbidity. The GDG noted that many of the interventions included in the review included an exercise component, and suggested that this may have resulted in clinical benefit for outcomes related to physical activity. The GDG discussed whether self-management is a realistic option for many, particularly those who are more elderly or frail or have cognitive problems. Given the prevalence of these characteristics among people whose multimorbidity is difficult to manage the GDG agreed that there is a need for research to develop and evaluate self-management interventions for people with multimorbidity.
Economic considerations	No economic evaluations were identified from the published literature. Although some interventions (which included exercise components) may have a clinical benefit, overall the GDG concluded that the evidence was insufficient to support a conclusion of clinical benefit and therefore no further economic considerations were deemed necessary by the GDG.
Quality of evidence	<u>Self-management programmes aiming to improve patients' management of their health conditions</u> In patients with comorbid physical health conditions, the evidence had the following quality ratings: health-related quality of life at very low quality due to risk of bias and inconsistency; self-rated health at moderate quality due to risk of bias; disability at moderate quality due to risk of bias; psychological wellbeing at moderate quality due to risk of bias; positive and active engagement with life at low quality due to risk of bias; role activities limitations at moderate quality due to risk

of bias; social role activities at low quality due to risk of bias; mortality at very low quality due to risk of bias and imprecision; some and great difficulty with ADL (bathing/dressing/eating/ toileting/transferring/walking) all at very low quality due to risk of bias, and imprecision; health distress at low quality due to risk of bias; Nottingham Extended ADL at very low quality due to risk of bias and imprecision; Canadian Occupational Performance Measures satisfaction/performance both at very low quality due to risk of bias and imprecision; Community Healthy Activities Model Program for Seniors (CHAMPS) score less than 6 at low quality due to risk of bias and imprecision; hospital admissions at low quality due to risk of bias and imprecision; and self-efficacy at very low quality due to risk of bias and imprecision.

In people with comorbid physical and mental health conditions, the evidence had the following quality ratings: health related quality of life physical component at low quality due to risk of bias and imprecision; health related quality of life mental component at high quality; health related quality of life (overall) at high quality; physical activity at very low quality due to risk of bias, indirectness and imprecision; walking at moderate quality due to risk of bias, moderate or vigorous activity at moderate quality due to imprecision, use of emergency department at very low quality due to risk of bias and imprecision, self-efficacy at low quality due to risk of bias and imprecision; and patient activation at moderate quality due to imprecision

In people with comorbid physical health conditions including people with dementia, the evidence had the following quality ratings: self-rated health at very low quality due to risk of bias and imprecision; and self-efficacy at very low quality due to risk of bias and imprecision.

<u>Self-management programmes aiming to improve patients' management of their treatment</u>

In people with comorbid physical and mental health problems for the comparison of self-management and usual care, the evidence had the following quality ratings: self-efficacy at moderate quality due to imprecision.

In people with comorbid physical and mental health problems for the comparison of self-management and control intervention (safety for older adults), the evidence had the following quality ratings: self-efficacy at low quality due to risk of bias and imprecision.

Other considerations

The GDG discussed the importance of empowering a person in managing their health conditions and treatment. The evidence about self-management programmes was not convincing and the GDG agreed that empowering people did not necessarily require a specific programme. They considered that empowering a person to consider their medicines and treatments for example could be achieved by healthcare professionals working with individual people and explaining their choices to them. Information, verbal or written, about conditions, medicines and other treatments would be valuable in helping people take active part in decisions about their health.

12 Format of encounters

12.1 Introduction

People with multimorbidity have by definition more than one condition and the epidemiological evidence indicated that many people with multimorbidity are taking increasing numbers of prescribed medicines. General practice consultations have traditionally been planned around reactive care provided when people present with individual problems. This model of care is increasingly inappropriate for people with multiple conditions, personal characteristics such as frailty and being prescribed multiple medicines. This chapter reports on an evidence review that sought to explore other ways of organising care for this group of people.

12.2 Review question: What format of encounters with healthcare professionals improves outcomes for people with multimorbidity?

For full details see review protocol in Appendix C.

Table 159: PICO characteristics of review question

Population	Adults (aged 18 years and over) with multimorbidity
Intervention(s)	Formats of healthcare encounters targeted at improving outcomes for patients with multimorbidity, as specified in papers. For example, interventions comparing: • Time allocated for consultations (including inpatient care; for example longer time allocation) • Planned recall and structured review • Method of communication (for example face to face, telephone, email, virtual) • Methods of arranging appointments (for example advanced booking, booking with chosen healthcare professional) • Methods to involve patient in planning content of appointments (for example patient setting agenda) • Multi-professional appointments (including ward rounds/clinics) • Setting of encounter (for example community visits) • Combination of the above
Comparison(s)	• Compared to each other • Standard care
Outcomes	**Critical:** • Health-related quality of life • Mortality • Functional outcomes (for example mobility, activities of daily living, FIM, or Barthel index, performance status) • Patient and carer satisfaction • Length of hospital stay (including annualised stay) • Unscheduled care (for example crisis appointments, hospital admission, readmission) **Important:** • Continuity of care (for example information, relationship/provider continuity, appropriate discharge/quality of transition) • Admission to care facility • Patient/carer treatment burden

Study design	RCTs; cohort studies if no RCTs retrieved

12.3 Clinical evidence

Four randomised controlled trials (RCTs) were included in the review,[88,113,180,187,224,225,236] which assessed the efficacy of telemonitoring compared with usual care; these are summarised in Table 160 below. One trial[113] evaluated the effect of telemonitoring only compared to usual care. One trial[187,224,225,236] evaluated the effect of telemonitoring with an alert system (that prompted follow-up) compared to usual care. One trial[180] evaluated the effect of telemonitoring with an alert system and nurse case management compared to usual care with nurse case management. One trial[88]evaluated the effect of telemonitoring, plus self-management, compared to usual care and psychoeducation. Due to variation between the interventions, no data has been pooled together.

Evidence from these studies is summarised in the clinical evidence profiles below.

See also the study selection flow chart in Appendix E, study evidence tables in Appendix H, forest plots in Appendix K, GRADE tables in Appendix J and excluded studies list in Appendix L.

Table 160: Summary of studies included in the review

Study	Intervention and comparison	Population	Outcomes	Comments
Integrated Telehealth Education and Activation of Mood (I-TEAM) study trial: Gellis 2014[88]	Intervention (n=57): Telemonitoring, duration of 3 months. • Daily monitoring of weight, blood pressure, pulse, oxygen saturation and temperature • Nurse contacted participants with abnormal readings for follow-up evaluation • Concomitant treatment: problem-solving treatment; tailored counselling, including medication use, psychoeducation, problem solving strategy and behavioural activation Comparison (n=58): Usual care plus psychoeducation	Older adults (aged 65 years or older) Multimorbidity: 100% (depression and either comorbid heart failure (HF) or chronic obstructive pulmonary disease (COPD)) USA	SF-12 mental component SF-12 physical component Emergency department visits Episodes of care Hospital days Patient satisfaction	
Hopp 2006[113]	Intervention (n=18): Telemonitoring, duration 6 months. Regular video contact between patients and clinical staff, including: • Discussion of the patient's overall health status • Review of medications • Discussion of any health concerns by the patient • Nurse reminders concerning the appropriate self-care	Adults (mean age: intervention 69.8±11.6 years, comparison 69.5±12.7 years) Multimorbidity: number of people with multimorbidity not reported (mean number of conditions:	SF-36V physical component SF-36V mental component Mortality Mean emergency department visits Mean hospital admissions Mean hospital days Patient satisfaction (General Home	

Study	Intervention and comparison	Population	Outcomes	Comments
	behaviours (diet, exercise, monitoring of symptoms including blood pressure and weight) Comparison (n=19): Usual care	intervention 3±1.8, comparison 3.8±1.7; 92% had 1 or more chronic conditions; 68% had 3 or more) USA	Satisfaction Scale)	
Noel 2004[180]	Intervention (n=47): Telemonitoring with alerts, duration 6 months. Plus nurse case management • Collects data for temperature, blood pressure, pulse, blood glucose, 3-lead electrocardiogram, stethoscope for heart and lung sounds, pulse oximetry, and weight • Pain level self-reported • Out of range patient data triggers alert, which is picked up by nurse case manager • Normal range of data determined using patient intake form Comparison (n=19): Usual care plus nurse case management	Adults (18-65 years; mean age 71) Multimorbidity: 79% had 2 or more comorbid conditions USA	OARS Multidimensional Functional assessment: functional level OARS Multidimensional Functional assessment: cognitive status OARS Multidimensional Functional assessment: patient satisfaction OARS Multidimensional Functional assessment: self-rated health status	
Tele-ERA study trial: Takahashi 2012A [187,224,225,236]	Intervention (n=102): Telemonitoring, duration 12 months. • Daily monitoring of symptoms and biometric information • Research nurse contacted participants via telephone or videoconference if alerts arose for follow-up evaluation Comparison (n=103): Usual care	Older adults (aged 60 years or older, mean age 80.3±8.2 years) Multimorbidity: number of people with multimorbidity not reported (mean chronic conditions 3±1.1) USA	SF-12 physical component SF-12 mental component Mortality Barthel ADL index Emergency department (ED) visits Mean ED visits Hospital admissions Mean number of hospital admissions Mean hospital days Mean hospice visits Mean hospice days Time to hospice entry	

Narrative findings

One study[88] measures SF-12 mental component but does not report the findings.

12.3.1 Telemonitoring versus usual care

Table 161: Clinical evidence summary: Telemonitoring versus usual care

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with telemedicine (95% CI)
Mortality	37 (1 study) 6 months	VERY LOW[b,c] due to indirectness, imprecision	RR 1.06 (0.17 to 6.72)	105 per 1000	6 more per 1000 (from 90 fewer to 393 more)
Quality of life (physical component), change score SF-36V. Scale from: 0 to 100, better indicated by higher scores	17 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	The mean quality of life (physical component) in the control groups was 0.64	The mean quality of life (physical component) in the intervention groups was 0.92 higher (6.25 lower to 8.09 higher)
Quality of life (mental component), change score SF-36V. Scale from: 0 to 100.	17 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	The mean quality of life (mental component) in the control groups was -4.11	The mean quality of life (mental component) in the intervention groups was 8.16 higher (1.31 lower to 17.63 higher)
Mean ED visits	37 (1 study) 6 months	MODERATE[b] due to indirectness	-	The mean ED visits in the control groups was 2.11	The mean ED visits in the intervention groups was 1.11 lower (2.55 lower to 0.33 higher)
Mean hospital admissions	37 (1 study) 6 months	LOW[b,c] due to indirectness, imprecision	-	The mean hospital admissions in the control groups was 1.26	The mean hospital admissions in the intervention groups was 0.59 lower (1.61 lower to 0.43 higher)
Mean hospital days	37 (1 study) 6 months	MODERATE[b] due to indirectness	-	The mean hospital days in the control groups was 7.11	The mean hospital days in the intervention groups was 4.28 lower

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with usual care	Risk difference with telemedicine (95% CI)
					(10.37 lower to 1.81 higher)
Patient satisfaction General Home Care Satisfaction Scale, change score No range reported.	37 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	The mean patient satisfaction in the control groups was -1.56	The mean patient satisfaction in the intervention groups was 0.56 higher (2.28 lower to 3.4 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

12.3.2 Telemonitoring with alerts versus usual care

Table 162: Clinical evidence summary: Telemonitoring with alerts versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Usual care	Risk difference with Telemedicine with alerts (95% CI)
Mortality	205 (1 study) 1 years	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	RR 3.79 (1.3 to 11.02)	39 per 1000	108 more per 1000 (from 12 more to 389 more)
Quality of life (physical health) SF-12. Scale from: 0 to 100. Better indicated by higher values.	180 (1 study) 1 years	VERY LOW [a,b,c] due to risk of bias, indirectness, imprecision	-	The mean quality of life (physical health) in the control groups was 34.2	The mean quality of life (physical health) in the intervention groups was 1.4 lower (4.48 lower to 1.68 higher)
Quality of life (mental health) SF-12. Scale from: 0 to 100. Better indicated by higher values.	166 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean quality of life (mental health) in the control groups was 58.1	The mean quality of life (mental health) in the intervention groups was 2.1 lower (4.64 lower to 0.44 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Usual care	Risk difference with Telemedicine with alerts (95% CI)
Activities of daily living Barthel ADL Index. Scale from: 0 to 100. Better indicated by higher values.	166 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean activities of daily living in the control groups was 93.1	The mean activities of daily living in the intervention groups was 2.6 lower (7.22 lower to 2.02 higher)
ER visits	205 (1 study) 1 years	LOW [a,b] due to indirectness, imprecision	RR 1.25 (0.84 to 1.88)	282 per 1000	70 more per 1000 (from 45 fewer to 248 more)
Mean ER visits	205 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean ER visits in the control groups was 0.45	The mean ER visits in the intervention groups was 0.26 higher (0.04 lower to 0.56 higher)
Hospital admissions	205 (1 study) 1 years	LOW [a,b] due to indirectness, imprecision	RR 1.19 (0.89 to 1.59)	437 per 1000	83 more per 1000 (from 48 fewer to 258 more)
Mean hospital admissions	205 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean hospital admissions in the control groups was 0.83	The mean hospital admissions in the intervention groups was 0.27 higher (0.13 lower to 0.67 higher)
Length of hospital stay	205 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean length of hospital stay in the control groups was 6.1 days	The mean length of hospital stay in the intervention groups was 2 lower (6.19 lower to 2.19 higher)
Mean hospice visits	194 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean hospice visits in the control groups was 14.5	The mean hospice visits in the intervention groups was 0.7 lower (6.7 lower to 5.3 higher)
Length of hospice stay	194 (1 study) 1 years	LOW [a,b] due to risk of bias, indirectness	-	The mean length of hospice stay in the control groups was 119.3 days	The mean length of hospice stay in the intervention groups was 61.4 lower

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Usual care	Risk difference with Telemedicine with alerts (95% CI)
					(92.88 to 29.92 lower)
Time to hospice entry	13 (1 study) 1 years	VERY LOW [a,b,c] due to risk of bias, indirectness, imprecision	HR 1.28 (0.94 to 1.74)	1000 per 1000	-

(a) Downgraded by 1 increment as the majority of evidence was at high risk of bias
(b) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

12.3.3 Telemonitoring with alerts (plus case management) versus usual care (plus case management)

Table 163: Clinical evidence summary: Telemonitoring with alerts plus case management versus usual care plus case management

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Usual care plus case management	Standardised mean difference with Telemedicine plus case management (95% CI)
Functional level OARS Multidimensional Functional assessment. Scale from: 0 to 75. Better indicated by higher values.	104 (1 study) 6 months	VERY LOW[a,b] due to risk of bias, indirectness	-	The mean functional level in the control groups was 40.19	The mean functional level in the intervention groups was 0.3 standard deviations lower (0.69 lower to 0.09 higher)
Cognitive status OARS Multidimensional Functional assessment. Scale from: 0 to 50. Better indicated by higher values.	104 (1 study) 6 months	VERY LOW[a,b,c] due to risk of bias, indirectness, imprecision	-	The mean cognitive status in the control groups was 19.68	The mean cognitive status in the intervention groups was 0.02 standard deviations lower (0.36 lower to 0.41 higher)
Patient satisfaction OARS Multidimensional Functional assessment. Scale	104 (1 study) 6 months	VERY LOW[a,b] due to risk of bias, indirectness	-	The mean patient satisfaction in the control groups was 97.14	The mean patient satisfaction in the intervention groups was 0.47 standard deviations higher

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Usual care plus case management	Standardised mean difference with Telemedicine plus case management (95% CI)
from: 0 to 140. Better indicated by higher values.					(0.08 to 0.86 higher)
Self-rated health OARS Multidimensional Functional assessment. Scale from: 0 to 185. Better indicated by higher values.	104 (1 study) 6 months	VERY LOW[a,b] due to risk of bias, indirectness	-	The mean self-rated health in the control groups was 85.14	The mean self-rated health in the intervention groups was 0.18 standard deviations lower (0.57 lower to 0.21 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded because the majority of the evidence included an indirect population (downgrade by 1 increment) or a very indirect population (downgrade by 2 increments)
(c) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

12.3.4 Telemonitoring (plus self-management) versus usual care (plus psychoeducation)

Table 164: Clinical evidence summary: Telemonitoring (plus self-management) versus usual care (plus psychoeducation)

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with telemonitoring (plus problem-solving and counselling) (95% CI)
Quality of life (mental component) SF-12. Scale from: 0 to 100. Better indicated by higher values.	94 (1 study) 6 months	LOW[a,b] due to risk of bias, imprecision	-	The mean quality of life (mental component) in the control groups was 40.3	The mean quality of life (mental component) in the intervention groups was 11.8 higher (1.34 to 22.26 higher)
Mean ER visits	94 (1 study) 12 months	MODERATE[b] due to imprecision	-	The mean ER visits in the control groups was 1.4	The mean ER visits in the intervention groups was 0.8 lower (1.37 to 0.23 lower)
Mean hospital days	94 (1 study)	MODERATE[b] due to	-	The mean hospital days in the control groups was	The mean hospital days in the intervention groups was

Outcomes	Number of participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with telemonitoring (plus problem-solving and counselling) (95% CI)
	12 months	imprecision		10.5	3 lower (5.22 to 0.78 lower)
Mean episodes of care	94 (1 study) 12 months	HIGH	-	The mean episodes of care in the control groups was 1.8	The mean episodes of care in the intervention groups was 0.5 lower (1.01 lower to 0.01 higher)
Patient satisfaction No scale reported. Better indicated by higher values.	94 (1 study) 3 months	LOW[a] due to risk of bias	-	The mean patient satisfaction in the control groups was 4.5	The mean patient satisfaction in the intervention groups was 0.1 lower (0.65 lower to 0.45 higher)

(a) Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
(b) Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

12.4 Economic evidence

Published literature

No relevant economic evaluations were identified.

One economic evaluation was identified but selectively excluded due to a combination of limited applicability and serious methodological limitations[185]. These are listed in Appendix M, with reasons for exclusion given.

See also the economic article selection flow chart in Appendix F.

12.5 Evidence statements

Clinical

Telemonitoring versus usual care

- Very low quality evidence from 1 RCT comprising of 37 participants demonstrated no clinical difference in mortality between telemonitoring and usual care.
- Very low quality evidence from 1 RCT comprising of 17 participants demonstrated no clinical difference between telemonitoring and usual care with regards to quality of life (SF-36V, physical component).
- Very low quality evidence from 1 RCT comprising of 17 participants demonstrated a clinical benefit of telemonitoring compared to usual care with regards to quality of life (SF-36V, mental component).
- Moderate quality evidence from 1 RCT comprising of 37 participants demonstrated no clinical difference between telemonitoring and usual care with regards to unscheduled care (ED visits and hospitalisations).
- Low quality evidence from 1 RCT comprising of 37 participants demonstrated no clinical difference between telemonitoring and usual care with regards to length of hospital stay.
- Very low quality evidence from 1 RCT comprising of 37 participants demonstrated no clinical difference between telemonitoring and usual care with regards to patient satisfaction.

Telemonitoring with alerts versus usual care

- Very low quality evidence from 1 RCT comprising of 205 participants demonstrated a clinical harm of telemonitoring with alerts compared with usual care with regards to mortality.
- Very low quality evidence from 1 RCT comprising of 180 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to quality of life (SF-12 physical and mental components).
- Low quality evidence from 1 RCT comprising of 166 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to functional outcomes (Barthel ADL index).
- Low quality evidence from 1 RCT comprising of 205 participants demonstrated a clinical harm of telemonitoring with alerts and usual care with regards to ED visits.
- Low quality evidence from 1 RCT comprising of 205 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to mean ED visits.

- Low quality evidence from 1 RCT comprising of 205 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to hospital admissions and mean hospital admissions.
- Low quality evidence from 1 RCT comprising of 205 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to hospital admissions and mean hospital admissions.
- Low quality evidence from 1 RCT comprising of 205 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to length of hospital stay.
- Low quality evidence from 1 RCT comprising of 194 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to hospice visits.
- Low quality evidence from 1 RCT comprising of 194 participants demonstrated no clinical difference between telemonitoring with alerts and usual care with regards to length of hospice stay.
- Very low quality evidence from 1 RCT comprising of 13 participants demonstrated a clinical harm of telemonitoring with alerts and usual care with regards to time to hospice entry.

Telemonitoring with alerts plus case management versus usual care plus case management

- Very low quality evidence from 1 RCT comprising of 104 participants demonstrated no clinical difference between telemonitoring with alerts plus case management and usual care plus case management with regards to functional level.
- Very low quality evidence from 1 RCT comprising of 104 participants demonstrated a clinical harm of telemonitoring with alerts plus case management compared with usual care plus case management with regards to cognitive status.
- Very low quality evidence from 1 RCT comprising of 104 participants demonstrated no clinical difference between telemonitoring with alerts plus case management and usual care plus case management with regards to patient satisfaction.
- Very low quality evidence from 1 RCT comprising of 104 participants demonstrated no clinical difference between telemonitoring with alerts plus case management and usual care plus case management with regards to self-rated health.

Telemonitoring (plus self-management) versus usual care (plus psychoeducation)

- Low quality evidence from 1 RCT comprising of 94 participants demonstrated a clinical benefit between telemonitoring plus self-management and usual care plus psychoeducation with regards to quality of life (SF-36 mental component).
- Moderate quality evidence from 1 RCT comprising of 94 participants demonstrated no clinical difference between telemonitoring plus self-management and usual care plus psychoeducation with regards to ED visits.
- Moderate quality evidence from 1 RCT comprising of 94 participants demonstrated no clinical difference between telemonitoring plus self-management and usual care plus psychoeducation with regards to ‘episodes of care’.
- High quality evidence from 1 RCT comprising of 94 participants demonstrated a clinical benefit between telemonitoring plus self-management and usual care plus psychoeducation with regards to length of hospital stay.
- Low quality evidence from 1 RCT comprising of 94 participants demonstrated no clinical difference between telemonitoring plus self-management and usual care plus psychoeducation with regards to patient satisfaction.

Economic

- No relevant economic evaluations were included.

12.5.1 Recommendations and link to evidence

Recommendations	**No recommendations.**
Research recommendations	**4. What is the clinical and cost effectiveness of alternative approaches to organising primary care compared with usual care for people with multimorbidity?**
Relative values of different outcomes	The GDG considered health-related quality of life, mortality, functional outcomes, patient and carer satisfaction, length of hospital stay and unscheduled care as critical outcomes. The GDG considered continuity of care, admission to care facility and patient or carer treatment burden as important outcomes.
Trade-off between clinical benefits and harms	There was no evidence for the majority of interventions listed in the protocol. Evidence was identified evaluating the clinical effectiveness of telemonitoring for people with multimorbidity only. The clinical evidence for telemonitoring alone compared with usual care showed that there was a clinical benefit in increasing quality of life (mental component)(critical outcome). There was no clinical difference between telemonitoring alone compared with usual care for: mortality, quality of life (physical component), ED visits, hospital admissions, length of hospital stay and patient satisfaction. No clinical harms were identified. The clinical evidence for telemonitoring with alerts versus usual care showed there was a clinical harm with regards to mortality, ED visits (critical outcomes) and to hospice entry. There was no clinical difference between telemonitoring with alerts compared with usual care for: quality of life (physical component), quality of life (mental component), ADLs, hospital admissions, length of hospital stay, hospice admissions or length of hospice stay. No clinical harms were identified. The clinical evidence for telemonitoring with alerts plus case management versus case management alone showed that there was no clinical difference in functional status, cognitive status, patient satisfaction and self-rated health with regards to telemonitoring with alerts versus case management. The clinical evidence for telemonitoring with alerts and self-management components versus usual care and psychoeducation showed that there was a clinical benefit in increasing quality of life (mental component)(critical outcome) and in reducing length of hospital stay (critical outcome). There was no clinical difference for: ED visits, length of hospital stay, episodes of care and patient satisfaction. No clinical harms were identified. Overall, there was limited evidence on what formats of encounters with health professionals would improve health outcomes for people with multimorbidity. The only clinical evidence identified within this area was on the effectiveness of telehealth interventions. The GDG noted that the telemonitoring interventions generally showed little clinical benefit and that there was not sufficient clinical evidence to recommend that telemonitoring with alerts should be used in clinical

	practice for people with multimorbidity.
Economic considerations	No relevant economic evaluations were identified from the published literature on this topic. The GDG discussed that there will be costs associated with telemonitoring (such as the equipment and time of a nurse to review the data). However, given the lack of evidence of clinical benefit for telemonitoring and that no recommendation is being made, the health economics were not considered further by the GDG.
Quality of evidence	The majority of the evidence evaluating telemedicine was of low to very low quality, and the GDG noted that many of the studies had small sample sizes. The majority of studies were at a serious risk of bias, due to selection bias and high missing data rates, and showed serious imprecision. One study has a population of people with multimorbidity. In one study 79% of the people had multimorbidity, and so was downgraded for indirectness. Of the remaining studies, one was an older adult population and one was an adult population with an estimation of 68-92% of people with multimorbidity, but the number of people with multimorbidity was not reported and so they were downgraded for indirectness. All of the evidence came from studies conducted in the USA. The GDG expressed uncertainty about whether evidence from a USA population could be generalised to the UK, as there are significant differences in the structure of the healthcare system. However, the GDG did not strongly believe that this would impact on the effectiveness of telemedicine and therefore agreed not to downgrade the studies for indirectness.
Other considerations	As an alternative to telehealth, the GDG discussed the potential benefit of regular monitoring of people with multimorbidity without a telemonitoring system. The GDG agreed that regular monitoring may be beneficial, for example, when a person self-monitors and notifies the relevant health professional when certain pre-determined thresholds are breached. The GDG suggested that interventions to encourage self-monitoring of symptoms in people with multimorbidity should include clear information for people about when to contact a GP or other health professional. The GDG had agreed a very wide ranging review protocol to include as much information as possible to inform recommendations. They were aware of trials that had started in the UK considering different formats of routine care but were concerned at the paucity of evidence for such an important area. They were aware that GP practices often gave people with more complex care double appointments and that practices used advanced nurse practitioners and people with other skills to augment general practice care. Because of the importance of this area they prioritised it as a research recommendation; further details can be found in Appendix O.

13 Reference list

1 Review Manager (RevMan) [Computer program]. Version 5. Copenhagen. The Nordic Cochrane Centre, The Cochrane Collaboration, 2015. Available from: http://tech.cochrane.org/Revman

2 PROBAST. Personal communication - publication expected for 2016. 2016. Available from: http://www.systematic-reviews.com/probast [Last accessed: 25 February 2016]

3 Abbatecola AM, Spazzafumo L, Corsonello A, Sirolla C, Bustacchini S, Guffanti E. Development and validation of the HOPE prognostic index on 24-month posthospital mortality and rehospitalization: Italian National Research Center on Aging (INRCA). Rejuvenation Research. 2011; 14(6):605-613

4 Ahmad R, Bath PA. Identification of risk factors for 15-year mortality among community-dwelling older people using Cox regression and a genetic algorithm. Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2005; 60(8):1052-1058

5 Alkema GE, Wilber KH, Shannon GR, Allen D. Reduced mortality: the unexpected impact of a telephone-based care management intervention for older adults in managed care. Health Services Research. 2007; 42:1632-1650

6 Allen D, Badro V, Denyer-Willis L, Ellen Macdonald M, Pare A, Hutchinson T et al. Fragmented care and whole-person illness: Decision-making for people with chronic end-stage kidney disease. Chronic Illness. 2015; 11(1):44-55

7 Altfeld SJ, Shier GE, Rooney M, Johnson TJ, Golden RL, Karavolos K et al. Effects of an enhanced discharge planning intervention for hospitalized older adults: a randomized trial. Gerontologist. 2013; 53(3):430-440

8 American Geriatrics Society Expert Panel on the Care of Older Adults with Multimorbidity. Guiding principles for the care of older adults with multimorbidity: an approach for clinicians. Journal of the American Geriatrics Society. 2012; 60(10):E1-E25

9 Applegate WB, Miller ST, Graney MJ, Elam JT, Burns R, Akins DE. A randomized, controlled trial of a geriatric assessment unit in a community rehabilitation hospital. New England Journal of Medicine. 1990; 322(22):1572-1578

10 Arnett DK, Goodman RA, Halperin JL, Anderson JL, Parekh AK, Zoghbi WA. AHA/ACC/HHS strategies to enhance application of clinical practice guidelines in patients with cardiovascular disease and comorbid conditions from the American heart association, American college of cardiology, and US department of health and human services. Circulation. 2014; 130(18):1662-1667

11 Asplund K, Gustafson Y, Jacobsson C, Bucht G, Wahlin A, Peterson J et al. Geriatric-based versus general wards for older acute medical patients: a randomized comparison of outcomes and use of resources. Journal of the American Geriatrics Society. 2000; 48(11):1381-1388

12 Auyeung TW, Lee JS, Leung J, Kwok T, Woo J. The selection of a screening test for frailty identification in community-dwelling older adults. Journal of Nutrition, Health and Aging. 2014; 18(2):199-203

13 Bardach SH, Schoenberg NE. Primary care physicians' prevention counseling with patients with multiple morbidity. Qualitative Health Research. 2012; 22(12):1599-1611

14 Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet. 2012; 380(9836):37-43

15 Battersby MW, Beattie J, Pols RG, Smith DP, Condon J, Blunden S. A randomised controlled trial of the Flinders Program of chronic condition management in Vietnam veterans with co-morbid alcohol misuse, and psychiatric and medical conditions. Australian and New Zealand Journal of Psychiatry. 2013; 47(5):451-462

16 Beck A, Scott J, Williams P, Robertson B, Jackson D, Gade G et al. A randomized trial of group outpatient visits for chronically ill older hmo members: The cooperative health care clinic. Journal of the American Geriatrics Society. 1997; 45(5):543-549

17 Behm L, Wilhelmson K, Falk K, Eklund K, Ziden L, Dahlin-Ivanoff S. Positive health outcomes following health-promoting and disease-preventive interventions for independent very old persons: long-term results of the three-armed RCT Elderly Persons in the Risk Zone. Archives of Gerontology and Geriatrics. 2014; 58(3):376-383

18 Beland SG, Tannenbaum C, Ducruet T, Preville M, Berbiche D, Moride Y. Effect of external variables on the performance of the geriatric comorbidity score derived from prescription claims in the community-dwelling elderly. Drugs and Aging. 2012; 29(11):891-897

19 Berglund H, Hasson H, Kjellgren K, Wilhelmson K. Effects of a continuum of care intervention on frail older persons' life satisfaction: a randomized controlled study. Journal of Clinical Nursing. 2015; 24(7-8):1079-1090

20 Bernabeu-Wittel M, Ollero-Baturone M, Moreno-Gavino L, Baron-Franco B, Fuertes A, Murcia-Zaragoza J et al. Development of a new predictive model for polypathological patients. The PROFUND index. European Journal of Internal Medicine. 2011; 22(3):311-317

21 Black DM, Reid IR, Boonen S, Bucci-Rechtweg C, Cauley JA, Cosman F et al. The effect of 3 versus 6 years of zoledronic acid treatment of osteoporosis: a randomized extension to the HORIZON-Pivotal Fracture Trial (PFT). Journal of Bone and Mineral Research. 2012; 27(2):243-254

22 Black DM, Reid IR, Cauley JA, Cosman F, Leung PC, Lakatos P et al. The effect of 6 versus 9 years of zoledronic acid treatment in osteoporosis: a randomized second extension to the HORIZON-Pivotal Fracture Trial (PFT). Journal of Bone and Mineral Research. 2015; 30(5):934-944

23 Black DM, Schwartz AV, Ensrud KE, Cauley JA, Levis S, Quandt SA et al. Effects of continuing or stopping alendronate after 5 years of treatment: the Fracture Intervention Trial Long-term Extension (FLEX): a randomized trial. JAMA. 2006; 296(24):2927-2938

24 Blakeman T, Blickem C, Kennedy A, Reeves D, Bower P, Gaffney H et al. Effect of information and telephone-guided access to community support for people with chronic kidney disease: randomised controlled trial. PLoS ONE. 2014; 9:e109135

25 Boeckxstaens P, Vaes B, Van Pottelbergh G, De Sutter A, Legrand D, Adriaensen W et al. Multimorbidity measures were poor predictors of adverse events in patients aged >80 years: a prospective cohort study. Journal of Clinical Epidemiology. 2015; 68(2):220-227

26 Boorsma M, Frijters DHM, Knol DL, Ribbe ME, Nijpels G, van Hout HPJ. Effects of multidisciplinary integrated care on quality of care in residential care facilities for elderly

people: a cluster randomized trial. CMAJ Canadian Medical Association Journal. 2011; 183(11):E724-E732

27 Boult C, Dowd B, McCaffrey D, Boult L, Hernandez R, Krulewitch H. Screening elders for risk of hospital admission. Journal of the American Geriatrics Society. 1993; 41(8):811-817

28 Boult C, Pacala JT, Boult LB. Targeting elders for geriatric evaluation and management: reliability, validity, and practicality of a questionnaire. Aging (Milan, Italy). 1995; 7(3):159-164

29 Boult C, Reider L, Frey K, Leff B, Boyd CM, Wolff JL et al. Early effects of "Guided Care" on the quality of health care for multimorbid older persons: a cluster-randomized controlled trial. Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2008; 63(3):321-327

30 Boult C, Reider L, Leff B, Frick KD, Boyd CM, Wolff JL et al. The effect of guided care teams on the use of health services: results from a cluster-randomized controlled trial. Archives of Internal Medicine. 2011; 171:460-466

31 Boult C, Leff B, Boyd CM, Wolff JL, Marsteller JA, Frick KD et al. A matched-pair cluster-randomized trial of guided care for high-risk older patients. Journal of General Internal Medicine. 2013; 28(5):612-621

32 Bouman A, van Rossum E, Evers S, Ambergen T, Kempen G, Knipschild P. Effects on health care use and associated cost of a home visiting program for older people with poor health status: a randomized clinical trial in the Netherlands. Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2008; 63(3):291-297

33 Boxer RS, Wang Z, Walsh SJ, Hager D, Kenny AM. The utility of the 6-minute walk test as a measure of frailty in older adults with heart failure. American Journal of Geriatric Cardiology. 2008; 17(1):7-12

34 Boyd CM, Reider L, Frey K, Scharfstein D, Leff B, Wolff J et al. The effects of guided care on the perceived quality of health care for multi-morbid older persons: 18-month outcomes from a cluster-randomized controlled trial. Journal of General Internal Medicine. 2010; 25:235-242

35 Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. JAMA. 2005; 294(6):716-724

36 Brettschneider C, Luck T, Fleischer S, Roling G, Beutner K, Luppa M et al. Cost-utility analysis of a preventive home visit program for older adults in Germany. BMC Health Services Research. 2015; 15:141

37 Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ : Canadian Medical Association Journal = Journal De L'Association Medicale Canadienne. 2010; 182(18):E839-E842

38 Castell MV, Sanchez M, Julian R, Queipo R, Martin S, Otero A. Frailty prevalence and slow walking speed in persons age 65 and older: implications for primary care. BMC Family Practice. 2013; 14:86

39 Chan TC, Luk JK, Chu LW, Chan FH. Validation study of Charlson Comorbidity Index in predicting mortality in Chinese older adults. Geriatrics and Gerontology International. 2014; 14(2):452-457

40 Chan TC, Shea YF, Luk KH, Chan HW, Chu LW. Development and validation of a prognostic index for 2-year mortality in Chinese older residents living in nursing homes. Geriatrics and Gerontology International. 2012; 12(3):555-562

41 Chow SK, Wong FK. A randomized controlled trial of a nurse-led case management programme for hospital-discharged older adults with co-morbidities. Journal of Advanced Nursing. 2014; 70:2257-2271

42 Christe V, Waeber G, Vollenweider P, Marques-Vidal P. Antihypertensive drug treatment changes in the general population: the CoLaus study. BMC Pharmacology and Toxicology. 2014; 15(1):20

43 Clegg A, Rogers L, Young J. Diagnostic test accuracy of simple instruments for identifying frailty in community-dwelling older people: a systematic review. Age and Ageing. 2015; 44(1):148-152

44 Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and Ageing. 2016; 45(3):353-360

45 Coburn KD, Marcantonio S, Lazansky R, Keller M, Davis N. Effect of a community-based nursing intervention on mortality in chronically ill older adults: a randomized controlled trial. PLoS Medicine. 2012; 9(7):e1001265

46 Cohen HJ, Feussner JR, Weinberger M, Carnes M, Hamdy RC, Hsieh F et al. A controlled trial of inpatient and outpatient geriatric evaluation and management. New England Journal of Medicine. 2002; 346(12):905-912

47 Coleman EA, Wagner EH, Grothaus LC, Hecht J, Savarino J, Buchner DM. Predicting hospitalization and functional decline in older health plan enrollees: are administrative data as accurate as self-report? Journal of the American Geriatrics Society. 1998; 46(4):419-425

48 Collard AF, Bachman SS, Beatrice DF. Acute care delivery for the geriatric patient: an innovative approach. QRB Quality Review Bulletin. 1985; 11(6):180-185

49 Counsell SR, Callahan CM, Clark DO, Tu W, Buttar AB, Stump TE et al. Geriatric care management for low-income seniors: a randomized controlled trial. JAMA. 2007; 298(22):2623-2633

50 Counsell SR, Holder CM, Liebenauer LL, Palmer RM, Fortinsky RH, Kresevic DM et al. Effects of a multicomponent intervention on functional outcomes and process of care in hospitalized older patients: a randomized controlled trial of Acute Care for Elders (ACE) in a community hospital. Journal of the American Geriatrics Society. 2000; 48(12):1572-1581

51 Courtney M, Edwards H, Chang A, Parker A, Finlayson K, Hamilton K. Fewer emergency readmissions and better quality of life for older adults at risk of hospital readmission: A randomized controlled trial to determine the effectiveness of a 24-week exercise and telephone follow-up program. Journal of the American Geriatrics Society. 2009; 57(3):395-402

52 Coventry PA, Fisher L, Kenning C, Bee P, Bower P. Capacity, responsibility, and motivation: a critical qualitative evaluation of patient and practitioner views about barriers to self-management in people with multimorbidity. BMC Health Services Research. 2014; 14:536

53 Covinsky KE, King JT, Jr., Quinn LM, Siddique R, Palmer R, Kresevic DM et al. Do acute care for elders units increase hospital costs? A cost analysis using the hospital perspective. Journal of the American Geriatrics Society. 1997; 45(6):729-734

54 Covinsky KE, Palmer RM, Kresevic DM, Kahana E, Counsell SR, Fortinsky RH et al. Improving functional outcomes in older patients: lessons from an acute care for elders unit. Joint Commission Journal on Quality Improvement. 1998; 24(2):63-76

55 Cowie L, Morgan M, White P, Gulliford M. Experience of continuity of care of patients with multiple long-term conditions in England. Journal of Health Services Research and Policy. 2009; 14(2):82-87

56 Curtis L. Unit costs of health and social care 2010. Canterbury: Personal Social Services Research Unit, University of Kent; 2010. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2010/

57 Curtis L. Unit costs of health and social care 2014. Canterbury: Personal Social Services Research Unit, University of Kent; 2014. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2014/

58 da Camara SM, Alvarado BE, Guralnik JM, Guerra RO, Maciel AC. Using the Short Physical Performance Battery to screen for frailty in young-old adults with distinct socioeconomic conditions. Geriatrics and Gerontology International. 2013; 13(2):421-428

59 Daniels R, van Rossum E, Beurskens A, van den Heuvel W, de Witte L. The predictive validity of three self-report screening instruments for identifying frail older people in the community. BMC Public Health. 2012; 12:69

60 Dent E, Visvanathan R, Piantadosi C, Chapman I. Use of the Mini Nutritional Assessment to detect frailty in hospitalised older people. Journal of Nutrition, Health and Aging. 2012; 16(9):764-767

61 Di Bari M, Profili F, Bandinelli S, Salvioni A, Mossello E, Corridori C et al. Screening for frailty in older adults using a postal questionnaire: rationale, methods, and instruments validation of the INTER-FRAIL study. Journal of the American Geriatrics Society. 2014; 62(10):1933-1937

62 Diez-Manglano J, Cabrerizo Garcia JL, Garcia-Arilla Calvo E, Jimeno Sainz A, Calvo Begueria E, Martinez-Alvarez RM et al. External validation of the PROFUND index in polypathological patients from internal medicine and acute geriatrics departments in Aragon. Internal and Emergency Medicine. 2015; 10(8):915-926

63 Donate-Martinez A, Garces FJ, Rodenas RF. Application of screening tools to detect risk of hospital readmission in elderly patients in Valencian Healthcare System (VHS) (Spain). Archives of Gerontology and Geriatrics. 2014; 59(2):408-414

64 Donnan PT. PEONY data, 2016

65 Donnan PT, Dorward DW, Mutch B, Morris AD. Development and validation of a model for predicting emergency admissions over the next year (PEONY): a UK historical cohort study. Archives of Internal Medicine. 2008; 168(13):1416-1422

66 Druss BG, Zhao L, von Esenwein SA, Bona JR, Fricks L, Jenkins-Tucker S et al. The Health and Recovery Peer (HARP) Program: a peer-led intervention to improve medical self-management for persons with serious mental illness. Schizophrenia Research. 2010; 118:264-270

67 Dunbar SB, Butts B, Reilly CM, Gary RA, Higgins MK, Ferranti EP et al. A pilot test of an integrated self-care intervention for persons with heart failure and concomitant diabetes. Nursing Outlook. 2014; 62:97-111

68 Dunbar SB, Reilly CM, Gary R, Higgins MK, Culler S, Butts B et al. Randomized clinical trial of an integrated self-care intervention for persons with heart failure and diabetes: quality of life and physical functioning outcomes. Journal of Cardiac Failure. 2015; 21(9):719-729

69 Eakin EG, Bull SS, Riley KM, Reeves MM, McLaughlin P, Gutierrez S. Resources for health: a primary-care-based diet and physical activity intervention targeting urban Latinos with multiple chronic conditions. Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association. 2007; 26(4):392-400

70 Edmans J, Bradshaw L, Franklin M, Gladman J, Conroy S. Specialist geriatric medical assessment for patients discharged from hospital acute assessment units: randomised controlled trial. BMJ. 2013; 347:f5874

71 Ekdahl AW, Wirehn AB, Alwin J, Jaarsma T, Unosson M, Husberg M et al. Costs and effects of an ambulatory geriatric unit (the AGe-FIT Study): a randomized controlled trial. Journal of the American Medical Directors Association. 2015; 16(6):497-503

72 Eklund K, Wilhelmson K, Gustafsson H, Landahl S, Dahlin-Ivanoff S. One-year outcome of frailty indicators and activities of daily living following the randomised controlled trial: "Continuum of care for frail older people". BMC Geriatrics. 2013; 13:76

73 Ell K, Katon W, Xie B, Lee PJ, Kapetanovic S, Guterman J et al. Collaborative care management of major depression among low-income, predominantly Hispanic subjects with diabetes: a randomized controlled trial. Diabetes Care. 2010; 33(4):706-713

74 Ellis G, Whitehead MA, O'Neill D, Langhorne P, Robinson D. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database of Systematic Reviews. 2011; Issue 7:CD006211. DOI:10.1002/14651858.CD006211.pub2

75 Ensrud KE, Barrett-Connor EL, Schwartz A, Santora AC, Bauer DC, Suryawanshi S et al. Randomized trial of effect of alendronate continuation versus discontinuation in women with low BMD: results from the Fracture Intervention Trial long-term extension. Journal of Bone and Mineral Research. 2004; 19(8):1259-1269

76 Epstein AM, Hall JA, Fretwell M, Feldstein M, DeCiantis ML, Tognetti J et al. Consultative geriatric assessment for ambulatory patients. A randomized trial in a health maintenance organization. JAMA. 1990; 263(4):538-544

77 Espino DV, Bazaldua OV, Palmer RF, Mouton CP, Parchman ML, Miles TP et al. Suboptimal medication use and mortality in an older adult community-based cohort: results from the Hispanic EPESE Study. Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2006; 61(2):170-175

78 Farmer C, K H. Q-BAST: a new checklist for evaluating risk of bias in questionnaires with no gold standard. G-I-N Conference 2015, Amsterdam, The Netherlands 2015;

79 Fortin M, Hudon C, Dubois MF, Almirall J, Lapointe L, Soubhi H. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. Health and Quality of Life Outcomes. 2005; 3:74

80 Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health and Quality of Life Outcomes. 2004; 2:51

81 Freis ED. Return of elevated blood pressure after withdrawal of antihypertensive drugs. Circulation. 1975; 51(6):1107-1113

82 Frese T, Deutsch T, Keyser M, Sandholzer H. In-home preventive comprehensive geriatric assessment (CGA) reduces mortality--a randomized controlled trial. Archives of Gerontology and Geriatrics. 2012; 55(3):639-644

83 Fretwell MD, Raymond PM, McGarvey ST, Owens N, Traines M, Silliman RA et al. The Senior Care Study. A controlled trial of a consultative/unit-based geriatric assessment program in acute care. Journal of the American Geriatrics Society. 1990; 38(10):1073-1081

84 Fried TR, McGraw S, Agostini JV, Tinetti ME. Views of older persons with multiple morbidities on competing outcomes and clinical decision-making. Journal of the American Geriatrics Society. 2008; 56(10):1839-1844

85 Friedman B, Li Y, Liebel DV, Powers BA. Effects of a home visiting nurse intervention versus care as usual on individual activities of daily living: a secondary analysis of a randomized controlled trial. BMC Geriatrics. 2014; 14:24

86 Gage H, Ting S, Williams P, Drennan V, Goodman C, Iliffe S et al. Nurse-led case management for community dwelling older people: an explorative study of models and costs. Journal of Nursing Management. United Kingdom 2013; 21(1):191-201

87 Garvey J, Connolly D, Boland F, Smith SM. OPTIMAL, an occupational therapy led self-management support programme for people with multimorbidity in primary care: a randomized controlled trial. BMC Family Practice. 2015; 16:59

88 Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. Journal of the American Geriatrics Society. 2014; 62(5):889-895

89 Gibbons CJ, Kenning C, Coventry PA, Bee P, Bundy C, Fisher L et al. Development of a multimorbidity illness perceptions scale (MULTIPleS). PLoS ONE. 2013; 8(12):e81852

90 Gill A, Kuluski K, Jaakkimainen L, Naganathan G, Upshur R, Wodchis WP. "Where do we go from here?" health system frustrations expressed by patients with multimorbidity, their caregivers and family physicians. Healthcare Policy. 2014; 9(4):73-89

91 Gitlin LN, Hauck WW, Dennis MP, Winter L, Hodgson N, Schinfeld S. Long-term effect on mortality of a home intervention that reduces functional difficulties in older adults: results from a randomized trial. Journal of the American Geriatrics Society. 2009; 57(3):476-481

92 Gitlin LN, Hauck WW, Winter L, Dennis MP, Schulz R. Effect of an in-home occupational and physical therapy intervention on reducing mortality in functionally vulnerable older people: preliminary findings. Journal of the American Geriatrics Society. 2006; 54(6):950-955

93 Gitlin LN, Winter L, Dennis MP, Corcoran M, Schinfeld S, Hauck WW. A randomized trial of a multicomponent home intervention to reduce functional difficulties in older adults. Journal of the American Geriatrics Society. 2006; 54(5):809-816

94 Gnjidic D, Hilmer SN, Blyth FM, Naganathan V, Waite L, Seibel MJ et al. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. Journal of Clinical Epidemiology. 2012; 65(9):989-995

95 Goldberg RW, Dickerson F, Lucksted A, Brown CH, Weber E, Tenhula WN et al. Living well: an intervention to improve self-management of medical illness for individuals with serious mental illness. Psychiatric Services. 2013; 64:51-57

96 Gomez C, Vega-Quiroga S, Bermejo-Pareja F, Medrano MJ, Louis ED, Benito-Leon J. Polypharmacy in the elderly: a marker of increased risk of mortality in a population-based prospective study (NEDICES). Gerontology. 2015; 61(4):301-309

97 Goodman RA, Boyd C, Tinetti ME, Von Kohorn I, Parekh AK, McGinnis JM. IOM and DHHS meeting on making clinical practice guidelines appropriate for patients with multiple chronic conditions. Annals of Family Medicine. 2014; 12(3):256-259

98 GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: http://www.gradeworkinggroup.org/ [Last accessed: 27 January 2016]

99 Greenberg G. Course of blood pressure in mild hypertensives after withdrawal of long term antihypertensive treatment. BMJ. 1986; 293(6553):988-992

100 Grimmer K, Milanese S, Beaton K, Atlas A. Additional measures do not improve the diagnostic accuracy of the Hospital Admission Risk Profile for detecting downstream quality of life in community-dwelling older people presenting to a hospital emergency department. Clinical Interventions In Aging. 2014; 9:233-242

101 Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ. 2012; 345:e6341

102 Harris RD, Henschke PJ, Popplewell PY, Radford AJ, Bond MJ, Turnbull RJ et al. A randomised study of outcomes in a defined group of acutely ill elderly patients managed in a geriatric assessment unit or a general medical unit. Australian and New Zealand Journal of Medicine. 1991; 21(2):230-234

103 Harvey P, Storer M, Berlowitz DJ, Jackson B, Hutchinson A, Lim WK. Feasibility and impact of a post-discharge geriatric evaluation and management service for patients from residential care: the Residential Care Intervention Program in the Elderly (RECIPE). BMC Geriatrics. 2014; 14:48

104 Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Annals of Internal Medicine. 2006; 144(6):427-437

105 Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Annals of Internal Medicine. 2013; 158(4):280-286

106 Heyworth ITM, Hazell ML, Linehan MF, Frank TL. How do common chronic conditions affect health-related quality of life? British Journal of General Practice. 2009; 59(568):e353-e358

107 Hippisley-Cox J. QAdmissions thresholds for practices, 2016

108 Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: Derivation and validation of QAdmissions score. BMJ Open. 2013; 3(8)

109 Hochhalter AK, Song J, Rush J, Sklar L, Stevens A. Making the most of your healthcare intervention for older adults with multiple chronic illnesses. Patient Education and Counseling. 2010; 81(2):207-213

110 Hogan DB, Fox RA, Badley BW, Mann OE. Effect of a geriatric consultation service on management of patients in an acute care hospital. CMAJ Canadian Medical Association Journal. 1987; 136(7):713-717

111 Hogg W, Lemelin J, Dahrouge S, Liddy C, Armstrong CD, Legault F et al. Randomized controlled trial of anticipatory and preventive multidisciplinary team care: for complex patients in a community-based primary care setting. Canadian Family Physician. 2009; 55(12):e76-e85

112 Hoogendijk EO, van der Horst HE, Deeg DJ, Frijters DH, Prins BA, Jansen AP et al. The identification of frail older adults in primary care: comparing the accuracy of five simple instruments. Age and Ageing. 2013; 42(2):262-265

113 Hopp F, Woodbridge P, Subramanian U, Copeland L, Smith D, Lowery J. Outcomes associated with a home care telehealth intervention. Telemedicine Journal and E-Health. 2006; 12:297-307

114 Hughes LD, McMurdo MET, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age and Ageing. 2013; 42(1):62-69

115 ISD Scotland. GP consultations: practice team information (PTI) statistics, 2013. Available from: http://www.isdscotland.org/Health-Topics/General-Practice/GP-consultations/

116 Jensen GL, Friedmann JM, Coleman CD, Smiciklas-Wright H. Screening for hospitalization and nutritional risks among community-dwelling older persons. American Journal of Clinical Nutrition. 2001; 74(2):201-205

117 Jones D, Song X, Mitnitski A, Rockwood K. Evaluation of a frailty index based on a comprehensive geriatric assessment in a population based study of elderly Canadians. Aging Clinical and Experimental Research. 2005; 17(6):465-471

118 Jowsey T, Jeon YH, Dugdale P, Glasgow NJ, Kljakovic M, Usherwood T. Challenges for co-morbid chronic illness care and policy in Australia: a qualitative study. Australia and New Zealand Health Policy. 2009; 6:22

119 Jyrkka J, Enlund H, Korhonen MJ, Sulkava R, Hartikainen S. Polypharmacy status as an indicator of mortality in an elderly population. Drugs and Aging. 2009; 26(12):1039-1048

120 Karppi P. Effects of a geriatric inpatient unit on elderly home care patients: a controlled trial. Aging (Milan, Italy). 1995; 7(3):207-211

121 Karppi P, Tilvis R. Effectiveness of a Finnish geriatric inpatient assessment. Two-year follow up of a randomized clinical trial on community-dwelling patients. Scandinavian Journal of Primary Health Care. 1995; 13(2):93-98

122 Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B et al. Collaborative care for patients with depression and chronic illnesses. New England Journal of Medicine. 2010; 363(27):2611-2620

123 Katon W, Russo J, Lin EHB, Schmittdiel J, Ciechanowski P, Ludman E et al. Cost-effectiveness of a multicondition collaborative care intervention: a randomized controlled trial. Archives of General Psychiatry. 2012; 69(5):506-514

124 Kay G, MacTavish M, Moffatt C, Lau G. Development and evaluation of a geriatric assessment unit in a community hospital. Perspectives (Gerontological Nursing Association (Canada)). 1992; 16(3):2-9

125 Kircher TT, Wormstall H, Muller PH, Schwarzler F, Buchkremer G, Wild K et al. A randomised trial of a geriatric evaluation and management consultation services in frail hospitalised patients. Age and Ageing. 2007; 36(1):36-42

126 Koch G, Wakefield BJ, Wakefield DS. Barriers and facilitators to managing multiple chronic conditions: A systematic literature review. Western Journal of Nursing Research. 2015; 37(4):498-516

127 Krause KJ. Mortality and prescription drug use: an analysis using the CHS limited access data. Journal of Insurance Medicine. 2007; 39(1):24-27

128 Kutner JS, Blatchford PJ, Taylor DH, Jr., Ritchie CS, Bull JH, Fairclough DL et al. Safety and benefit of discontinuing statin therapy in the setting of advanced, life-limiting illness: a randomized clinical trial. JAMA Internal Medicine. 2015; 175(5):691-700

129 Lampela P, Hartikainen S, Lavikainen P, Sulkava R, Huupponen R. Effects of medication assessment as part of a comprehensive geriatric assessment on drug use over a 1-year period: a population-based intervention study. Drugs and Aging. 2010; 27(6):507-521

130 Lampela P, Lavikainen P, Huupponen R, Leskinen E, Hartikainen S. Comprehensive geriatric assessment decreases prevalence of orthostatic hypotension in older persons. Scandinavian Journal of Public Health. 2013; 41(4):351-358

131 Landefeld CS, Palmer RM, Kresevic DM, Fortinsky RH, Kowal J. A randomized trial of care in a hospital medical unit especially designed to improve the functional outcomes of acutely ill older patients. New England Journal of Medicine. 1995; 332(20):1338-1344

132 Legrain S, Tubach F, Bonnet-Zamponi D, Lemaire A, Aquino JP, Paillaud E et al. A new multimodal geriatric discharge-planning intervention to prevent emergency visits and rehospitalizations of older adults: the optimization of medication in AGEd multicenter randomized controlled trial. Journal of the American Geriatrics Society. 2011; 59(11):2017-2028

133 Lewin G, Allan J, Patterson C, Knuiman M, Boldy D, Hendrie D. A comparison of the home-care and healthcare service use and costs of older Australians randomised to receive a restorative or a conventional home-care service. Health and Social Care in the Community. 2014; 22(3):328-336

134 Li CM, Chen CY, Li CY, Wang WD, Wu SC. The effectiveness of a comprehensive geriatric assessment intervention program for frailty in community-dwelling older people: a randomized, controlled trial. Archives of Gerontology and Geriatrics. 2010; 50 Suppl 1:S39-S42

135 Lihavainen K, Sipila S, Rantanen T, Seppanen J, Lavikainen P, Sulkava R et al. Effects of comprehensive geriatric intervention on physical performance among people aged 75 years and over. Aging Clinical and Experimental Research. 2012; 24(4):331-338

136 Lihavainen K, Sipila S, Rantanen T, Kauppinen M, Sulkava R, Hartikainen S. Effects of comprehensive geriatric assessment and targeted intervention on mobility in persons aged 75 years and over: a randomized controlled trial. Clinical Rehabilitation. 2012; 26(4):314-326

137 Lorig KR, Sobel DS, Stewart AL, Brown BW, Jr., Bandura A, Ritter P et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Medical Care. 1999; 37(1):5-14

138 Ludman EJ, Peterson D, Katon WJ, Lin EH, Von Korff M, Ciechanowski P et al. Improving confidence for self care in patients with depression and chronic illnesses. Behavioral Medicine. 2013; 39:1-6

139 MacNeil Vroomen JL, Boorsma M, Bosmans JE, Frijters DH, Nijpels G, van Hout HP. Is it time for a change? A cost-effectiveness analysis comparing a Multidisciplinary Integrated Care model for residential homes to usual care. PLoS ONE. Netherlands 2012; 7(5):e37444

140 Maland LJ, Lutz LJ, Castle CH. Effects of withdrawing diuretic therapy on blood pressure in mild hypertension. Hypertension. 1983; 5(4):539-544

141 Marek KD, Stetzer F, Ryan PA, Bub LD, Adams SJ, Schlidt A et al. Nurse care coordination and technology effects on health status of frail older adults via enhanced self-management of medication: randomized clinical trial to test efficacy. Nursing Research. 2013; 62(4):269-278

142 Martinez-Velilla N, Cambra-Contin K, Ibanez-Beroiz B. Comorbidity and prognostic indices do not improve the 5-year mortality prediction of components of comprehensive geriatric assessment in hospitalized older patients. BMC Geriatrics. 2014; 14:64

143 May C, Montori VM, Mair FS. We need minimally disruptive medicine. BMJ. 2009; 339:b2803

144 Mazzaglia G, Roti L, Corsini G, Colombini A, Maciocco G, Marchionni N et al. Screening of older community-dwelling people at risk for death and hospitalization: the Assistenza Socio-Sanitaria in Italia project. Journal of the American Geriatrics Society. 2007; 55(12):1955-1960

145 McGregor M, Lin EHB, Katon WJ. TEAMcare: an integrated multicondition collaborative care program for chronic illnesses and depression. Journal of Ambulatory Care Management. 2011; 34(2):152-162

146 Md Yusof MY, Horan MA, Jones M, McInnes L, Rabbitt PMA, Pendleton N. Developing a self-reported comorbidity index to predict mortality of community-dwelling older adults. Archives of Gerontology and Geriatrics. 2010; 50(3):e63-e67

147 Melis RJ, van Eijken MI, Teerenstra S, Van Achterberg T, Parker SG, Borm GF et al. A randomized study of a multidisciplinary program to intervene on geriatric syndromes in vulnerable older people who live at home (Dutch EASYcare Study). Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2008; 63(3):283-290

148 Metzelthin SF, van Rossum E, de Witte LP, Ambergen AW, Hobma SO, Sipers W et al. Effectiveness of interdisciplinary primary care approach to reduce disability in community dwelling frail older people: cluster randomised controlled trial. BMJ. 2013; 347:f5264

149 Michalska D, Stepan JJ, Basson BR, Pavo I. The effect of raloxifene after discontinuation of long-term alendronate treatment of postmenopausal osteoporosis. Journal of Clinical Endocrinology and Metabolism. 2006; 91(3):870-877

150 Miller PD, Watts NB, Licata AA, Harris ST, Genant HK, Wasnich RD et al. Cyclical etidronate in the treatment of postmenopausal osteoporosis: efficacy and safety after seven years of treatment. American Journal of Medicine. 1997; 103(6):468-476

151 Miller ST, Applegate WB, Elam JT, Graney MJ. Influence of diagnostic classification on outcomes and charges in geriatric assessment and rehabilitation. Journal of the American Geriatrics Society. 1994; 42(1):11-15

152 Min L, Yoon W, Mariano J, Wenger NS, Elliott MN, Kamberg C et al. The vulnerable elders-13 survey predicts 5-year functional decline and mortality outcomes in older ambulatory care patients. Journal of the American Geriatrics Society. 2009; 57(11):2070-2076

153 Monteserin R, Brotons C, Moral I, Altimir S, San Jose A, Santaeugenia S et al. Effectiveness of a geriatric intervention in primary care: a randomized clinical trial. Family Practice. 2010; 27(3):239-245

154 Mosley DG, Peterson E, Martin DC. Do hierarchical condition category model scores predict hospitalization risk in newly enrolled Medicare advantage participants as well as probability of repeated admission scores? Journal of the American Geriatrics Society. 2009; 57(12):2306-2310

155 Muth C, van den Akker M, Blom JW, Mallen CD, Rochon J, Schellevis FG et al. The Ariadne principles: How to handle multimorbidity in primary care consultations. BMC Medicine. 2014; 12(1)

156 National Clinical Guideline Centre. Chronic heart failure: the management of chronic heart failure in adults in primary and secondary care. NICE clinical guideline 108. London. National Clinical Guideline Centre, 2010. Available from: http://guidance.nice.org.uk/CG108/

157 National Clinical Guideline Centre. Hypertension: the clinical managment of primary hypertension in adults: update of clinical guidelines 18 and 34. NICE clinical guideline 127. London. National Clinical Guideline Centre, 2011. Available from: http://guidance.nice.org.uk/CG127

158 National Clinical Guideline Centre. The management of stable angina. NICE clinical guideline 126. London. National Institute for Health and Clinical Excellence, 2011. Available from: http://guidance.nice.org.uk/CG126

159 National Clinical Guideline Centre. Headaches: diagnosis and management of headaches in young people and adults. NICE clinical guideline 150. London. National Clinical Guideline Centre, 2012. Available from: http://guidance.nice.org.uk/CG150

160 National Clinical Guideline Centre. Patient experience in adult NHS services: improving the experience of care for people using adult NHS services. NICE clinical guideline 138. London. National Clinical Guideline Centre, 2012. Available from: http://www.nice.org.uk/CG138

161 National Clinical Guideline Centre. Atrial fibrillation: the management of atrial fibrillation. NICE clinical guideline 180. London. National Clinical Guideline Centre, 2014. Available from: http://guidance.nice.org.uk/CG180

162 National Clinical Guideline Centre. Chronic kidney disease: early identification and management of chronic kidney disease in adults in primary and secondary care. NICE clinical guideline 182. London. National Clinical Guideline Centre, 2014. Available from: http://guidance.nice.org.uk/CG182

163 National Clinical Guideline Centre. Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. NICE clinical guideline 181. London. National Clinical Guideline Centre, 2014. Available from: http://guidance.nice.org.uk/CG181

164 National Clinical Guideline Centre. Type 2 diabetes in adults: management. NICE guideline 28. London. National Clinical Guideline Centre, 2015. Available from: https://www.nice.org.uk/guidance/ng28

165 National Collaborating Centre for Chronic Conditions. Hypertension: management in adults in primary care: pharmacological update. NICE clinical guideline 34. London. Royal College of Physicians, 2006. Available from: http://guidance.nice.org.uk/CG34

166 National Collaborating Centre for Mental Health. Depression: the treatment and management of depression in adults. NICE clinical guideline 90. London. Royal College of Psychiatrists and The British Psychological Society, 2009. Available from: http://www.nice.org.uk/CG90

167 National Collaborating Centre for Mental Health. The treatment and management of depression in adults with chronic physical health problems (partial update of CG23). NICE clinical guideline 91. Leicester and London. Royal College of Psychiatrists and The British Psychological Society, 2009. Available from: http://guidance.nice.org.uk/CG91

168 National Collaborating Centre for Mental Health. Psychosis and schizophrenia in adults. NICE clinical guideline 178. London. Royal College of Psychiatrists and The British Psychological Society, 2014. Available from: http://guidance.nice.org.uk/CG178

169 National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

170 National Institute for Health and Care Excellence. Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes. NICE guideline 5. Manchester. National Institute for Health and Care Excellence, 2015. Available from: https://www.nice.org.uk/guidance/ng5

171 National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. 2nd edition. London: National Institute for Health and Clinical Excellence; 2008. Available from: http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf

172 National Institute for Health and Clinical Excellence. Medicines adherence: involving patients in decisions about prescribed medicines and supporting adherence. NICE clinical guideline 76. London. Royal College of General Practitioners, 2009. Available from: http://guidance.nice.org.uk/CG76

173 National Institute for Health and Clinical Excellence. The guidelines manual. London: National Institute for Health and Clinical Excellence; 2012. Available from: http://publications.nice.org.uk/the-guidelines-manual-pmg6/

174 National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal 2013. 2nd edition. London: National Institute for Health and Clinical Excellence; 2013. Available from: http://publications.nice.org.uk/pmg9

175 Naughton BJ, Moran MB, Feinglass J, Falconer J, Williams ME. Reducing hospital costs for the geriatric patient admitted from the emergency department: a randomized trial. Journal of the American Geriatrics Society. 1994; 42(10):1045-1049

176 Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. Journal of the American Geriatrics Society. 2004; 52(5):675-684

177 Ng JH, Elliott MN, Scholle SH, Ahmed K, Collins RL, Bierman AS. Identifying older adults at high risk of mortality using the Medicare health outcomes survey. Journal of Ambulatory Care Management. 2012; 35(4):277-291

178 NHS Scotland. Polypharmacy guidance, 2012. Available from: http://www.central.knowledge.scot.nhs.uk/upload/Polypharmacy%20full%20guidance%20v2.pdf

179 Nikolaus T, Specht-Leible N, Bach M, Oster P, Schlierf G. A randomized trial of comprehensive geriatric assessment and home intervention in the care of hospitalized patients. Age and Ageing. Germany 1999; 28(6):543-550

180 Noel HC, Vogel DC, Erdos JJ, Cornwall D, Levin F. Home telehealth reduces healthcare costs. Telemedicine Journal and E-Health. 2004; 10(2):170-183

181 North of England Hypertension Guideline Development Group. Hypertension (persistently high blood pressure) in adults. NICE clinical guideline 18, 2004. Available from: http://guidance.nice.org.uk/CG18

182 Nunes DP, Duarte YAdO, Santos JLF, Lebrao ML. Screening for frailty in older adults using a self-reported instrument. Revista De Saude Publica. 2015; 49:2

183 Office for National Statistics. Life tables. 2013. Available from: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-365199 [Last accessed: 11 February 2016]

184 Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2008. Available from: http://www.oecd.org/std/ppp [Last accessed: 17 February 2016]

185 Pare G, Poba-Nzaou P, Sicotte C. Home telemonitoring for chronic disease management: an economic assessment. International Journal of Technology Assessment in Health Care. 2013; 29(2):155-161

186 Park YH, Chang H. Effect of a health coaching self-management program for older adults with multimorbidity in nursing homes. Patient Preference and Adherence. 2014; 8:959-970

187 Pecina JL, Hanson GJ, Van Houten H, Takahashi PY. Impact of telemonitoring on older adults health-related quality of life: the Tele-ERA study. Quality of Life Research. 2013; 22:2315-2321

188 Phibbs CS, Holty JE, Goldstein MK, Garber AM, Wang Y, Feussner JR et al. The effect of geriatrics evaluation and management on nursing home use and health care costs: results from a randomized trial. Medical Care. 2006; 44(1):91-95

189 Pilotto A, Ferrucci L, Franceschi M, D'Ambrosio LP, Scarcelli C, Cascavilla L et al. Development and validation of a multidimensional prognostic index for one-year mortality from

comprehensive geriatric assessment in hospitalized older patients. Rejuvenation Research. 2008; 11(1):151-161

190 Pozzi C, Lapi F, Mazzaglia G, Inzitari M, Boncinelli M, Geppetti P et al. Is suboptimal prescribing a risk factor for poor health outcomes in community-dwelling elders? The ICARe Dicomano study. Pharmacoepidemiology and Drug Safety. 2010; 19(9):954-960

191 Purser JL, Kuchibhatla MN, Fillenbaum GG, Harding T, Peterson ED, Alexander KP. Identifying frailty in hospitalized older adults with significant coronary artery disease. Journal of the American Geriatrics Society. 2006; 54(11):1674-1681

192 Radley DC, Gottlieb DJ, Fisher ES, Tosteson AN. Comorbidity risk-adjustment strategies are comparable among persons with hip fracture. Journal of Clinical Epidemiology. 2008; 61(6):580-587

193 Reuben DB, Borok GM, Wolde-Tsadik G, Ershoff DH, Fishman LK, Ambrosini VL et al. A randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. New England Journal of Medicine. 1995; 332(20):1345-1350

194 Richardson K, Ananou A, Lafortune L, Brayne C, Matthews FE. Variation over time in the association between polypharmacy and mortality in the older population. Drugs and Aging. 2011; 28(7):547-560

195 Ritt M, Schwarz C, Kronawitter V, Delinic A, Bollheimer LC, Gassmann KG et al. Analysis of Rockwood et al's clinical frailty scale and Fried et al's frailty phenotype as predictors of mortality and other clinical outcomes in older patients who were admitted to a geriatric ward. Journal of Nutrition, Health and Aging. 2015; 19(10):1043-1048

196 Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I et al. A global clinical measure of fitness and frailty in elderly people. CMAJ Canadian Medical Association Journal. 2005; 173(5):489-495

197 Rubenstein LZ, Josephson KR, Harker JO, Miller DK, Wieland D. The Sepulveda GEU Study revisited: long-term outcomes, use of services, and costs. Aging (Milan, Italy). 1995; 7(3):212-217

198 Rubenstein LZ, Josephson KR, Wieland GD, English PA, Sayre JA, Kane RL. Effectiveness of a geriatric evaluation unit. A randomized clinical trial. New England Journal of Medicine. 1984; 311(26):1664-1670

199 Rubenstein LZ, Wieland GD, Josephson KR, Rosbrook B, Sayre J, Kane RL. Improved survival for frail elderly inpatients on a geriatric evaluation unit (GEU): who benefits? Journal of Clinical Epidemiology. 1988; 41(5):441-449

200 Rubin CD, Sizemore MT, Loftis PA, Adams-Huet B, Anderson RJ. The effect of geriatric evaluation and management on Medicare reimbursement in a large public hospital: a randomized clinical trial. Journal of the American Geriatrics Society. 1992; 40(10):989-995

201 Rubin CD, Sizemore MT, Loftis PA, de Mola NL. A randomized, controlled trial of outpatient geriatric evaluation and management in a large public hospital. Journal of the American Geriatrics Society. 1993; 41(10):1023-1028

202 Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. British Journal of General Practice. 2011; 61(582):e12-e21

203 Saltvedt I, Jordhoy M, Opdahl Mo ES, Fayers P, Kaasa S, Sletvold O. Randomised trial of in-hospital geriatric intervention: impact on function and morale. Gerontology. 2006; 52(4):223-230

204 Saltvedt I, Saltnes T, Mo ES, Fayers P, Kaasa S, Sletvold O. Acute geriatric intervention increases the number of patients able to live at home. A prospective randomized study. Aging Clinical and Experimental Research. 2004; 16(4):300-306

205 Saltvedt I, Spigset O, Ruths S, Fayers P, Kaasa S, Sletvold O. Patterns of drug prescription in a geriatric evaluation and management unit as compared with the general medical wards: a randomised study. European Journal of Clinical Pharmacology. 2005; 61(12):921-928

206 Saltvedt I, Mo ESO, Fayers P, Kaasa S, Sletvold O. Reduced mortality in treating acutely sick, frail older patients in a geriatric evaluation and management unit. A prospective randomized trial. Journal of the American Geriatrics Society.: Blackwell Science Inc. 2002; 50(5):792-798

207 Sancarlo D, D'Onofrio G, Franceschi M, Scarcelli C, Niro V, Addante F et al. Validation of a Modified-Multidimensional Prognostic Index (m-MPI) including the Mini Nutritional Assessment Short-Form (MNA-SF) for the prediction of one-year mortality in hospitalized elderly patients. Journal of Nutrition, Health and Aging. 2011; 15(3):169-173

208 Sancarlo D, Pilotto A, Panza F, Copetti M, Longo MG, D'Ambrosio P et al. A Multidimensional Prognostic Index (MPI) based on a comprehensive geriatric assessment predicts short- and long-term all-cause mortality in older hospitalized patients with transient ischemic attack. Journal of Neurology. 2012; 259(4):670-678

209 Sandberg M, Kristensson J, Midlov P, Jakobsson U. Effects on healthcare utilization of case management for frail older people: a randomized controlled trial (RCT). Archives of Gerontology and Geriatrics. 2015; 60(1):71-81

210 Savva GM, Donoghue OA, Horgan F, O'Regan C, Cronin H, Kenny RA. Using timed up-and-go to identify frail members of the older population. Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2013; 68(4):441-446

211 Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. American Journal of Epidemiology. 2001; 154(9):854-864

212 Schoenberg NE, Bardach SH, Manchikanti KN, Goodenow AC. Appalachian residents' experiences with and management of multiple morbidity. Qualitative Health Research. 2011; 21(5):601-611

213 Schoon Y, Bongers K, Van Kempen J, Melis R, Olde Rikkert M. Gait speed as a test for monitoring frailty in community-dwelling older people has the highest diagnostic value compared to step length and chair rise time. European Journal of Physical and Rehabilitation Medicine. 2014; 50(6):693-701

214 Senior HEJ, Parsons M, Kerse N, Chen MH, Jacobs S, Hoorn SV et al. Promoting independence in frail older people: a randomised controlled trial of a restorative care service in New Zealand. Age and Ageing. 2014; 43(3):418-424

215 Shamian J, Clarfield AM, Maclean J. A randomized trial of intra-hospital relocation of geriatric patients in a tertiary-care teaching hospital. Journal of the American Geriatrics Society. 1984; 32(11):794-800

216 Silliman RA, McGarvey ST, Raymond PM, Fretwell MD. The Senior Care Study. Does inpatient interdisciplinary geriatric assessment help the family caregivers of acutely ill older patients? Journal of the American Geriatrics Society. 1990; 38(4):461-466

217 Sinnott C, Hugh S, Browne J, Bradley C. GPs' perspectives on the management of patients with multimorbidity: systematic review and synthesis of qualitative research. BMJ Open. 2013; 3:e003610

218 Slaets JP, Kauffmann RH, Duivenvoorden HJ, Pelemans W, Schudel WJ. A randomized trial of geriatric liaison intervention in elderly medical inpatients. Psychosomatic Medicine. 1997; 59:585-591

219 Smets IH, Kempen GI, Janssen-Heijnen ML, Deckx L, Buntinx FJ, van den Akker M. Four screening instruments for frailty in older patients with and without cancer: a diagnostic study. BMC Geriatrics. 2014; 14:26

220 Sommers LS, Marton KI, Barbaccia JC, Randolph J. Physician, nurse, and social worker collaboration in primary care for chronically ill seniors. Archives of Internal Medicine. 2000; 160(12):1825-1833

221 Soong J, Poots AJ, Scott S, Donald K, Bell D. Developing and validating a risk prediction model for acute care based on frailty syndromes. BMJ Open. 2015; 5(10):e008457

222 Spector WD, Limcangco R, Williams C, Rhodes W, Hurd D. Potentially avoidable hospitalizations for elderly long-stay residents in nursing homes. Medical Care. 2013; 51(8):673-681

223 Susser SR, McCusker J, Belzile E. Comorbidity information in older patients at an emergency visit: self-report vs. administrative data had poor agreement but similar predictive validity. Journal of Clinical Epidemiology. 2008; 61(5):511-515

224 Takahashi PY, Hanson GJ, Thorsteinsdottir B, Van Houten HK, Shah ND, Naessens JM et al. The impact of telemonitoring upon hospice referral in the community: a randomized controlled trial. Clinical Interventions In Aging. 2012; 7:445-451

225 Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Archives of Internal Medicine. 2012; 172:773-779

226 Tanajewski L, Franklin M, Gkountouras G, Berdunov V, Edmans J, Conroy S et al. Cost-effectiveness of a specialist geriatric medical intervention for frail older people discharged from acute medical units: economic evaluation in a two-centre randomised controlled trial (AMIGOS). PLoS ONE. 2015; 10(5):e0121340

227 Thomas DR, Brahan R, Haywood BP. Inpatient community-based geriatric assessment reduces subsequent mortality. Journal of the American Geriatrics Society. 1993; 41(2):101-104

228 Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007; 8:16

229 Townsend A, Hunt K, Wyke S. Managing multiple morbidity in mid-life: a qualitative study of attitudes to drug use. BMJ. 2003; 327(7419):837

230 Townsend A, Wyke S, Hunt K. Frequent consulting and multiple morbidity: a qualitative comparison of 'high' and 'low' consulters of GPs. Family Practice. 2008; 25(3):168-175

231 Tran VT, Harrington M, Montori VM, Barnes C, Wicks P, Ravaud P. Adaptation and validation of the Treatment Burden Questionnaire (TBQ) in English using an internet platform. BMC Medicine. 2014; 12:109

232 Tran VT, Montori VM, Eton DT, Baruch D, Falissard B, Ravaud P. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. BMC Medicine. 2012; 10:68

233 Trentini M, Semeraro S, Motta M. Effectiveness of geriatric evaluation and care. One-year results of a multicenter randomized clinical trial. Aging (Milan, Italy). 2001; 13(5):395-405

234 Tribess S, Virtuoso Junior JS, da Silva CAR, de Paulo TRS, Romo-Perez V, de Oliveira RJ. The age marker as a predictive indicator for frailty syndrome in the elderly. Acta Scientiarum - Health Sciences. 2013; 35(2):273-277

235 Tribess S, Virtuoso Junior JS, Oliveira RJ. Physical activity as a predictor of absence of frailty in the elderly. Revista Da Associacao Medica Brasileira. 2012; 58(3):341-347

236 Upatising B, Hanson GJ, Kim YL, Cha SS, Yih Y, Takahashi PY. Effects of home telemonitoring on transitions between frailty states and death for older adults: a randomized controlled trial. International Journal of General Medicine. 2013; 6:145-151

237 Von Korff M, Vitiello MV, McCurry SM, Balderson BH, Moore AL, Baker LD et al. Group interventions for co-morbid insomnia and osteoarthritis pain in primary care: the lifestyles cluster randomized trial design. Contemporary Clinical Trials. 2012; 33:759-768

238 Wagner JT, Bachmann LM, Boult C, Harari D, von Renteln-Kruse W, Egger M et al. Predicting the risk of hospital admission in older persons--validation of a brief self-administered questionnaire in three European countries. Journal of the American Geriatrics Society. 2006; 54(8):1271-1276

239 Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. Medical Care. 2014; 52(8):751-765

240 Wallis SJ, Wall J, Biram RWS, Romero-Ortuno R. Association of the clinical frailty scale with hospital outcomes. QJM. 2015; 108(12):943-949

241 Wang R, Chen L, Fan L, Gao D, Liang Z, He J et al. Incidence and effects of polypharmacy on clinical outcome among patients aged 80+: a five-year follow-up study. PLoS ONE. 2015; 10(11):e0142123

242 White SJ, Powers JS, Knight JR, Harrell D, Varnell L, Vaughn C et al. Effectiveness of an inpatient geriatric service in a university hospital. Journal of the Tennessee Medical Association. 1994; 87(10):425-428

243 Widagdo IS, Pratt N, Russell M, Roughead EE. Predictive performance of four frailty measures in an older Australian population. Age and Ageing. 2015; 44(6):967-972

244 Williams A. Patients with comorbidities: perceptions of acute care services. Journal of Advanced Nursing. 2004; 46(1):13-22

245 Winograd CH, Gerety MB, Lai NA. A negative trial of inpatient geriatric consultation. Lessons learned and recommendations for future research. Archives of Internal Medicine. 1993; 153(17):2017-2023

246 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Archives of Internal Medicine. 2002; 162(20):2269-2276

247 Zekry D, Loures Valle BH, Graf C, Michel JP, Gold G, Krause KH et al. Prospective comparison of 6 comorbidity indices as predictors of 1-year post-hospital discharge institutionalization, readmission, and mortality in elderly individuals. Journal of the American Medical Directors Association. 2012; 13(3):272-278

248 Zeng C, Ellis JL, Steiner JF, Shoup JA, McQuillan DB, Bayliss EA. Assessment of morbidity over time in predicting health outcomes. Medical Care. 2014; 52 Suppl 3:S52-S59

249 Zuckerman IH, Langenberg P, Baumgarten M, Orwig D, Byrns PJ, Simoni-Wastila L et al. Inappropriate drug use and risk of transition to nursing homes among community-dwelling older adults. Medical Care. 2006; 44(8):722-730

14 Acronyms and abbreviations

Acronym or abbreviation	Description
ACE	Acute Care for Elders
ACGA	Abbreviated Comprehensive Geriatric Assessment
ACT	Australia Capital Territory
ADDQUOL	Audit of Diabetes Dependent Quality of Life
ADL	Activities of Daily Living
AGREE II	Appraisal of Guidelines for Research and Evaluation II
AQUOL	Assessment of Quality of Life
AUC	Area under the ROC Curve
BIPQ	Brief Illness Perception Questionnaire
BISEP	Burden of Illness Score for Elderly Persons
BMD	Bone Mineral Density
BME	Black and Minority Ethnicities
CA Program	Care Advocate Program
CARS	Community Assessment Risk Screen
CCI	Charlson Comorbidity Index
CDS	Chronic Disease Score
CDSMP	Chronic Disease Self-Management programme
CDSMP	Chronic disease Self-Management Programme
CERQUAL	Confidence in the Evidence from Reviews of Qualitative research
CGA	Comprehensive Geriatric Assessment
CHS	Cardiovascular Health Study
CHS	Cardiovascular Health Study
CSHA	Canadian Study for Health and Aging
CIRS	Cumulative Illness Rating
COPD	Chronic obstructive pulmonary disease
COPD	Chronic Obstructive Pulmonary Disease
COPM	Canadian Occupational Performance Measure
CPGs	Clinical Practice Guidelines
CT Scan	Computed Tomography
EPP	Expert Patient Programme
ESRD	End Stage Renal Disease
FCI	Functional Comorbidity Index
GCfbS	Geriatric Comorbidity Score
GEMU	Geriatric Evaluation and Management Unit
GFI	Groningen Frailty Indicator
GRACE support team	Geriatric Resources for Assessment and Care of Elders
GSCU	Geriatric Special Care Unit
HADS	Hospital Anxiety and Depression Scale
HARP	Hospital Admission Risk Profile
HF	Heart Failure
HOPE	Hospitalised Older Patient Examination

Acronym or abbreviation	Description
HSCIC	Health & Social Care Information Centre
HVNs	Home Visiting Nurses
IADL	Instrumental Activities of Daily Living
IPAQ	International Physical Activity Questionnaire
ISAR	Identifying Seniors at Risk
I-TEAM	Integrated Tele-health Education and Activation of Mood
MCS	Mental Component Score
MLWHF	Minnesota Living with Heart Failure
MMMSE	Modified Mini Mental State Examination
MMSE	Mini Mental State Examination
MPI	Multidimensional Prognostic Index
MRI	Magnetic Resonance Imaging
NCM	Nurse Case Manager
NEADL	Nottingham Extended Activities of Daily Living
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OARS	Older Americans Resources and Services
PACIC	Patient Satisfaction –Patient Assessment of Chronic Illness
PASE	Physical Activity Scale for the Elderly
PCS	Physical Component Score
PDSMS	Perceived Diabetes Self-Management Scale
PEONY	Predicting Emergency Admissions Over the Next Year
PSA screen	Prostate-specific Antigen screen
RGN	Registered General Nurse
RIGAMA	Risk Index for geriatric Acute Medical Admissions
SCDSMP	Stanford Chronic Disease Self-Management Group Programme
SCHFI	Self-care in Heart Failure Index
SOF	Study of Osteoporotic Fractures
STOPP/START	Screening Tool of Older People's potentially inappropriate Prescriptions/ Screening Tool to Alert doctors to Right Treatments
TBQ	Treatment Burden Questionnaire
TUG	Timed up and go test

15 Glossary

The NICE Glossary can be found at www.nice.org.uk/glossary.

15.1 Guideline-specific terms

Term	Definition
Anti-hypertensives	A class of drugs that are used to treat hypertension (high blood pressure).
Bisphosphonates	Drugs used to slow down or prevent bone damage.
Cardiovascular disease	A general term that describes a disease of the heart or blood vessels.
Care co-ordination	Either 1 individual (case management by a key worker) or 1 organisation (care management) takes the lead in organising a patient's person's care and support. This may include liaising with other healthcare professionals (for example, GP or specialist services) and, in general, focuses on continuity of care. The key worker or organisation may not necessarily be responsible for delivering any additional intervention.
Care plan	A care plan is an agreement between patient and health or social care professional to support management of day to day health and symptoms by the patient and other healthcare professionals and/or to organise care. It can be a written document and/or something recorded in patient notes.
Coaching	A method of patient education that guides and prompts a patient to be an active participant in behaviour change.
Collaborative care	A complex intervention with 4 key components: a multidisciplinary approach to patient care (including the use of non-medical case-managers); structured patient care plans; scheduled follow-ups; enhanced inter-professional communication.
Comprehensive Geriatric Assessment (CGA)	A comprehensive geriatric assessment is an interdisciplinary diagnostic process to determine the medical, psychological and functional capability of someone who is frail and old. The aim is to develop a coordinated, integrated plan for treatment and long-term support.
Creatinine	Creatinine is a waste product from the normal breakdown of muscle tissue. As creatinine is produced, it's filtered through the kidneys and excreted in urine.
Disease burden	The symptoms of the conditions a patient has may impact on patient functioning, quality of life and wellbeing.
Dyslipidemia	An abnormal amount of lipids (for example, cholesterol or fat) in the blood.
Frailty	The condition of being weak and delicate.
Geriatrics	Speciality that focuses on health care of elderly people. It aims to promote health by preventing and treating diseases and disabilities in older adults.
Gerontology	The scientific study of the biological, psychological, and sociological phenomena associated with old age and aging.
Haemoglobin	The iron-containing oxygen-transport metalloprotein in the red blood cells.
Home follow-up	As with telephone follow-up, home follow-up is defined here as pre-arranged visits by a healthcare professional to a patient's (or carer's) home, either as a one-off or on multiple occasions. The aim of follow-up will vary according to the patients' person's needs and context.
Holistic assessment	A trained healthcare professional performs a comprehensive assessment of a patient's person's physical health, mental health, social situation and functional ability (see Comprehensive Geriatric Assessment)
Hypercholesterolaemia	An excess of cholesterol in the bloodstream.

Term	Definition
Hyperlipidemia	Excess lipids (cholesterol, triglycerides, or both) in the blood. Hyperlipidemia increases the risk of high blood pressure, heart disease, and stroke.
Hypoglycemia	A condition characterized by abnormally low blood glucose (blood sugar) levels, usually less than 70 mg/dl.
Individualised management plan	Individualised management plan is a management plan covering clinical aspects of person's care such as medicines they are taking, the services they are attending and including information about which areas of care are most important to the patient and whether treatments have been stopped to reduce treatment burden.
Inpatient	A patient who lives in hospital while under treatment.
Medicines	Medicines includes topical treatments such as ointments, creams and drops, as well as medicines taken by mouth or injection.
Medicines management	A healthcare professional works collaboratively with the patient, and if necessary other members of the MDT, to optimise safe, effective and appropriate drug therapy. This may involve the healthcare professional checking patient's medicine-taking behaviour, concerns about side effects and reviewing the indications for medicines.
Multidisciplinary care	When professionals from a range of disciplines work together to deliver comprehensive care that addresses as many of the patient's needs as possible. This can be delivered by a range of professionals functioning as a team under 1 organisational umbrella or by professionals from a range of organisations, including private practice, brought together as a unique team.
Multimorbidity	Multimorbidity refers to the presence of 2 or more long-term health conditions, which can include: • defined physical and mental health conditions such as diabetes or schizophrenia • on-going conditions such as learning disability • symptom complexes such as frailty or chronic pain • sensory impairment such as sight or hearing loss • alcohol and substance misuse. The management of risk factors for future disease can be a major treatment burden for people with multimorbidity and should be carefully considered when optimising care. This guideline covers the optimisation of care for: • adults with 2 or more long-term physical health conditions • adults with 1 or more mental health condition and at least 1 physical health condition.
Multimorbidity approach to care (an approach to care that takes account of multimorbidity)	An approach to care that takes account of multimorbidity, involves personalised assessment and the development of an individualised management plan. The aim is to improve quality of life by reducing treatment burden, adverse events, and unplanned or uncoordinated care. The approach takes account of a person's individual needs, preferences for treatments, health priorities and lifestyle. It aims to improve coordination of care across services, particularly if this has become fragmented.
Occupational therapy	An intervention for people whose health prevents them from doing certain activities (for example dressing or getting to the shops), where an occupational therapist works with them to implement practical solutions (for example changing their environment, using new equipment).
Osteoporosis	A medical condition in which bones become brittle and fragile from loss of tissue, usually as a result of hormonal changes or calcium/vitamin D deficiency.

Term	Definition
Outpatient	A patient who attends a hospital for treatment without staying there overnight.
Patient-centred care	Care that takes into account a patient's individual needs and preferences, where patients have the opportunity to make informed decisions about their care and treatment, in partnership with their healthcare professionals.
Patient decision aid (PDA)	A tool that presents evidence-based estimates of the benefits and risks of the available treatment options in sufficient detail that people are better able to judge their value. PDAs are tailored to a person's health status and help them to make specific, personal choices about their treatment.
Peer support	Support from someone who has the same condition and has trained to be a peer support worker.
Polypharmacy	The administration of multiple drugs at the same time
Primary prevention	Treatments aimed at preventing the onset of conditions.
Prophylactic treatment	A treatment to prevent a disease from occurring.
Psychoeducation	Education offered to individuals with a mental health condition and their families to help empower them and deal with their condition in an optimal way.
Secondary prevention	Treatments aimed at preventing the exacerbation of existing conditions.
Self-management	An intervention aimed at increasing a patient's person's ability to manage their own condition without the need for intensive support from healthcare professionals.
Self-monitoring	Where patients observe and keep a record of their own symptoms.
Shared decision making	Shared decision-making is a process in which patients are actively involved in decision making, supported by healthcare professionals to make fully informed choices about investigations, treatment and care that reflect their preferences.
Statins	A group of medicines that can help lower the level of low-density lipoprotein (LDL) cholesterol in the blood.
Stepped care	Stepped care provides a framework in which to organise the provision of services supporting patients, carers and healthcare professionals in identifying and accessing the most effective interventions. Stepped care is a system for delivering and monitoring treatment with the explicit aim of providing the most effective, yet least burdensome, treatment to the a person first, and which has a self-correcting mechanism built in so if a person does not benefit from an initial intervention they are 'stepped up' to a more complex intervention. Typically, stepped care starts by providing low-intensity interventions.
Tapering regimen	A regime where doses of medication is reduced gradually over a period of time.
Telehealth	Telehealth is the delivery of health-related services and information via telecommunications technologies.
Telemonitoring	Telemonitoring involves remotely monitoring patients who are not at the same location as the health care provider
Telephone follow-up	In this review, telephone-follow up is defined as a pre-arranged telephone call to a patient (or carer) by a healthcare professional. This may be a one-off or multiple telephone calls depending on the aim of follow-up.
Treatment Burden	The self-care practices that patients with chronic illness must perform to respond to the requirements of their healthcare providers, as well as the impact that these practices have on patient functioning, quality of life and wellbeing.

15.2 General terms

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in an RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Annualised data	Annualised data (such as annualised numbers needed to treat (NNT) and annualised absolute difference) is an adjustment in the effect estimate so that it represents the effect that would occur over a 12-month follow-up period. If an effect estimate has been derived from studies with a shorter or longer follow-up than 12-months, the effect estimate is adjusted using the assumption that the effect will be constant over time.
Applicability	How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered.
Arm (of a clinical study)	Subsection of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between 2 or more events, characteristics or other variables. The relationship may or may not be causal.
Base case analysis	In an economic evaluation, this is the main analysis based on the most plausible estimate of each input. In contrast, see Sensitivity analysis.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Bayesian analysis	A method of statistics, where a statistic is estimated by combining established information or belief (the 'prior') with new evidence (the 'likelihood') to give a revised estimate (the 'posterior').
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Influences on a study that can make the results look better or worse than they really are. (Bias can even make it look as if a treatment works when it does not.) Bias can occur by chance, deliberately or as a result of systematic errors in the design and execution of a study. It can also occur at different stages in the research process, for example, during the collection, analysis, interpretation, publication or review of research data. For examples see selection bias, performance bias, information bias, confounding factor, and publication bias.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias. A single-blinded study is one in which patients do not know which study group they are in (for example whether they are taking the experimental drug or a placebo). A double-blinded study is one in which

Term	Definition
	neither patients nor the researchers and doctors know which study group the patients are in. A triple blind study is one in which neither the patients, clinicians or the people carrying out the statistical analysis know which treatment patients received.
Carer (caregiver)	Someone who looks after family, partners or friends in need of help because they are ill, frail or have a disability.
Case–control study	A study to find out the cause(s) of a disease or condition. This is done by comparing a group of patients who have the disease or condition (cases) with a group of people who do not have it (controls) but who are otherwise as similar as possible (in characteristics thought to be unrelated to the causes of the disease or condition). This means the researcher can look for aspects of their lives that differ to see if they may cause the condition. For example, a group of people with lung cancer might be compared with a group of people the same age that do not have lung cancer. The researcher could compare how long both groups had been exposed to tobacco smoke. Such studies are retrospective because they look back in time from the outcome to the possible causes of a disease or condition.
Case series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical effectiveness	How well a specific test or treatment works when used in the 'real world' (for example, when used by a doctor with a patient at home), rather than in a carefully controlled clinical trial. Trials that assess clinical effectiveness are sometimes called management trials. Clinical effectiveness is not the same as efficacy.
Clinician	A healthcare professional who provides patient care. For example, a doctor, nurse or physiotherapist.
Cochrane Review	The Cochrane Library consists of a regularly updated collection of evidence-based medicine databases including the Cochrane Database of Systematic Reviews (reviews of randomised controlled trials prepared by the Cochrane Collaboration).
Cohort study	A study with 2 or more groups of people – cohorts – with similar characteristics. One group receives a treatment, is exposed to a risk factor or has a particular symptom and the other group does not. The study follows their progress over time and records what happens. See also observational study.
Comorbidity	A disease or condition that someone has in addition to the health problem being studied or treated.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confidence interval (CI)	There is always some uncertainty in research. This is because a small group of patients is studied to predict the effects of a treatment on the wider population. The confidence interval is a way of expressing how

Term	Definition
	certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. The CI is usually stated as '95% CI', which means that the range of values has a 95 in a 100 chance of including the 'true' value. For example, a study may state that "based on our sample findings, we are 95% certain that the 'true' population blood pressure is not higher than 150 and not lower than 110". In such a case the 95% CI would be 110 to 150. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment – often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Confounding factor	Something that influences a study and can result in misleading findings if it is not understood or appropriately dealt with. For example, a study of heart disease may look at a group of people that exercises regularly and a group that does not exercise. If the ages of the people in the 2 groups are different, then any difference in heart disease rates between the 2 groups could be because of age rather than exercise. Therefore age is a confounding factor.
Consensus methods	Techniques used to reach agreement on a particular issue. Consensus methods may be used to develop NICE guidance if there is not enough good quality research evidence to give a clear answer to a question. Formal consensus methods include Delphi and nominal group techniques.
Construct validity	A metric of whether a questionnaire measures the intended construct. Usually assessed by whether the questionnaire is correlated with constructs in a way that would be expected; i.e. highly correlated with other validated measures of the same or related construct, and poorly correlated with other validated measures of unrelated constructs.
Control group	A group of people in a study who do not receive the treatment or test being studied. Instead, they may receive the standard treatment (sometimes called 'usual care') or a dummy treatment (placebo). The results for the control group are compared with those for a group receiving the treatment being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the treatment group, to make it as easy as possible to detect any effects due to the treatment.
Cost-consequences analysis (CCA)	Cost-consequences analysis is one of the tools used to carry out an economic evaluation. This compares the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of a test or treatment with a suitable alternative. Unlike cost-benefit analysis or cost-effectiveness analysis, it does not attempt to summarise outcomes in a single measure (like the quality-adjusted life year) or in financial terms. Instead, outcomes are shown in their natural units (some of which may be monetary) and it is left to decision-makers to determine whether, overall, the treatment is worth carrying out.
Cost-effectiveness analysis (CEA)	Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical

Term	Definition
	decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost-utility analysis (CUA)	Cost-utility analysis is one of the tools used to carry out an economic evaluation. The benefits are assessed in terms of both quality and duration of life, and expressed as quality-adjusted life years (QALYs). See also utility.
Credible interval (CrI)	The Bayesian equivalent of a confidence interval.
Decision analysis	An explicit quantitative approach to decision-making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.
Deterministic analysis	In economic evaluation, this is an analysis that uses a point estimate for each input. In contrast, see Probabilistic analysis
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disutility	The loss of quality of life associated with having a disease or condition. See Utility
Dominance	A health economics term. When comparing tests or treatments, an option that is both less effective and costs more is said to be 'dominated' by the alternative.
Drop-out	A participant who withdraws from a trial before the end.
Economic evaluation	An economic evaluation is used to assess the cost-effectiveness of healthcare interventions (that is, to compare the costs and benefits of a healthcare intervention to assess whether it is worth doing). The aim of an economic evaluation is to maximise the level of benefits – health effects – relative to the resources available. It should be used to inform and support the decision-making process; it is not supposed to replace the judgement of healthcare professionals. There are several types of economic evaluation: cost-benefit analysis, cost-consequences analysis, cost-effectiveness analysis, cost-minimisation analysis and cost-utility analysis. They use similar methods to define and evaluate costs, but differ in the way they estimate the benefits of a particular drug, programme or intervention.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	A measure that shows the magnitude of the outcome in one group compared with that in a control group. For example, if the absolute risk reduction is shown to be 5% and it is the outcome of interest, the effect size is 5%. The effect size is usually tested, using statistics, to find out how likely it is that the effect is a result of the treatment and has not just happened by chance (that is, to see if it is statistically significant).
Effectiveness	How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.
Efficacy	How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.
Epidemiological study	The study of a disease within a population, defining its incidence and

Term	Definition
	prevalence and examining the roles of external influences (for example, infection, diet) and interventions.
EQ-5D (EuroQol 5 dimensions)	A standardised instrument used to measure health-related quality of life. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals or patients).
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Extrapolation	An assumption that the results of studies of a specific population will also hold true for another population with similar characteristics.
Follow-up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health-related variables.
Generalisability	The extent to which the results of a study hold true for groups that did not participate in the research. See also external validity.
Gold standard	A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease.
GRADE, GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	Study or analysis of the cost of using and distributing healthcare resources.
Health-related quality of life (HRQoL)	A measure of the effects of an illness to see how it affects someone's day-to-day life.
Heterogeneity or Lack of homogeneity	The term is used in meta-analyses and systematic reviews to describe when the results of a test or treatment (or estimates of its effect) differ significantly in different studies. Such differences may occur as a result of differences in the populations studied, the outcome measures used or because of different definitions of the variables involved. It is the opposite of homogeneity.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The extra cost linked to using one test or treatment rather than another. Or the additional cost of doing a test or providing a treatment more frequently.
Incremental cost-effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.

Term	Definition
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a clinical trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully complied with the treatment or switched to an alternative treatment. Intention-to-treat analyses are often used to assess clinical effectiveness because they mirror actual practice: that is, not everyone complies with treatment and the treatment people receive may be changed according to how they respond to it.
Internal reliability	A measure of how well individual items in a scale correlate with each other; therefore indicating whether items are assessing the same construct. Typically assessed using Cronbach's alpha.
Interpretability	A measure of how easily a questionnaire can be used and interpreted in clinical practice. For example, how clinicians and patients should interpret scores on the questionnaire, and what change in scores is clinically meaningful.
Intervention	In medical terms this could be a drug treatment, surgical procedure, diagnostic or psychological therapy. Examples of public health interventions could include action to help someone to be physically active or to eat a more healthy diet.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by (1 minus specificity).
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Logistic regression or Logit model	In statistics, logistic regression is a type of analysis used for predicting the outcome of a binary dependent variable based on one or more predictor variables. It can be used to estimate the log of the odds (known as the 'logit').
Loss to follow-up	A patient, or the proportion of patients, actively participating in a clinical trial at the beginning, but whom the researchers were unable to trace or contact by the point of follow-up in the trial
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
Meta-analysis	A method often used in systematic reviews. Results from several studies of the same test or treatment are combined to estimate the overall effect of the treatment.
Minimally important difference (MID)	The minimally important difference (MID) is the threshold at which a change in the effect estimate is decided to be clinically meaningful; that is, it indicates a clinical benefit or harm that is important to patients.

Term	Definition
Multivariate model	A statistical model for analysis of the relationship between 2 or more predictor (independent) variables and the outcome (dependent) variable.
Negative predictive value (NPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a negative test result who do not have the disease, and can be interpreted as the probability that a negative test result is correct. It is calculated as follows: NPV = number of true negatives/(number of true negatives + number of false negatives)
Number needed to treat (NNT)	The average number of patients who need to be treated to get a positive outcome. For example, if the NNT is 4, then 4 patients would have to be treated to ensure 1 of them gets better. The closer the NNT is to 1, the better the treatment. For example, if you give a stroke prevention drug to 20 people before 1 stroke is prevented, the number needed to treat is 20. See also number needed to harm, absolute risk reduction.
Observational study	Individuals or groups are observed or certain factors are measured. No attempt is made to affect the outcome. For example, an observational study of a disease or treatment would allow 'nature' or usual medical care to take its course. Changes or differences in one characteristic (for example, whether or not people received a specific treatment or intervention) are studied without intervening. There is a greater risk of selection bias than in experimental studies.
Odds ratio	Odds are a way to represent how likely it is that something will happen (the probability). An odds ratio compares the probability of something in one group with the probability of the same thing in another. An odds ratio of 1 between 2 groups would show that the probability of the event (for example a person developing a disease, or a treatment working) is the same for both. An odds ratio greater than 1 means the event is more likely in the first group. An odds ratio less than 1 means that the event is less likely in the first group. Sometimes probability can be compared across more than 2 groups – in this case, one of the groups is chosen as the 'reference category', and the odds ratio is calculated for each group compared with the reference category. For example, to compare the risk of dying from lung cancer for non-smokers, occasional smokers and regular smokers, non-smokers could be used as the reference category. Odds ratios would be worked out for occasional smokers compared with non-smokers and for regular smokers compared with non-smokers. See also confidence interval, risk ratio.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	The impact that a test, treatment, policy, programme or other intervention has on a person, group or population. Outcomes from interventions to improve the public's health could include changes in knowledge and behaviour related to health, societal changes (for example, a reduction in crime rates) and a change in people's health and wellbeing or health status. In clinical terms, outcomes could include the number of patients who fully recover from an illness or the number of hospital admissions, and an improvement or deterioration in someone's health, functional ability, symptoms or situation. Researchers should decide what outcomes to measure before a study

Term	Definition
	begins.
P value	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that one seems more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance) it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Placebo	A fake (or dummy) treatment given to participants in the control group of a clinical trial. It is indistinguishable from the actual treatment (which is given to participants in the experimental group). The aim is to determine what effect the experimental treatment has had – over and above any placebo effect caused because someone has received (or thinks they have received) care or attention.
Polypharmacy	The use or prescription of multiple medications.
Posterior distribution	In Bayesian statistics this is the probability distribution for a statistic based after combining established information or belief (the prior) with new evidence (the likelihood).
Positive predictive value (PPV)	In screening or diagnostic tests: A measure of the usefulness of a screening or diagnostic test. It is the proportion of those with a positive test result who have the disease, and can be interpreted as the probability that a positive test result is correct. It is calculated as follows: PPV = number of true positives/(number of true positives + number of false positives)
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Pre-test probability	In diagnostic tests: The proportion of people with the target disorder in the population at risk at a specific time point or time interval. Prevalence may depend on how a disorder is diagnosed.
Prevalence	See Pre-test probability.
Prior distribution	In Bayesian statistics this is the probability distribution for a statistic based on previous evidence or belief.
Primary care	Healthcare delivered outside hospitals. It includes a range of services provided by GPs, nurses, health visitors, midwives and other healthcare professionals and allied health professionals such as dentists, pharmacists and opticians.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Probabilistic analysis	In economic evaluation, this is an analysis that uses a probability distribution for each input. In contrast, see Deterministic analysis.
Product licence	An authorisation from the MHRA to market a medicinal product.
Prognosis	A probable course or outcome of a disease. Prognostic factors are patient or disease characteristics that influence the course. Good prognosis is associated with low rate of undesirable outcomes; poor prognosis is associated with a high rate of undesirable outcomes.

Term	Definition
Prospective study	A research study in which the health or other characteristic of participants is monitored (or 'followed up') for a period of time, with events recorded as they happen. This contrasts with retrospective studies.
Publication bias	Publication bias occurs when researchers publish the results of studies showing that a treatment works well and don't publish those showing it did not have any effect. If this happens, analysis of the published results will not give an accurate idea of how well the treatment works. This type of bias can be assessed by a funnel plot.
Quality of life	See 'Health-related quality of life'.
Quality-adjusted life year (QALY)	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to 1 year of life in perfect health. QALYS are calculated by estimating the years of life remaining for a patient following a particular treatment or intervention and weighting each year with a quality of life score (on a scale of 0 to 1). It is often measured in terms of the person's ability to perform the activities of daily life, freedom from pain and mental disturbance.
Randomisation	Assigning participants in a research study to different groups without taking any similarities or differences between them into account. For example, it could involve using a random numbers table or a computer-generated random sequence. It means that each individual (or each group in the case of cluster randomisation) has the same chance of receiving each intervention.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug or treatment. One group (the experimental group) receives the treatment being tested, the other (the comparison or control group) receives an alternative treatment, a dummy treatment (placebo) or no treatment at all. The groups are followed up to see how effective the experimental treatment was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.
RCT	See 'Randomised controlled trial'.
Receiver operated characteristic (ROC) curve	A graphical method of assessing the accuracy of a diagnostic test. Sensitivity is plotted against 1 minus specificity. A perfect test will have a positive, vertical linear slope starting at the origin. A good test will be somewhere close to this ideal.
Reference standard	The test that is considered to be the best available method to establish the presence or absence of the outcome – this may not be the one that is routinely used in practice.
Reporting bias	See 'Publication bias'.
Reproducibility	A measure of the reliability of a questionnaire. Assessed by whether participants' scores on a questionnaire are stable over time (test-retest reliability).
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Responsiveness	A measure of whether a questionnaire is able to detect changes in the intended construct; for example, if changes in the construct occur following an intervention.
Retrospective study	A research study that focuses on the past and present. The study examines past exposure to suspected risk factors for the disease or condition. Unlike prospective studies, it does not cover events that

Term	Definition
	occur after the study group is selected.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Risk ratio (RR)	The ratio of the risk of disease or death among those exposed to certain conditions compared with the risk for those who are not exposed to the same conditions (for example, the risk of people who smoke getting lung cancer compared with the risk for people who do not smoke). If both groups face the same level of risk, the risk ratio is 1. If the first group had a risk ratio of 2, subjects in that group would be twice as likely to have the event happen. A risk ratio of less than 1 means the outcome is less likely in the first group. The risk ratio is sometimes referred to as relative risk.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	Selection bias occurs if: a) The characteristics of the people selected for a study differ from the wider population from which they have been drawn, or b) There are differences between groups of participants in a study in terms of how likely they are to get better.
Sensitivity	How well a test detects the thing it is testing for. If a diagnostic test for a disease has high sensitivity, it is likely to pick up all cases of the disease in people who have it (that is, give a 'true positive' result). But if a test is too sensitive it will sometimes also give a positive result in people who don't have the disease (that is, give a 'false positive'). For example, if a test were developed to detect if a woman is 6 months pregnant, a very sensitive test would detect everyone who was 6 months pregnant, but would probably also include those who are 5 and 7 months pregnant. If the same test were more specific (sometimes referred to as having higher specificity), it would detect only those who are 6 months pregnant, and someone who was 5 months pregnant would get a negative result (a 'true negative'). But it would probably also miss some people who were 6 months pregnant (that is, give a 'false negative'). Breast screening is a 'real-life' example. The number of women who are recalled for a second breast screening test is relatively high because the test is very sensitive. If it were made more specific, people who don't have the disease would be less likely to be called back for a second test but more women who have the disease would be missed.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariate analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): 2 or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or

Term	Definition
	below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (for example, Monte Carlo simulation).
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p<0.05$).
Specificity	The proportion of true negatives that are correctly identified as such. For example in diagnostic testing the specificity is the proportion of non-cases correctly diagnosed as non-cases. See related term 'Sensitivity'. In terms of literature searching a highly specific search is generally narrow and aimed at picking up the key papers in a field and avoiding a wide range of papers.
Stakeholder	An organisation with an interest in a topic that NICE is developing a clinical guideline or piece of public health guidance on. Organisations that register as stakeholders can comment on the draft scope and the draft guidance. Stakeholders may be: • manufacturers of drugs or equipment • national patient and carer organisations • NHS organisations • organisations representing healthcare professionals.
Systematic review	A review in which evidence from scientific studies has been identified, appraised and synthesised in a methodical way according to predetermined criteria. It may include a meta-analysis.
Test-retest reliability	See reproducibility
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Transition probability	In a state transition model (Markov model), this is the probability of moving from one health state to another over a specific period of time.
Treatment allocation	Assigning a participant to a particular arm of a trial.
Univariate	Analysis which separately explores each variable in a data set.
User friendliness	A measure of how easy and acceptable a questionnaire is to complete; for example, whether the items are unambiguous and the scales easy to complete.
Utility	In health economics, a 'utility' is the measure of the preference or value that an individual or society places upon a particular health state. It is generally a number between 0 (representing death) and 1 (perfect health). The most widely used measure of benefit in cost–utility analysis is the quality-adjusted life year, but other measures include disability-adjusted life years (DALYs) and healthy year equivalents (HYEs).