NICE real-world evidence framework

# Overview

## Key messages

The [NICE Strategy 2021 to 2026](https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026) states our ambition to use real-world data to resolve gaps in knowledge and drive forward access to innovations for patients. Real-world data is essential to enabling rapid, robust, and responsive technology evaluations and dynamic, living guidelines.

* We developed the Real-World Evidence Framework to help deliver on this ambition. It does this by:
* Identifying when real-world data can be used to reduce uncertainties and improve guidance, including technology evaluations and guidelines
* clearly describing best-practices for planning, conducting and reporting real-world evidence studies to improve the quality and transparency of evidence.
* The framework provides in-depth guidance and tools to support the implementation of these core principles across different uses of real-world evidence. It is structured as follows:
* the [introduction](#_Introduction_to_real-world) provides background on real-world data and real-world evidence, discusses its strengths and weaknesses, and summarises current and potential uses within NICE guidance
* [study conduct](#_Conduct_of_quantitative_1) describes NICE’s expectations for planning, conducting and reporting real-world evidence studies, recognising that acceptability of evidence will depend on the type of evidence and other contextual factors
* [assessing data suitability](#_Assessing__data) describes the information needed to assess data provenance and its quality and relevance for addressing specific research questions
* [methods for real-world studies of comparative effects](#_Methods_for_real-world_1) provides more specific recommendations for conducting non-randomised studies. These include traditional observational studies as well as clinical trials that use real-world data to form an external control.
* [Table 1](#_Table_1) summarises key considerations for conducting real-world evidence studies. The following core principles should be followed to generate high-quality and trusted real-world evidence:
* ensure data is of good provenance, relevant and of sufficient quality to answer the research question
* generate evidence in a transparent way and with integrity from study planning through to study conduct and reporting
* use analytical methods that minimise the risk of bias and characterise uncertainty.
* The framework is a living framework that will be updated periodically to reflect user feedback, learnings from implementation including exemplar case studies, developments in real-world evidence methodology, and to extend its scope to include additional guidance on priority topics.
* We encourage companies planning to use real-world data in their submissions to engage early with [NICE Scientific Advice](https://www.nice.org.uk/about/what-we-do/life-sciences/scientific-advice) on how to make best use of real-world data as part of their evidence-generation plans.

## Table 1

Summary of key considerations in planning, conducting and reporting real-world evidence studies

| Stage of evidence generation | Key considerations |
| --- | --- |
| [Planning](#_Study_planning_1) | * Clearly define the research question including, as relevant, definitions of population eligibility criteria, interventions, outcomes and the target quantity of estimation * Plan the study prospectively and make protocols publicly available * Choose data that is of good provenance and of sufficient quality and relevance to address the research question * Justify the need for further primary data collection, weighing up the burden on patients and healthcare professionals against the value of additional data * Use data in accordance with local law, governance processes, codes of practice and the requirements of the data owner |
| [Conduct](#_Study_conduct_2) | * Use a study design and statistical methods appropriate to the research question, considering the key risks of bias * Use sensitivity and/or bias analysis to assess the robustness of studies to key risks of bias and uncertain data curation or analytical decisions * Do quality assurance to ensure the integrity and quality of the study |
| [Reporting](#_Study_reporting) | * Report study design and analytical methods in sufficient detail to enable independent researchers to fully understand what was done and why, critically appraise the study and reproduce it * Reporting should also cover:   + provenance, quality, and relevance of the data (see [assessing data suitability](#_Chapter__))   + data curation   + patient attrition from initial data to the final analyses   + characteristics of patients (including missing data) and details of follow up overall and across key population groups   + results for all planned and conducted analyses (clearly indicating any analyses that were not pre-planned)   + assessment of risk of bias and generalisability to the target population in the NHS |

## Real-world data and its role in NICE guidance

* Real-world data refers to data collected outside of highly-controlled trials. It can come from many different sources including patient health records, administrative records, patient registries, surveys, observational cohort studies and digital health technologies.
* Real-world data is already widely used to inform NICE guidance to, for example:
* characterise health conditions, interventions, care pathways and patient outcomes and experiences
* design, populate and validate economic models (including estimates of resource use, quality of life, event rates, prevalence, incidence and long-term outcomes)
* develop or validate digital health technologies
* identify, characterise and address [health inequalities](https://www.nice.org.uk/Glossary?letter=H#Health%20inequalities)
* understand the safety of medical technologies including medicines, devices and interventional procedures
* assess the impact of interventions (including tests) on service delivery and decisions about care
* assess the applicability of clinical trials to patients in the NHS.
* Real-world data that represents the population of interest is NICE’s preferred source of evidence for most of these applications. Such data is regularly used for these purposes in NICE guidance but its use could be more commonplace, especially of routinely collected data.
* Randomised controlled trials are the preferred source of evidence on the effectiveness of interventions. Randomisation ensures that any differences in baseline characteristics between groups are because of chance. Blinding (if applied) prevents knowledge of treatment allocation from influencing behaviours. However, randomised trials are sometimes unavailable or are not directly relevant to decisions about patient care in the NHS.
* Randomised trials may not be available for several reasons, including:
* randomisation is considered unethical or unfeasible (for instance, for some rare or severe diseases with unmet need)
* technical challenges make randomisation impractical, which is most common for medical devices and interventional procedures
* funding is not available for a trial (for example, when the intervention is already used in routine practice).
* Even if randomised evidence is available, it may not be sufficient for decision making in the NHS for several reasons including:
* the comparator does not reflect standard of care in the NHS
* relevant population groups are excluded
* there are major differences in patient behaviours, care pathways or settings that differ from implementation in routine practice
* follow up is limited
* unvalidated surrogate outcomes are used
* learning effects are present
* trials were of poor quality.
* Non-randomised studies are already widely used to estimate the effects of medical devices and procedures and public health interventions, for which trials are less common. They are becoming more widely used in initial assessments of medicines, as more are granted regulatory approval based on uncontrolled single-arm trials. Finally, the increased focus on the lifecycle of technologies and lived experiences of patients relies on non-randomised studies after initial approvals. The most common non-randomised studies using real-world data to assess comparative effects are observational cohort studies and single-arm trials with real-world external control.
* Real-world data could be used to a greater extent to fill evidence gaps and speed up patient access but there are important barriers to its use, including:
* challenges in accessing high-quality data in a timely manner
* concerns about data provenance and quality
* risk of bias from information limitations, selection into or exit from studies, and confounding by indication in comparative studies
* limited trust in the integrity of some real-world evidence studies because of the complexity of evidence generation and opportunity to ‘cherry-pick’ results.

We are communicating our view on best-practices for the conduct of real-world evidence studies to ensure they are generated transparently and are of good quality. This is essential to improving trust in real-world evidence studies and their use in decision making.

# Introduction to real-world evidence in NICE decision making

## What is real-world data?

We define real-world data as data collected outside the context of a highly-controlled [clinical trial](https://www.nice.org.uk/Glossary?letter=C#Clinical%20trial). Real-world data can be routinely collected during the delivery of health or social care. It can also be collected prospectively, to address 1 or more specific research questions. Most real-world data sources are observational (or non-interventional), that is, any interventions (or exposures) are not determined by a study protocol. But medical interventions are decided by patients and healthcare professionals. And in public health or social care, interventions may be determined by individual behaviours, environmental exposures or policy makers.

Some interventional studies, such as pragmatic clinical trials, can also produce real-world evidence. Such trials may also make use of real-world data sources to design trials, recruit participants or collect outcome data. For more information, see the [UK Medicines and Healthcare products Regulatory Agency’s (MHRA) guideline on randomised controlled trials using real-world data to support regulatory decisions](https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guideline-on-randomised-controlled-trials-using-real-world-data-to-support-regulatory-decisions).

[Table 2](#Table1) describes common sources of non-interventional real-world data. These include original data collections (such as patient health records) and curated data sources (such as retrospective chart reviews). While each type of data source has some general strengths and weaknesses, the value for a given research question will depend on the characteristics of the specific data (for further information, see [assessing data suitability](#_Assessingment_of_Data)). Different sources of real-world data can be combined to improve data quality and coverage, potentially allowing additional research questions to be answered.

Real-world data can be quantitative or qualitative. Common data types include patient demographics, health behaviours, medical history, clinical outcomes (including patient-reported outcomes), patient or user experiences, resource use, costs, omics, laboratory measurements, imaging, free text, test results and patient-generated data. We consider both national data collections and international data in decision making.

## Table 2

Common sources of real-world data

| Data source | Description | Examples |
| --- | --- | --- |
| Electronic health records | Computerised individual patient records. These are typically used to inform the clinical management of patients.  These sometimes integrate data from other information systems including laboratory, genomic, and imaging systems | The [Clinical Practice Research Datalink (CPRD)](https://www.cprd.com/) GOLD contains demographic and clinical information on patients enrolled in participating general practices across the UK. |
| Administrative data | Data collected for administrative purposes. | The [Hospital Episode Statistics](https://datadictionary.nhs.uk/supporting_information/hes_data_dictionary.html) Admitted Patient Care dataset contains information on diagnoses and procedures done for all patients admitted to NHS hospitals or NHS funded treatments in private hospitals. Its primary purpose is to inform the reimbursement of hospitals through payment by results and other operational activities. |
| Claims data | Data on healthcare service use collected from insurance-based systems such as in the US. | The [Medicare data](https://www.cms.gov/Research-Statistics-Data-and-Systems/Research-Statistics-Data-and-Systems) contains data on individuals in receipt of Medicare services derived from reimbursement information or payment of bills. |
| Patient registries | Registries are organised systems that collect uniform data (clinical and other) to identify specified outcomes for a population defined by a particular disease, condition or exposure.  Registries can serve several purposes including research, clinical care or policy. Registries can include interventional studies. | The [Systemic Anti-Cancer Therapy (SACT) registry](https://pubmed.ncbi.nlm.nih.gov/31340008/) contains information on all patients treated with anticancer therapies from NHS England providers. This data is widely used within NICE to provide information on drugs approved for use within the Cancer Drugs Fund.  The [UK Cystic Fibrosis Registry](https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry) collects data on consenting people with cystic fibrosis across specialist centres in the UK. The registry data is used to improve the health of people with cystic fibrosis by facilitating research, guiding quality improvement at care centres and monitoring the safety of new drugs. |
| Patient generated health data | Data generated directly by patients or their carers including from wearable medical or personal devices, mobile apps and social media. Data can be collected actively or passively. | Pulse oximeters used to monitor people with COVID-19 treated at home to alert to need for hospital admission ([Greenhalgh et al. 2021](https://pubmed.ncbi.nlm.nih.gov/33766809/)).  Self-reported data on COVID and long-COVID symptoms from the [ZOE app](https://covid.joinzoe.com/). |
| Chart reviews | Data extracted retrospectively from a review of patient health records (including paper or electronic records).  Chart reviews are widely used in natural history studies. They may allow the extraction of data not reported in routine data sources. | Retrospective chart reviews are especially common in studies of rare diseases to model natural history of disease and treatment pathways ([Garbade et al. 2021](https://pubmed.ncbi.nlm.nih.gov/32845020/)). |
| Audit and service evaluation | Clinical audits are done to understand how current standards of care measure against best practice or a set standard, and subsequently inform quality improvement. Data can be collected prospectively or retrospectively  Service evaluations are done to define and judge current care. | The [Healthcare Quality Improvement Partnership](https://www.hqip.org.uk/) manages national clinical audit programmes such as the [Sentinel Stroke National Audit Programme (SSNAP)](https://www.strokeaudit.org/). SSNAP is used to assess the quality of the organisation and delivery of multidisciplinary inpatient stroke health services in England, Wales and Northern Ireland. |
| Observational cohorts with primary data collection | Traditional prospective studies designed to answer one or more research questions. | The [UK Biobank](https://www.ukbiobank.ac.uk/enable-your-research/about-our-data) collects data on patient medical histories and genetics. It links to patient records for health outcomes. It was not designed for a specific research question but to enable epidemiological research.  [EMBRACE-I](https://pubmed.ncbi.nlm.nih.gov/33794207/) is a multicentre prospective cohort study to evaluate local tumour control and morbidity in patient undergoing MRI-based IAGBT (image guided adaptive brachytherapy) for locally advanced cervical tumours. |
| Health surveys, interviews and focus groups | Health surveys involve systematic collection of data about health and disease in a human population through surveys. They have various purposes including understanding trends in health in a population or understanding patients’ experiences of care.  Interviews and focus groups are done to collect qualitative data such as patient perception and experiences. | The [Health Survey for England](https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england) is an annual representative household survey measuring trends in health in England.  The [‘Living with Lipoedema’ 2021 survey](https://www.lipoedema.co.uk/liposuction/?s=#:~:text=Lipoedema%20UK%E2%80%99s%20May%202021%20Survey%20%E2%80%98Living%20With%20Lipoedema-,Treatments%E2%80%99%20was%20completed%20by%20933%20women%20with%20lipoedema.) by patient charity Lipoedema UK collects patient experience data from individuals with lipoedema. It evaluates experiences of patients having non-cosmetic liposuction or other treatments for lipoedema. |

## What is real-world evidence?

We define real-world evidence as evidence generated from the analysis of real-world data. It can cover a large array of evidence types including disease epidemiology, health service research or causal estimation (see [use cases for real-world data in NICE guidance](#_Use_cases_for)). It can be generated from a large range of study designs and analytical methods (including quantitative and [qualitative methods](https://www.nice.org.uk/Glossary?letter=Q#Qualitative%20research)) depending on the research question or use case. A real-world evidence study may use routinely collected data, bespoke data collection, or a combination of the two. We consider single-arm trials that use real-world data sources to create an external control to be real-world evidence studies.

## Uses of real-world evidence in NICE guidance

### NICE guidance

NICE has several guidance products that use the best available evidence to develop recommendations that guide decisions in health, public health and social care, including:

* guidelines for clinical, social care and public health topics, which offer advisory guidance to health and social care professionals
* evaluations of medical technologies including medicines, diagnostics, medical devices, digital health technologies and interventional procedures.

Guidelines are developed internally by NICE. Technology evaluations are usually informed by company submissions but may also use evidence submitted by manufacturers or other stakeholders or research commissioned from independent academic centres.

The processes and methods for technology evaluations differ across NICE’s programmes. The technology appraisal programme evaluates mostly medicines (including highly specialised technologies) but can also include medical devices and diagnostics. The technology appraisal and diagnostic guidance programmes both consider the cost-effectiveness of medical technologies. The medical technology evaluation programme evaluates medical technologies including medical devices, digital health technologies, and diagnostics that are expected to be cost-saving or cost-neutral and uses cost-consequence analysis considering patient and system outcomes. The interventional procedures programme evaluates the efficacy and safety of interventional procedures without analysis of cost.

When NICE recommends a treatment 'as an option' through its technology appraisal programmes, the NHS must make sure it is available within 3 months (unless otherwise specified) of its date of publication. If a technology is potentially cost-effective but there is substantial and resolvable uncertainty about its value, it can be recommended for use in a managed access agreement. After a specified period of data collection the technology is reassessed through the technology appraisal programme. Selected devices, diagnostic or digital technologies that are recommended in NICE guidance and are likely to be affordable and produce cost savings within 12 months of adoption can be funded through NHS England’s [MedTech funding mandate](https://www.england.nhs.uk/aac/what-we-do/how-can-the-aac-help-me/the-medtech-funding-mandate/).

### Use cases for real-world data

The differences between NICE’s guidance programmes lead to variation in the uses and acceptability of real-world evidence.

Some common uses of real-world data across NICE programmes, with examples from previous NICE guidance, are:

* characterising health conditions, interventions, care pathways, and patient outcomes and experiences including natural history ([Onasemnogene abeparvovec for treating spinal muscular atrophy [HST15]](https://www.nice.org.uk/guidance/hst15) used multiple sources of real-world data to characterise spinal muscular atrophy)
* economic burden ([Benralizumab for treating severe eosinophilic asthma [TA565]](https://www.nice.org.uk/guidance/ta565) reported data from CPRD GOLD linked to HES)
* design, populate and validate [economic models](https://www.nice.org.uk/Glossary?letter=E#Economic%20modelling). Common types of evidence include:
* patient starting characteristics ([QAngio XA 3D QFR and CAAS vFFR imaging software for assessing coronary stenosis during invasive coronary angiography [DG43]](https://www.nice.org.uk/guidance/DG43) reported data from the IRIS registry)
* baseline rates of events ([Chronic obstructive pulmonary disease in over 16s: diagnosis and management [NG115]](https://www.nice.org.uk/guidance/ng115) reported data from CPRD GOLD on baseline COPD exacerbation rates by disease severity)
* transition probabilities between health states or disease progression ([Chronic obstructive pulmonary disease in over 16s: diagnosis and management [NG115]](https://www.nice.org.uk/guidance/ng115) used THIN data to model COPD disease progression)
* resource use and costs ([HeartFlow FFRCT for estimating fractional flow reserve from coronary CT angiography [MTG32]](https://www.nice.org.uk/guidance/mtg32) used cost data on coronary revascularisation from NHS Reference Costs)
* patient-reported outcomes, including quality of life ([Elosulfase alfa for treating mucopolysaccharidosis type IVa [HS2]](https://www.nice.org.uk/guidance/hst2) used quality of life data from a survey)
* extrapolation ([Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy [TA520]](https://www.nice.org.uk/guidance/ta520) used Flatiron data)
* patient experience ([Liposuction for chronic lymphoedema [IPG588]](https://www.nice.org.uk/guidance/ipg588) used data from the Living with Lipidaemia survey)
* development and validation of digital health technologies including prognostic models (see the [NICE Evidence Standards Framework for Digital Health Technologies](https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies) for further information)
* identify, characterise and address health inequalities ([Crizanlizumab for preventing sickle cell crises in sickle cell disease [TA743]](https://www.nice.org.uk/guidance/ta743) reported evidence from the National Haemoglobinopathy Registry on the health and disproportionate burden of sickle cell disease in certain minority ethnic groups)
* test accuracy or reproducibility of test results such as biomarkers ([Zio XT for detecting cardiac arrhythmias [MTG52]](https://www.nice.org.uk/guidance/mtg52) reported data from a retrospective observational cohort study)
* device or procedure failure rates ([Joint replacement: hip, knee and shoulder [NG157]](https://www.nice.org.uk/guidance/ng157) used data from the national joint registry on revision rates of knee replacements)
* impact of interventions (including tests) on service delivery and decisions about care ([Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer [DG34]](https://www.nice.org.uk/guidance/dg34) reported results from several prospective observational studies).

Real-world data can also be used to assess the applicability of trial results to patients in the NHS or even to estimate intervention effects (for further information, see [estimating intervention effects using real-world data](#_Estimating_intervention_effects)).

Real-world data, if representative of the target population and of sufficient quality, is the most appropriate source of evidence for most of these use cases. For instance, background event rates or natural history data from trials may over or underestimate event rates in the target population because of selective recruitment ([Bloudek et al. 2021](https://icer.org/wp-content/uploads/2020/10/ICER_Aetion_RWE_HAE_Report_082421.pdf)). Real-world evidence is already widely used in cost-effectiveness and cost-consequence studies informing NICE guidance ([Leahy et al. 2020](https://pubmed.ncbi.nlm.nih.gov/32698805/), [Makady et al. 2018](https://pubmed.ncbi.nlm.nih.gov/29214389/)). However, its use could be more commonplace. In some cases, there may be value in performing studies using routinely collected data rather than relying on published evidence that has lower applicability to the research question.

## Estimating intervention effects using real-world data

### Randomised controlled trials

[Randomised controlled trials](https://www.nice.org.uk/Glossary?letter=R#Randomised%20controlled%20trial) are the preferred study design for estimating the effects of interventions including on safety, effectiveness, resource use and costs. This is because randomisation ensures that any differences in known and unknown baseline characteristics between groups are because of chance. Blinding (if applied) prevents knowledge of treatment allocation from influencing behaviours, and standardised protocols ensure consistent data collection.

However, randomised controlled trials are not always available or may not be sufficient to address the research question of interest.

Randomised trials may not be available for several reasons, including:

* randomisation is considered unethical
* patients are unwilling to be allocated to one of the interventions in the trial
* limited number of eligible patients
* financial or technical constraints on studies
* all treatment combinations (including treatment sequences) cannot be directly assessed.

Randomised controlled trials may be especially difficult to do in rare diseases because of small patient numbers, unmet need or variations in clinical practice. Similarly, high-quality randomised controlled trials can be challenging for medical devices and interventional procedures because of the difficulty of blinding, the importance of learning effects, changes to standard of care making the choice of comparator challenging, changes to the characteristics of the technology over time that may impact on performance, and limited research capacity or access to funding ([Bernard et al. 2014](https://pubmed.ncbi.nlm.nih.gov/25285025/)).

Even if trials are available, they may not be directly applicable to the research question or to routine care in the NHS because of:

* use of comparators that do not represent the standard of care in the NHS (including [placebo](https://www.nice.org.uk/Glossary?letter=P#Placebo) control)
* use of unvalidated surrogate outcomes
* limited follow up
* exclusion of eligible population groups (for example, individuals with co-morbidities)
* differences in populations, care pathways, or settings that impact on the transferability of results to the target population in the NHS
* differences in patient’s use of a technology
* clinical support that differs from routine practice
* learning effects (that is, the effect of an intervention changes over time as users become more experienced)
* methods used to address post-randomisation events such as treatment switching, loss to follow up or missing data.

Some of these challenges, such as the use of comparators that do not represent the standard of care in the NHS, can be addressed through other approaches such as [network meta-analysis](https://www.nice.org.uk/Glossary?letter=N#Network%20meta-analysis) under certain assumptions about the comparability of the trials. See the [NICE Decision Support Unit report on](https://nicedsu.sites.sheffield.ac.uk/methods-development/chte2020-sources-and-synthesis-of-evidence) sources and synthesis of evidence for further information.

Some definitions of real-world evidence include evidence from randomised controlled trials that use real-world data in their design or for measuring outcomes, such as pragmatic clinical trials. Such trials may provide substantial value in combining the internal validity from randomisation with the greater generalisability of data from routine practice. The UK MHRA has published [guidance on producing real-world evidence from randomised controlled trials](https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guideline-on-randomised-controlled-trials-using-real-world-data-to-support-regulatory-decisions).

### Real-world evidence

Because of limited availability of high-quality randomised controlled trials to answer all relevant questions in the development of NICE guidance, and recognition of the importance of learning from the lived experiences of patients in routine settings, there is a renewed interest in using real-world evidence to contextualise or estimate intervention effects. Of course, the absence of relevant and reliable evidence from trials does not automatically mean that real-world evidence offers value. This will depend on the ability of the study to overcome the risks of bias, and the relevance of the data.

#### Contextualisation

Contextualisation involves assessing whether the results from trials will translate well to the target population in the NHS. While this is an important use of real-world data across NICE programmes, NICE does require the collection of further data through managed access arrangements for medicines that are potentially cost-effective and if uncertainties can be addressed through further data collection. This data is often used to understand the relevance of trials to the NHS.

Real-world data has been used in NICE guidance to contextualise clinical trials including for:

* differences in eligible population in the NHS, treatment pathways, care settings and outcomes ([Pegcetacoplan for treating paroxysmal nocturnal haemoglobinuria [TA778]](https://www.nice.org.uk/guidance/ta778) used UK registry data to show that urinary haemoglobin levels in UK practice were in line with the eligibility threshold for the randomised controlled trial)
* modelling the relationship between surrogate outcomes and final [outcomes](https://www.nice.org.uk/Glossary?letter=O#Outcomes) (including patient-reported outcomes)
* measuring the use of, and adherence to, interventions ([Abemaciclib with fulvestrant for treating hormone receptor-positive, HER2-negative advanced breast cancer after endocrine therapy](https://www.nice.org.uk/guidance/ta725) used data from SACT to assess treatment duration)
* assessing the appropriateness of assumptions about long-term outcomes or treatment effects beyond trial periods ([Nintedanib for treating progressive fibrosing interstitial lung diseases [TA747]](https://www.nice.org.uk/guidance/ta747) used registry data to validate extrapolations of long-term outcomes).

For example, [Osimertinib for treating EGFR T790M mutation-positive advanced non-small-cell lung cancer [TA653]](https://www.nice.org.uk/guidance/ta653) used SACT data to assess the relevance of results from the AURA3 trial to NHS patients. In particular, SACT was used to compare:

* overall survival
* differences in patient characteristics including age, ethnicity, performance status and treatment history.

#### Estimation

Effects can be estimated for a range of different outcomes, including:

* Patient outcomes – clinical outcomes, biomarkers, patient-reported outcomes, behaviour change, user satisfaction and engagement
* System outcomes – resource use, costs and processes of care.

The potential uses of real-world data for estimating effects of interventions depend on the stage in their life cycle.

For new interventions (for example, those with recent marketing authorisation in the UK), there will be limited real-world data on their use and outcomes in the NHS. As such, the uses of real-world data are likely restricted to:

* creating a comparator arm (that is, external control) to estimate effects against a single-arm trial or to add to controls from a randomised controlled trial (see [Daratumumab monotherapy for treating relapsed and refractory multiple myeloma](https://www.nice.org.uk/guidance/gid-ta10874/documents/final-appraisal-determination-document?utm_medium=social&utm_source=twitter&utm_campaign=daratumumab150322) used SACT data to form an external control to a single arm phase II trial)
* estimating comparative effects in other countries in which the technology was available earlier than in the UK ([Jonsson et al. 2021](https://www.valueinhealthjournal.com/article/S1098-3015(21)00478-2/fulltext))
* predicting outcomes and treatment effects in routine settings, for example, by reweighting results from trials to reflect characteristics of all eligible patients ([Pembrolizumab with carboplatin and paclitaxel for untreated metastatic squamous non-small-cell lung cancer [TA770]](https://www.nice.org.uk/guidance/ta770) used prescribing data from the Cancer Drugs Fund to estimate outcomes weighted by subgroup prevalence).

Once medical technologies are used routinely or in pilot projects, the opportunities for real-world data are greater and include:

* estimating effects of interventions in routine settings (see [DyeVert Systems for reducing the risk of acute kidney injury in coronary and peripheral angiography [MTG60]](https://www.nice.org.uk/guidance/mtg60))
* providing head-to-head comparisons with preferred comparators ([Douros et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31175610/), [EU-PAS 45073](https://www.encepp.eu/encepp/viewResource.htm?id=45109))
* estimating effects in populations excluded from, or under-represented in, the available randomised controlled trials, or extrapolating results from trials ([Wing et al. 2021](https://www.journalslibrary.nihr.ac.uk/hta/hta25510/#/abstract))
* exploring heterogeneity in intervention effects
* estimating effects on final [outcomes](https://www.nice.org.uk/Glossary?letter=O#Outcomes) of interest (rather than surrogate outcomes) and over longer time periods
* estimating effects for combination therapies (including sequences) or decision strategies not examined in randomised controlled trials ([Fu et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34844936/))
* incorporating into evidence synthesis, for example, informing priors, increasing power or filling evidence gaps in a [network meta-analysis](https://www.nice.org.uk/Glossary?letter=N#Network%20meta-analysis) ([NICE Decision Support Unit 2020](https://nicedsu.sites.sheffield.ac.uk/methods-development/chte2020-sources-and-synthesis-of-evidence), [Sarri et al. 2020](https://pubmed.ncbi.nlm.nih.gov/33298465/)).

### The validity of real-world evidence for estimating intervention effects

A growing body of literature aims to understand the internal [validity](https://www.nice.org.uk/Glossary?letter=V#Validity) of real-world evidence (or, more generally, non-randomised studies) in comparison with randomised controlled trials. This includes meta-epidemiological studies, which compare results from studies of different designs addressing the same question ([Woolacott et al. 2017](https://pubmed.ncbi.nlm.nih.gov/28709997/)), individual case studies ([Dickerman et al. 2020](https://pubmed.ncbi.nlm.nih.gov/32989456/)) and systematic replication studies such as RCT Duplicate ([Franklin et al. 2020](https://pubmed.ncbi.nlm.nih.gov/33327727/)). These studies have enhanced our understanding of the study design and analytical techniques and contexts in which estimates from non-randomised studies are likely to be most valid. Some key learnings from these studies are summarised in [Table 3](#Table3) (please note, these are general learnings and do not apply universally).

## Table 3

Examples of when non-randomised studies are generally more or less likely to produce valid estimates of comparative effects

| Context | Generally lower risk of bias | Generally higher risk of bias |
| --- | --- | --- |
| Study design | Comparative observational cohort studies with concurrent control | Single-arm trials with external control |
| Study design | New users of intervention (including interventions that are established standard of care) | Prevalent users |
| Lifecycle of intervention | Established use of intervention | Newly approved intervention |
| Comparators | Active comparator | Untreated controls |
| Comparators | Concurrent | Historical |
| Comparators | Many alternative treatments for same indication | No other relevant treatment options |
| Comparators | Same type of intervention (such as drug compared with drug) | Different type of intervention (such as drug compared with interventional procedure) |
| Level of clinical equipoise (or uncertainty) | Substantial | Limited |
| Outcomes | Objective clinical endpoints | Self-reported outcomes |
| Outcomes | Unintended effects (such as adverse events) | Intended effects (such as effectiveness) |
| Data quality | High | Low |

## Challenges in generating real-world evidence

Real-world data has great potential for improving our understanding of the value of interventions in routine settings. However, there are important challenges that must be addressed to generate robust results and improve trust in the evidence. We describe key challenges below.

### Trust in real-world evidence studies

Real-world data is often complex and requires substantial preparation before it can be analysed. Also, for some applications, such as the estimation of comparative effects, the methods of analysis can be advanced. Researchers often have access to data before finalising their statistical analysis plans and data preparation and analytical decisions can have important effects on the resulting estimates. Therefore, concerns about the integrity and trustworthiness of the resulting evidence (for example, resulting from data dredging or cherry-picking) need to be addressed. Concerns about the appropriate use of data have been highlighted by the retraction of high-profile studies about the effectiveness of repurposed medicines for treating COVID-19 from prominent medical journals.

Several methods have been proposed to improve trust in the integrity of real-world evidence studies:

* register the study protocol before implementing the study ([RWE Transparency Initiative](https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-transparency-initiative))
* reporting checklists or tools (see [EQUATOR network](https://www.equator-network.org/reporting-guidelines/))
* author statements to confirm the integrity of data access and study conduct ([The Editors of the Lancet Group](https://pubmed.ncbi.nlm.nih.gov/32950071/) 2020)
* open publishing of data, code lists and analytical code.

See guidance on [planning, conducting and reporting real-world evidence studies](#_Conduct_of_quantitative) to generate real-world evidence.

### Data quality and relevance

There are several common challenges with using real-world data. Some types of data are often, though not always, absent from real-world data sources (such as measures of tumour size or functional status). Other variables may be collected at an insufficiently granular level. For instance, a study may need knowledge of a specific drug or medical device, but the data may include only drug or device class. Similarly, a study may need to distinguish between haemorrhagic and ischaemic strokes while a data source may contain data on all strokes without further detail. Even if relevant items are collected with the needed granularity, the data may be missing or inaccurate, which can cause [information bias](#_Information_bias). In addition, there may be variation in data recording practices and quality across centres or individuals and in the quality management processes for different sources of data.

In addition to the availability of data on relevant study elements, the relevance of a given data source to a research question may be affected by several factors. This includes the representativeness of the study sample and similarities in treatment patterns and healthcare delivery to routine care in the NHS, the timeliness of data, sample size and length of follow up. The key questions are whether the data is sufficient to produce robust estimates relevant to the decision problem and whether results are expected to translate or generalise to the target population in the NHS.

See the section on [assessing data suitability](#_Chapter__) for further information.

### Risk of bias

Studies using real-world data are at risk of bias from a number of sources, depending on the use case. We describe key risks of bias that threaten internal validity in individual real-world evidence studies below. Detailed descriptions of risks of bias in non-randomised studies are available, such as [ENCEPP 2021](https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml) and [Sterne et al. 2021](https://training.cochrane.org/handbook/current/chapter-25).

#### Selection bias

In descriptive studies, selection bias occurs if the subjects studied are not representative of the target population. This might result from non-random [sampling](https://www.nice.org.uk/Glossary?letter=S#Sampling) of the source population or from non-response to a questionnaire. In comparative effect studies, selection bias occurs if the selection of participants or follow-up time is related to both the interventions and the outcomes. A lack of representativeness of the target population is not itself a cause of selection bias in comparative studies. Selection bias in comparative studies is distinct from [confounding](#_Confounding).

Common causes of selection bias at study entry include:

* including prevalent users of a technology compared with non-users (users who had already experienced the event or not tolerated the intervention would be excluded from analysis)
* excluding a period of follow up in which the outcome cannot occur (known as immortal time bias for survival outcomes)
* selection into the study based on a characteristic (for example, admission to hospital) that is related to the intervention and outcome.

Common causes of selection bias at study exit are loss to follow up and missing outcome data.

#### Information bias

Information bias may result from missing or inaccurate data on population eligibility criteria, interventions or exposures, outcomes and covariates (as relevant). These limitations may occur because of low data quality, care patterns or data collection processes. They may also result from misspecification of the follow-up period. The consequences of these issues depend on factors including the study type, whether limitations vary across intervention groups, whether they are random or systematic (that is, the missing data mechanism), the magnitude of the limitation and in which variables they occur. One common cause of differential misclassification across groups is detection bias. This occurs when the processes of care differ according to intervention status such that outcomes are more likely to identified in 1 group than in another.

#### Confounding

[Confounding](https://www.nice.org.uk/Glossary?letter=C#Confounding) occurs when there are common causes of the choice of intervention and the outcome. This is expected to be common in healthcare because health professionals and patients make decisions about treatment initiation and continuation based on their expectations of benefits and risks (known as confounding by indication or channelling bias). Confounding bias may be intractable when comparing treatments with different indications and across types of intervention (for example, interventional procedure compared with drug treatment) and for studies of environmental exposures.

Bias may also arise because of inappropriate adjustment for covariates, for example, if a study controls for covariates on the causal pathway (such as blood pressure in the effect of anti-hypertensive medication on stroke), colliders (a variable influenced independently by both the exposure and the outcome), or instruments (defined as a variable that is associated with the exposure but unrelated with the outcome except through the exposure).

#### Other forms of bias

Reverse causation (or protopathic bias) occurs when the intervention is a result of the outcome or a symptom of the outcome. This is most problematic in conditions with long latency periods such as several cancers. If present, this is a severe form of bias with major implications for internal validity.

Biases may also result from the statistical analysis of data (for example, model misspecification).

When assessing the body of literature on a research question there are further concerns about [publication bias](https://www.nice.org.uk/Glossary?letter=P#Publication%20bias) because of non-reporting of real-world evidence studies, especially if they show null results ([Chan et al. 2014](https://pubmed.ncbi.nlm.nih.gov/24411650/)).

# Conduct of quantitative real-world evidence studies

Key messages

|  |
| --- |
| * Transparent and reproducible generation of real-world evidence is essential to improve trust in the evidence and enable reviewers to critically appraise studies * The following principles underpin the conduct of real-world evidence studies: * Ensure data is of good provenance, relevant and of sufficient quality to answer the research question. * Generate evidence in a transparent way and with integrity from study planning through to study conduct and reporting. * Use analytical methods that minimise the risk of bias and characterise uncertainty. * The acceptable level of evidence may depend on the application and various [contextual factors](#_Considerations_for_the). |

## Introduction

### Overview

This section describes NICE’s preferred approaches for planning, conducting and reporting real-world evidence studies. The focus is on primary real-world evidence studies of quantitative data.

The following principles underpin the conduct of real-world evidence studies:

* Ensure data is of good and known provenance, relevant and of sufficient quality to answer the research question.
* Generate evidence in a transparent way and with integrity from study planning through to study conduct and reporting.
* Use analytical methods that minimise the risk of bias and characterise uncertainty.

### Considerations for the quality and acceptability of real-world evidence

All studies should aim for the highest level of transparency and rigour. However, the large number of real-world evidence studies that can inform a single piece of guidance means there may be reasonable trade-offs between the extent of analysis and reporting and the context of use, including:

* the contribution of the study to the final recommendation
* the impact of the recommendation on health and system outcomes
* other contextual factors.

The contribution of a particular type of evidence will vary across applications depending on the key drivers of uncertainty. For instance, in oncology, assumptions around long-term outcomes such as overall survival and the applicability of global trials to the NHS are often key ([Morrell et al. 2018](https://pubmed.ncbi.nlm.nih.gov/29855313/)). In cost-effectiveness or cost-consequence models, a number of different parameters could be important determinants of cost effectiveness including incidence, prevalence, natural history of disease, test performance, costs or quality of life.

In general, studies of clinical effectiveness will need higher levels of rigour and transparency than simple characterisation studies. Estimates of clinical effectiveness are usually a key driver of recommendations and non-randomised studies are typically at high risk of bias.

The contextual factors that influence the acceptability of evidence vary across NICE programmes but can include the level of decision uncertainty, disease prevalence, disease severity, impact on health inequalities and the possibility of generating high-quality evidence.

High-quality real-world evidence may be more difficult to generate in certain contexts such as in rare diseases and for many medical devices and interventional procedures, because of limited availability of high-quality, relevant data and methodological challenges.

Common challenges in the evaluation of medical devices and interventional studies using real-world data include:

* limited integrated national data collections of medical device use and outcomes
* lack of granularity in many routinely collected data sources to identify specific devices (and unique device identifiers)
* identifying appropriate comparators, changes to technologies over time and learning effects

Common challenges in rare diseases include:

* a lack of systematic identification of the target population
* small sample sizes or the need to combine multiple sources of data with different data models and data collection processes
* a lack of agreed common data elements
* substantial variation in natural history of disease.

## Study planning

### Defining the research question

Evidence developers should clearly specify their research question, including:

* conceptual definitions of key study variables including, as relevant, population eligibility criteria, interventions or exposures, outcomes (patient or system outcome) and covariates (including confounders and effect modifiers)
* subgroups
* the target quantity that is to be estimated, for example, disease prevalence or average effect of adhering to an intervention on overall survival.

For non-randomised studies of comparative effects, developers should provide clear justification for the study, considering reasons for the absence of randomised evidence, the limitations of existing trials and the ability to produce robust real-world evidence for the research question.

### Planning study conduct

Developers should plan studies prospectively as far as possible. [Protocols](https://www.nice.org.uk/Glossary?letter=P#Protocol) should fully describe all pre-planned analyses including subgroup and sensitivity analyses. We recognise that the complexity of data curation in many real-world evidence studies means not all analytical decisions can be pre-specified.

Prospective planning improves the quality of studies and reduces the risk of developers performing multiple analyses and selecting those producing the most favourable results.

Pre-specifying analysis plans is especially important for studies of comparative effects (see [analysis plan](#_Analysis_plan)). For such studies, we encourage publishing the study protocol on a publicly accessible platform, with any changes to the protocol registered and justified. We do not recommend a specific platform but options include [ClinicalTrials.gov](https://clinicaltrials.gov/), the [ISRCTN registry](https://www.isrctn.com/), the European Union electronic Register of Post-Authorisation Studies ([EU-PAS](https://www.encepp.eu/encepp/studiesDatabase.jsp)), and [OSF.io](https://osf.io/).

Further guidance on registration of study protocols is provided by the [Real-World Evidence Transparency Initiative](https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-transparency-initiative).

### Choosing fit-for-purpose data

Developers should justify the selection of the final data sources, ensuring the data are of good provenance and fit-for-purpose for the research question (see [assessing data suitability](#_Assessment_of_Data)).

We encourage developers to identify candidate data sources through a systematic, transparent and reproducible search. In the UK this may be supported by registries of data sources such as the Health Data Research UK [Innovation Gateway](https://www.healthdatagateway.org/). Candidate data sources would then be assessed prospectively against defined criteria. This can be informed by the considerations outlined in the section on assessing data suitability or by following external guidance ([Hall et al. 2012](https://pubmed.ncbi.nlm.nih.gov/22069180/), [Gatto et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34716990/)).

Data should be accessed and used in accordance with local law, governance arrangements, codes of practice and requirements of the data owner. In the UK, the Health Research Authority (HRA) provides guidance around research and use of data in accordance with the [UK Policy Framework for Health and Social Care Research](https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/uk-policy-framework-health-and-social-care-research/).

Developers should ensure they have appropriate ethical (or other) approval for the research study if needed. They should also develop a plan for sharing data with independent researchers and NICE collaborating centres, if possible.

### Data collection

For some use cases, primary data collection may be needed. Examples include:

* prospective [observational study](https://www.nice.org.uk/Glossary?letter=O#Observational%20study)
* retrospective chart review
* additional data collection to complement an existing data source, for example, adding a quality-of-life questionnaire to a patient registry or performing a subsample validation study
* health survey.

The need for further data collection should be justified and the burden on patients and healthcare professionals should be proportionate to the value of the additional data.

Data collection should follow a predefined protocol and quality assurance processes should be put in place to ensure the integrity and consistency of data collection. Data collection should follow best-practice standards for Findable, Accessible, Interoperable, and Reusable (FAIR) data using open data standards ([UK Health Data Research Alliance 2021](https://ukhealthdata.org/wp-content/uploads/2021/12/211124-White-Paper-Recommendations-of-Data-Standards-v2-1.pdf)).

Data should be collected, stored, processed and deleted in accordance with the current data protection laws with appropriate transparency information provided and safeguards implemented. Approvals from HRA or local organisation review and agreement as appropriate should be in place.

Please refer to [Health Research Authority guidance](https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/) on governance requirements and data protection regulation for research and non-research use of healthcare data.

## Study conduct

### Choosing study design and analytical methods

The study design and analytical methods used should be relevant to the research question and reflect the characteristics of the data, including:

* the distribution of the outcome variable
* sample size (number of events)
* the structure of the data including data hierarchies or clustering
* heterogeneity
* whether data is cross-sectional or longitudinal.

Diagnostic checks should be used to assess the appropriateness of the selected statistical model, if relevant.

Descriptive studies will be appropriate for many real-world evidence use cases including measuring disease prevalence or incidence, healthcare utilisation or costs, treatment pathways and patient characteristics.

### Minimising risk of bias

Threats to internal validity from sources of bias should be identified and addressed through data collection and analysis as appropriate. Key threats to internal validity come from selection, information, confounding and other biases depending on the use case (see [risk of bias](#_Risk_of_bias)).

The risk of bias from using a particular data source will be informed by the information considered during [data suitability assessment](#_Chapter__).

More detailed guidance on minimising bias in studies of comparative effects is provided in the section on [methods](#_Methods_for_real-world_1).

### Assessing robustness of study results

Developers should seek to minimise bias at the study design and analysis stages. However, because of the range of possible biases and the complexity of real-world data and analytical methods, it is inevitable that some concerns about residual bias will remain.

[Sensitivity analyses](https://www.nice.org.uk/Glossary?letter=S#Sensitivity%20analysis) should reflect areas with the greatest concerns about risk of bias, or when analytical decisions were made despite notable uncertainty. Common considerations include:

* varying operational definitions of key study variables
* differing time windows to define study variables and follow up
* using alternative patient eligibility criteria
* addressing missing data and measurement error
* alternative model specifications
* addressing treatment switching or loss to follow up
* adjusting for non-adherence.

If concerns about residual bias remain high and impact on the ability to make recommendations, developers could consider using quantitative bias analysis. These methods provide quantitative estimates of the impact of bias on study results ([Lash et al. 2014](https://pubmed.ncbi.nlm.nih.gov/25080530/)). If external data on bias is incorporated, this should be identified in a transparent and systematic way. For parameters of [economic models](https://www.nice.org.uk/Glossary?letter=E#Economic%20modelling) including relative effects, sensitivity analysis may consider the impact of bias on cost-effectiveness as well as the parameter value.

### Using proportionate quality assurance processes

Quality assurance of data management, analytical code and analysis is essential to ensure the integrity of the study and reduce the risk of coding errors. Quality assurance processes should be [proportionate](#_Considerations_for_the) to the use.

For general guidance on quality assurance please see the [Office for National Statistic’s Quality Assurance of Code for Analysis and Research](https://best-practice-and-impact.github.io/qa-of-code-guidance/intro.html) and the UK Government’s [Aqua Book](https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government). This may be supported by the use of validated analytical platforms.

## Study reporting

Reporting of studies should be sufficient to enable an independent researcher with access to the data to reproduce the study, interpret the results, and fully understand its strengths and limitations. Several reporting checklists identify key reporting items for:

* observational studies (see the [EQUATOR network](https://www.equator-network.org/reporting-guidelines/) for reporting checklists by study design, and the [STROBE guidelines](https://pubmed.ncbi.nlm.nih.gov/30930717/))
* observational studies of routinely collected data ([RECORD](https://www.record-statement.org/checklist.php)), and
* studies of comparative effects ([RECORD-PE](https://www.record-statement.org/checklist-pe.php); although this tool was initially designed for phamacoepidemiological studies the items are relevant to other comparative studies).

Also, the [START-RWE tool](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6R1KCA) has been developed to help the presentation of study data, methods and results across use cases.

Below we describe key issues across data sources, data curation, methods and results that are especially important to cover in reporting the study.

### Reporting on data sources

Sufficient information should be provided to understand the data source, its provenance, and quality and relevance in relation to the research questions. This should be informed by the considerations described in the section on [data suitability assessment](#_Chapter__).

Developers should provide additional information:

* Ethical (or other) approval for the research study or explain why such approval was not necessary.
* A statement that the data was accessed and used in accordance with approvals and information governance requirements.
* A description of how others can access the data (that is, a [data sharing statement](https://authors.bmj.com/policies/data-sharing/)).

### Reporting on data curation and analysis

Many real-world evidence studies, especially those using routinely collected data, need considerable processing (or curation) before analysis is done. The decisions made in data curation (including linkages, transformations and exclusions) may have substantial effects on study results. Data curation should be well described, such that reviewers can understand what was done and how it may impact on results. The software including the version system and any external packages used to perform analyses should be reported. Ideally, analytical code should follow best practice in code structure, formatting and comments and be publicly available (for example, through a code repository such as GitHub) or made available on request to enable reproduction.

Trust in the integrity of study conduct can be further improved by providing evidence that the study was done appropriately, for example, by showing an audit trail of the analysis. This could demonstrate, for instance, that developers prepared analysis and finalised protocols before the relevant results were revealed ([MacCoun and Perlmutter 2015](https://www.nature.com/articles/526187a)).

### Reporting on methods

Below we describe key items that should be reported. This information should be presented for all analyses including subgroup and sensitivity analyses. Methods should be consistent with the study protocol, and deviations should be identified and justified.

#### Study design

Clear operational definitions should be given for all study variables and details of follow up, if relevant. Study variables include patient eligibility criteria, interventions or exposures, outcomes and covariates.

For each variable, information should be provided on:

* the operational definition of study variables including code lists
* the time period over which information for each variable is sought, defined in relation to an index date (for example, the date of initiating treatment)
* the grace period between observations that are assumed to represent continued use of an intervention, if relevant.

For studies of comparative effects, the process by which potential confounders were identified should be described alongside assumptions about the causal relationships between study variables.

The following information on follow up should be described:

* the start and end of follow up in relation to the index date
* for interventions, assumptions about the minimum time between intervention and outcome occurrence (latency period) and the likely duration of effects (exposure effect window).

In [longitudinal studies](https://www.nice.org.uk/Glossary?letter=L#Longitudinal%20study), this information can be usefully summarised using a study design diagram ([Schneeweiss et al. 2019](https://pubmed.ncbi.nlm.nih.gov/30856654/)). The [REPEAT initiative’s project page](https://www.repeatinitiative.org/projects.html) hosts the paper and design diagram templates.

#### Statistical methods

The statistical methods used should be clearly described. Information should be sufficient to:

* understand what methods were used and why they were chosen
* demonstrate the validity of modelling assumptions
* understand how the analysis addresses different risks of bias including selection bias, information bias and, if relevant, confounding (also see the section on [quality appraisal](#_Quality_appraisal)).

### Reporting results

The following information should be presented in all studies:

* flow (or patient attrition) diagrams to report number of patients at each stage of the study from raw data to the final analytical sample with reasons for exclusion
* patient characteristics (including missing data) and details of follow up including event rates (or other distributional information on outcomes). For comparative studies these should be presented across groups or levels of exposure and, if relevant, before and after adjustment
* differences in patient characteristics in the analytical sample and target population.

Results should include central-point estimates, measures of precision and other relevant distributional information if needed. Results should be presented for the main analysis and all subgroup and sensitivity analyses. It should be clear which of these analyses were pre-specified and which were not. For adjusted analyses, unadjusted results should also be presented.

### Communicating real-world evidence studies clearly

Real-world evidence studies can be technically complex. To help readers understand them, studies should be documented clearly by:

following advice on writing understandable scientific material ([Gopen and Swan 1990](https://www.americanscientist.org/blog/the-long-view/the-science-of-scientific-writing), [Greene 2013](https://press.uchicago.edu/ucp/books/book/chicago/W/bo15288825.html))

explaining jargon and avoiding where possible

[avoiding abbreviations](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5081004/)

labelling tables, graphs, and other non-text content clearly and explaining how to interpret them.

# Assessing data suitability

Key messages

|  |
| --- |
| * Transparent reporting of data sources is essential to ensure trust in the data source and understand its fitness-for-purpose to address the research question * [Data provenance](#_Data_provenance) * Reporting on data sources should cover the characteristics of the data, data collection, coverage and governance * Data fitness-for-purpose can be summarised by the data quality and relevance * [data quality](#_Data_quality) relates to the completeness and accuracy of key study variables * [data relevance](#_Data_relevance) is determined by the data content, differences in patients, interventions and care settings between the data and the target population in the NHS, and characteristics of the data such as sample size and length of follow up * The Data Suitability Assessment Tool (DataSAT) may be used to provide consistent and structured information on data suitability (Appendix 1) * There are reasonable trade-offs between different data sources in terms of quality, size, clinical detail and locality * The acceptability of a given data source may depend on the application and various [contextual factors](#_Considerations_for_the) |

## Introduction

Data used to inform NICE guidance should be reported transparently and be of good provenance and fit-for-purpose in relation to the research question. The primary aims of this section of the framework are to:

* provide clear guidance to evidence developers about expectations for clear and transparent reporting on data and its fitness-for-purpose
* enable evidence reviewers and committees to understand data trustworthiness and suitability when critically appraising the study or developing recommendations.

We do not define minimum standards for data suitability beyond that the data should be used in accordance with national laws and regulations concerning data protection and information governance (see the section on [reporting on data sources](#_Reporting_on_data)). The considerations for data suitability are broadly applicable across different types of real-world data and use cases. The acceptability of a data source will depend on the use case and [contextual factors](#_Considerations_for_the). We recognise the need for trade-offs between different characteristics of data sources including quality, size, clinical detail and locality. For some applications international data will be appropriate.

We do not request a particular format for the overall presentation of this information. However, we have developed the Data Suitability Assessment Tool (DataSAT) to help the consistent and structured presentation of data suitability at the point of assessment. The concepts presented in the tool may also help developers choose between potential data sources and in performing feasibility studies, but this is not its primary purpose. The tool template and example applications are presented in appendix 1.

## Data provenance

A full understanding of data provenance is essential to create trust in the use of data and understand its fitness-for-purpose for a given application. In this section we present data provenance considerations across 4 themes: basic characteristics of the data source, data collection, coverage and governance.

Many real-world evidence studies will combine more than 1 data source, either by data linkage or data pooling. Data linkage is often done to extend the information available on individual patients, for example, by combining data from a prospective observational cohort study with hospital discharge or mortality records. Data pooling is used to extend sample size or coverage of data and is common in studies of rare diseases.

The reporting of data sources should primarily refer to the combined data used for the research study. However, important differences between contributing datasets should be clearly described.

### Basic characteristics of data

Information that allows identification of the data sources should be clearly reported. This includes the names of the overall and contributing data sources, versions (if available) and the dates of data extraction.

Common data models are used to standardise the structure and sometimes coding systems of different data sources. If data has been converted to a common data model, the model and its version should be reported and full details of the mapping made available. This information is essential to allow the study to be reproduced.

While complete and accurate data linkage will improve the quality and value of data, imperfect linkage could exclude patient records or lead to data misclassification. Therefore, when multiple sources of data are linked the following information should be reported:

* who did the linkage (for example, NHS Digital)
* methods of linkage including whether deterministic or probabilistic, and the variables used for linkage
* the performance characteristics of data linkage ([Government Analysis Function 2021](https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/quality-assessment-in-data-linkage)).

### Data collection

An understanding of a data source requires knowledge of the purpose and methods of data collection.

Information on the original purpose of data collection should include:

* whether the data was routinely collected or collected for a specific research purpose (or a combination)
* the type of data source and primary use, for example:
* electronic health records for patient care
* administrative data for reimbursement of providers
* registry for assessing medical device safety
* prospective observational cohort study to estimate quality of life following an intervention
* retrospective chart review to model the natural history of a condition.

Additional information on important data types should cover:

* which types of data were collected, for example clinical diagnoses, tests, procedures and prescriptions
* how these were coded or recorded, for example, using ICD-10 codes for clinical diagnoses, or free text data on cancer stage or biomarkers
* how data was collected, for example, directly by healthcare professionals in clinical examinations, by remote monitoring or by administrative staff. If data is captured by a digital health technology, the validity of the technology should be reported
* changes to data collection over time, for example
* addition of new data elements (for example, a quality-of-life questionnaire)
* removal of data elements
* changes to the method of data collection (for instance, a switch to routine monitoring of patient outcomes)
* changes to coding systems (for example, the switch from Read v2 to SNOMED-CT codes in UK primary care). Information on any mapping between coding systems should be made available
* software updates to data capture systems including digital health technologies that had substantial impacts on data capture
* quality assurance processes for data collection that were in place (including training or blinded review)

transformations performed on the data such as conversion to a common data model or other data standards.

Any differences between data providers in how data were collected should be described. This is especially important when data sources are pooled from different systems.

### Data coverage

Providing clear information on data coverage is essential, including the population, care settings, geography and time. Such information has important implications for [data relevance](#_Data_relevance) that can inform later assessments of data suitability.

Information should be provided on:

* The extent to which the data source captures the target population.
* If a data source does not include the full target population, the representativeness of the data captured should be noted.
* For studies involving prospective data collection including patient registries, information on patient accrual should be reported.
* The care settings in which data collection was based.
* This should distinguish between care settings, if relevant (for example, primary care compared with secondary care) and types of providers (for example, specialist medical centres compared with general hospitals) .
* If information was collected outside of the health or social care system, this should be described (for instance, remote monitoring of activities of daily living).
* The geographical coverage of the data including countries and regions, if relevant.
* The time period of data collection.

### Data governance

Information about data governance is important for understanding the maturity of data and its reliability. This should include the following information:

* the name of the data controller
* the funding source for data collection and maintenance
* data documentation including items such as a data dictionary and data model
* details of the quality assurance and data management process including audit.

## Data fitness-for-purpose

[Data provenance](#_Data_provenance) described important characteristics of data sources distinct from the planned study. In this section we focus on the fitness-for-purpose of data to answer specific research questions considering its quality and relevance. A dataset may be of value for 1 application but not another.

Substantial data curation including data cleaning, exclusions and transformations is needed to prepare original data sources for analysis. Data curation and quality assurance should be reported transparently as described in [study reporting](#_Reporting_on_data_1).

## Data quality

Limitations to data quality include missing data, measurement error, misclassification and incorrect reporting of dates. These issues can apply to all study variables including patient eligibility criteria, outcomes, interventions or exposures, and covariates. They can create information biases that cause real-world evidence studies to produce biased estimates. Transparent reporting of data quality is essential for reviewers to understand the risk of bias and whether it has been adequately addressed through data analysis or explored through sensitivity analysis. We focus on 2 main aspects of data quality: completeness and accuracy.

Information on completeness and accuracy should be provided for all key study variables. Study variables can be constructed by combining multiple data elements including both structured and unstructured data and may come from different linked data sources. The complexity of these study variables will vary according to the data sources and applications. For instance, in some applications an asthma exacerbation may be identified from a single data field (such as questionnaire), while in others it may need to be constructed from combinations of diagnostic codes, prescriptions, tests, free text or other data.

As described in the section on [study reporting](#_Reporting_on_methods), it is essential that clear and unambiguous definitions are given for each study variable including types of data, code lists, extraction from unstructured data, and time periods. These cohort definitions including ode lists should be made available to others and reused, if appropriate. The validity of an existing code list should be reviewed before use.

These considerations also apply to data from digital health technologies producing patient-generated data, including patient-reported outcomes and digital biomarkers. Further information on the validity of data generated from the technology and user accessibility should be provided.

To interpret study results, further information is needed on reasons for data missingness or inaccuracy and whether these are random or systematic. For comparative studies, it is important to understand the extent to which missingness or inaccuracy differ across intervention groups. The section on [addressing information bias](#_Addressing_information_bias) has further information on methods for dealing with missing data, measurement error and misclassification.

### Completeness

Data completeness refers to the percentage of records without missing data at a given time point. It does not provide information on the accuracy of that data. The percentage is easily calculated from the data source and should be calculated before excluding relevant data or imputation. Data completeness is not considered relevant for outcomes such as experiencing a myocardial infarction. The absence of an event (when it has occurred) is considered a data accuracy issue (misclassification) rather than missing data.

### Accuracy

Measuring accuracy, or how closely the data resemble reality, depends on the type of variable. Below we describe common metrics of accuracy for different types of variables:

* continuous or count variables (mean error, mean absolute error, mean squared error)
* categorical variables (diagnostic accuracy measures such as sensitivity, specificity, positive predictive value, and negative predictive value) ([Fox et al. 2022](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7750925/))
* time-to-event variables (difference between actual time of event and recorded time of event).

Gold standard approaches for measuring accuracy of the data include:

* comparison with an established gold standard source (for example UK Office for National Statistics mortality records)
* medical record review.

These approaches may be taken for a subset of the analytical population or be based on a previous study in the same or similar population and data source.

If gold standard approaches are not possible, other approaches that can show approximate accuracy include:

* comparing different variable definitions, for example, by using additional codes, requiring multiple codes, or combining different data types
* comparing sample distributions with population distributions or previous studies
* exploring plausibility of the data, informed by expert opinion
* checking consistency (agreement in patient status in records across the data sources)
* assessing conformance (whether the recording of data elements is consistent with the data source specifications)
* checking persistence (whether the data are consistent over time).

Transparent reporting of data accuracy for key study variables includes:

* quantitative information on accuracy, if available, including means and confidence intervals. Additional distributional information may also be valuable.
* describing the methods and processes used to quantify accuracy including any assumptions made. When this is based on previous studies, the applicability to the present analysis should be discussed, and may consider differences in study variable definitions, populations, data sources, time periods or other relevant considerations.

## Data relevance

The second component of data fitness-for-purpose is data relevancy. Key questions of data relevancy are whether:

* the data provides sufficient information to produce robust and relevant results
* the results are likely to generalise to patients in the NHS.

The assessment of data relevancy should be informed by the information provided on [data provenance](#_Data_provenance).

NICE prefers data relating directly to the UK population that reflects current care in the NHS. However, we recognise the potential value of international data if limited information is available for the NHS. In some applications there will be a trade-off between using local data and other important characteristics of data including quality, recency, clinical detail, sample size and follow up. International data is likely to be of particular value when an intervention has been available in another country before becoming available in the UK, or in the context of rare diseases. Similar considerations apply to using data from regional or specialist healthcare providers within the NHS.

We describe key aspects of data relevancy below distinguishing between data content, coverage and characteristics.

### Data content

There are 3 key considerations for understanding whether the data content is sufficient for a research question:

* Does the data source contain sufficient data elements to enable appropriate definitions of population eligibility criteria, outcomes, interventions and covariates, as relevant?
* Are the data elements collected with sufficient granularity (or detail)?
* Are measurements taken at relevant time points?

To help understand whether data elements are sufficient, it is useful to first define the target concept and judge the extent to which this can be proxied using real-world data. The implications of insufficient data will vary depending on the study variable and use case. Key endpoints necessary to answer the research questions should be available and should be sufficiently objective and detailed to support an evaluation. Insufficient information to define the population, interventions or outcomes appropriately will limit the relevance of the research findings. Insufficient information on confounders will limit the ability to produce valid findings.

The needed granularity of data will vary across research questions. For example, when considering the effect of knee replacement on quality of life we may be interested in the effect compared with physiotherapy alone, total versus partial knee replacement, or of different implanted devices. Similarly, any stroke may be appropriate as an outcome for some research questions, while others will need haemorrhagic and ischaemic strokes to be separated.

Finally, we may be interested in the effect of knee replacement on quality of life at 1 year after the procedure. In routinely collected data, the recording of such information does not follow a strict protocol and measurements, if present, may be taken at other time points.

### Data coverage

The generalisability of research findings to patients in the NHS will depend on several factors, including:

* the similarity in patient characteristics between the analytical sample and target population
* the similarity in care pathways and treatment settings
* changes in care pathways (including diagnostic tests) and outcomes over time.

The similarity of the analytical sample to the target population is especially important in descriptive studies, such as those estimating disease prevalence. In comparative studies this may be less important if the intervention effects are expected to transfer across patients with different characteristics and the emphasis should be on ensuring internal validity. If there is substantial heterogeneity in treatment effects across subgroups, similarity in patient characteristics becomes more important. Effect estimates on the relative scale usually transfer better across subgroups than estimates on absolute scales ([Roberts and Prieto-Merino 2014](https://pubmed.ncbi.nlm.nih.gov/25705414/)). In other applications, such as prognostic modelling, non-representative sampling may be preferred to ensure adequate representation of important patient subgroups.

Consideration needs to be given to how any differences in the treatment pathways or care settings seen in the analytical sample and the NHS may impact on the relevance of results. This is especially important when using international data. Even within the NHS, the data may relate to specific regions that are not representative of the country or focus on specialist providers rather than all providers. Finally, changes to care pathways including diagnostic tests as well as background trends in outcomes (such as mortality) may limit the value of historical data even from the NHS. These issues need to be carefully considered and reported when discussing the relevance of data for use in NICE guidance.

### Data characteristics

The final category of data relevancy concerns the size of the analytical sample and the length (and distribution) of follow up. The sample size should be large enough to produce robust estimates. The follow up should be long enough for the outcomes of interest to have occurred or accrued (for outcomes such as healthcare costs). The amount of data available before the start of follow up may also be important to provide information on confounders and identify new users of an intervention. Using more current data (that is, data with a lower time lag between data collection and availability for research) may extend follow up.

# Methods for real-world studies of comparative effects

Key messages

|  |
| --- |
| * [Study design](#_Study_design) * Design non-randomised studies to emulate the preferred randomised controlled trial ([target trial approach](#_The_target_trial)) * Avoid time-related biases because of deviations between time zero, eligibility and start of intervention * For studies using external control, minimise differences between data sources including availability and operational definitions of key study variables, data collection processes, patient characteristics, treatment settings, care pathways, and time periods, and consider the implications for study quality and relevance * [Analysis](#_Analysis) * Identify potential confounders using a systematic approach and clearly articulate causal assumptions (including time-varying confounders) * Use a statistical method that addresses confounding considering observed and unobserved confounders * Consider the impact of bias from informative censoring, missing data, and measurement error and address appropriately if needed * Use sensitivity and bias analysis to assess the robustness of results to main risks of bias and uncertain data curation and analysis decisions * [Reporting](#_Reporting) * Justify the need for non-randomised evidence * Provide a study protocol and statistical analysis plan before performing final analyses * Report studies in sufficient detail to enable independent researchers to reproduce the study and understand what was done and why * Assess the risk of bias and relevance of the study to the research question * The acceptable quality of evidence may depend on various [contextual factors](#_Considerations_for_the) |

## Introduction

We previously outlined principles for the robust and transparent [conduct of quantitative real-world evidence studies](#_Use__) across different use cases. In this section we provide more detailed recommendations for the conduct of studies of comparative effects using real-world data. This includes traditional observational studies based on primary or secondary data collection and trials in which real-world data is used to form an external control. We do not provide specific considerations for purely interventional studies (whether randomised or not) or external control studies using only interventional data.

Randomised controlled trials are the preferred study design for estimating comparative effects. Non-randomised evidence may add value if randomised controlled trials are absent, not directly relevant to the research question or of poor quality (see [challenges in randomised controlled trials](#_Use_and_challenges)).

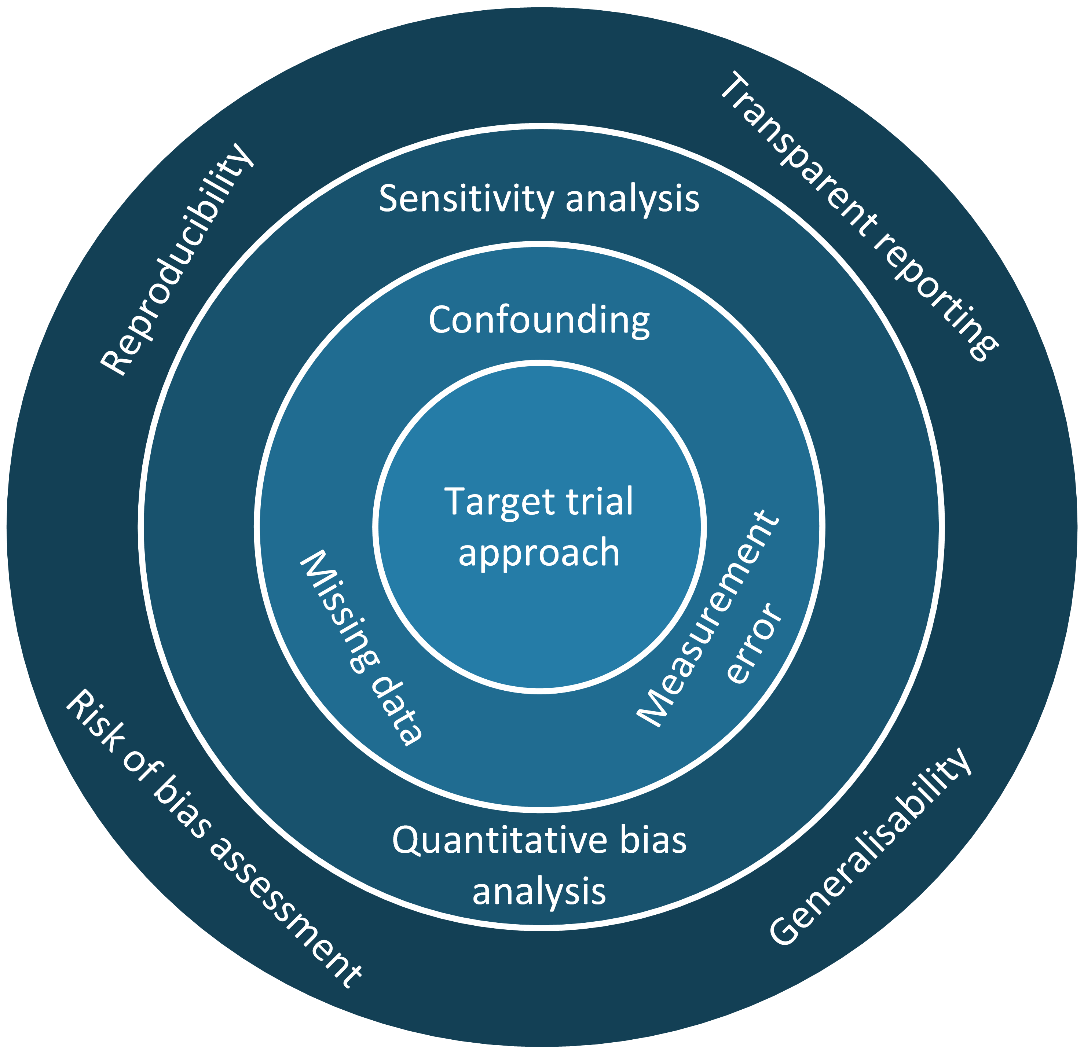
If real-world evidence on comparative effects may improve the evidence base, it is essential that studies are done using robust and transparent methods. We recommend designing real-world evidence studies to emulate the randomised trial that would ideally have been done (see [study design](#_Study_design)), using appropriate statistical methods to address confounding and informational biases (see [analysis](#_Analysis)), and assessing the robustness of results using sensitivity and bias analysis (see [assessing robustness](#_Assessing_robustness_of)). The preferred approach is summarised in [figure 1](#Figure2).

The recommendations provided here are intended to improve the quality of real-world studies of comparative effects, both in terms of methodological quality and validity, and the transparency of study conduct. They were derived from best-practice guidance from the published literature, international research consortia, and international regulatory and payer bodies, and will be updated regularly in line with developing methodologies. They build on the [NICE Decision Support Unit Technical Support Document (TSD) 17](https://nicedsu.sites.sheffield.ac.uk/tsds/observational-data-tsd), which presents statistical methods for analysing observational data.

We recognise that not all studies will be able to meet all recommendations in full. The acceptability and contribution of specific studies to decisions will depend on the NICE programme as well as several contextual factors as described in [conduct of quantitative real-world evidence studies](#_Use__).

## Figure 1

Visual summary of key considerations for planning and reporting a study of comparative effects



## Types of non-randomised study design

### Overview

A large variety of study designs can be used to estimate the effects of interventions, exposures or policies. The preferred study design will be context dependent. It may depend on whether variation in the exposure is within individuals over time, between individuals, or between other groups such as healthcare providers. In general, confidence in non-randomised study results is strengthened if results are replicated using different study designs, known as triangulation ([Lawlor et al. 2016](https://pubmed.ncbi.nlm.nih.gov/28108528/)).

One important distinction is between interventional and observational studies. In interventional studies, individuals (or groups of individuals) are allocated to one or more interventions according to a protocol. Allocation to interventions can be random, quasi-random or non-random. In observational studies, interventions are not determined by a protocol but instead according to the preferences of health and social care professionals and patients. Hybrid studies may make use of both interventional and observational data. In this section we focus on observational and hybrid studies only.

Both interventional and observational studies can be uncontrolled. Uncontrolled studies are appropriate only in rare cases, in which the natural course of the disease is well understood and highly predictable and the treatment effect is very large ([ICH E10](https://www.ema.europa.eu/en/ich-e10-choice-control-group-clinical-trials), [Deeks et al. 2003](https://pubmed.ncbi.nlm.nih.gov/14499048/)). In most cases a comparison group is needed to generate reliable and informative estimates of treatment effects. Controlled studies can make use of variation in exposures and outcomes across individuals (or groups), within individuals (or groups) over time, or both. In this section we focus on controlled studies.

Below we discuss types of comparative studies. Some taxonomies distinguish between prospective studies (involving primary data collection) and retrospective studies (based on already collected data). This distinction does not necessarily convey information about study quality and so we advise against its use ([Dekkers and Groenewold 2020](https://pubmed.ncbi.nlm.nih.gov/33055302/)).

### Cohort studies

In controlled cohort studies, we identify individuals based on their exposures and compare outcomes during follow up. Usually, cohort studies will compare individuals subject to different exposures within the same data source. However, they can also combine data from different sources including from interventional and observational data sources. In this case, the observational data is used to form an external control to the intervention used in the trial. The trial will often be an uncontrolled single-arm trial but could also be an arm from a controlled trial. External data can also be used to augment concurrent controls within a trial.

External controls can also be formed from data from previous clinical trials. A potential advantage of such studies is greater similarity in patient inclusion criteria, follow up and outcome determination. Often only aggregate rather than individual patient-level data will be available from previous trials. [NICE's technical support document 18](https://nicedsu.sites.sheffield.ac.uk/tsds/population-adjusted-indirect-comparisons-maic-and-stc) describes methods for unanchored indirect comparisons with aggregated data. In this section, we restrict attention to external control studies using individual patient-level data from observational data sources.

### Self-controlled studies

Self-controlled designs make use of variation in exposure status within individuals over time. They are most appropriate for transient exposures with acute-onset events ([Hallas and Pottegard 2014](https://pubmed.ncbi.nlm.nih.gov/24635348/)). While primarily used in studies of adverse effects of medicines (including vaccines), they have been used to assess the effects of oncology medicines using the experiences of individuals on prior lines of therapy ([Hatswell and Sullivan 2020](https://pubmed.ncbi.nlm.nih.gov/30698076/)). This is most relevant if appropriate standard-of-care comparators are not available.

A key advantage of self-controlled methods is the ability to control for confounders (including unmeasured or unknown confounders) that do not vary over time, such as genetic inheritance, or vary slowly like many health behaviours. However, it is still necessary to adjust for covariates that may change over time (for example, disease severity). Such methods generally either assume no time-based trends in outcomes or try to model the trend statistically. These approaches can often be strengthened by the addition of control groups of people not exposed to the interventions.

### Cross-sectional studies

In [cross-sectional studies](https://www.nice.org.uk/Glossary?letter=C#Cross-sectional%20study) information on current exposures and outcomes is collected at a single time point. While they can be used to estimate intervention effects, they are less reliable than longitudinal studies (such as cohort studies) if there is a clear temporal separation of exposures and outcomes.

### Case-control studies

In [case-control studies](https://www.nice.org.uk/Glossary?letter=C#Case-control%20studies) individuals are selected based on outcomes, and rates of exposures are compared. Case-control studies embedded within an underlying cohort are known as nested case-cohort studies. Case-control studies conducted within existing database studies are generally not recommended because they use less information than cohort studies ([Schuemie et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31436848/)). Case-control studies are most useful for rare outcomes or if there is a need to collect further information on exposures.

### Quasi-experimental studies

[Quasi-experimental studies](https://www.nice.org.uk/Glossary?letter=Q#Quasi-experimental%20study) and natural experiments exploit external variation that impacts treatment assignment. In quasi-experimental studies, the assignment can be proxied as random conditional on an instrument (a variable strongly correlated with the exposure of interest but not the outcome, except through the exposure). This provides some unmeasured confounding control. Common designs include instrumental variable regression, regression discontinuity and interrupted time series ([Reeves et al. 2017](https://pubmed.ncbi.nlm.nih.gov/28351692/)). These studies are only appropriate if a strong instrument is available. They are most widely used in public health settings in which randomisation may not be feasible.

## Study design

### The target trial approach

Non-randomised studies should be designed to mimic the randomised trial that would ideally have been done ([Hernan and Robins 2016](https://pubmed.ncbi.nlm.nih.gov/26994063/), [Gomes et al. 2022](https://pubmed.ncbi.nlm.nih.gov/35332434/)). This process, known as the target trial approach (or trial emulation), requires developers to clearly articulate the study design. It helps avoid selection bias because of poor design ([Bykov et al. 2022](https://pubmed.ncbi.nlm.nih.gov/34260087/)) and improves transparency for reviewers.

Studies should aim to replicate the target trial as closely as possible and, if this is not possible, trade-offs should be clearly described. In some cases, a data source may not be of sufficient relevance or quality to allow trial replication. This can be particularly problematic for studies using a real-world data external control if differences between data sources in terms of patients, settings, care, data collection and time period limit the comparability of data sources ([Gray et al. 2020](https://pubmed.ncbi.nlm.nih.gov/32440847/), [Pocock 1976](https://pubmed.ncbi.nlm.nih.gov/770493/)). Sometimes it will not be possible to adequately replicate a target trial with real-world data and bespoke data collection may be needed.

The target trial approach is applicable to many non-randomised study designs including traditional observational cohort studies, single-arm trials with external control, case-control and quasi-experimental studies. It is used within the Cochrane ROBINS-I risk of bias tool for non-randomised studies ([Sterne et al. 2016](https://pubmed.ncbi.nlm.nih.gov/27733354/)).

The target trial can be defined across 7 dimensions: eligibility criteria, treatment strategies, assignment procedure, follow-up period, outcomes, causal effect of interest and analysis plan. We describe each dimension below and provide considerations for those developing evidence to inform NICE guidance.

#### Eligibility criteria

For most studies, the eligibility criteria should mimic a hypothetical pragmatic trial and reflect the clinical pathways (including diagnostic tests) and patients seen in routine care in the NHS. For external control studies, the focus should be on matching the eligibility criteria from the interventional study rather than the broader target population. As in a trial, eligibility criteria should be based on variables recorded before starting the intervention.

If heterogeneity is anticipated in the intervention effects, [subgroup analysis](https://www.nice.org.uk/Glossary?letter=S#Subgroup%20analysis) can be done. The subgroups should be defined upfront when planning the study.

#### Treatment strategies

Treatment strategies include the intervention of interest and any comparators. Comparators could be different levels of an exposure (for example, different doses of a medicine), a different intervention, or the absence of intervention. In observational data it is very difficult to emulate a placebo-controlled trial because of higher risk of selection bias and intractable confounding.

Active comparators for the same treatment indication are preferred. Active comparators reduce the risk of confounding by indication by ensuring greater similarity of patients having different interventions. If routine follow-up procedures are similar across interventions this also reduces the risk of detection bias. The active comparator should ideally reflect established practice in the NHS.

For studies of interventions, new (or incident) user designs are generally preferred to studies of prevalent users because of the lower risk of selection bias and better replication of trial designs. Prevalent users have, by definition, remained on-treatment and survived for some period of follow up. In secondary database studies, new users are typically defined using a washout period in which the individual was not observed to use the intervention of interest. A further advantage of new-user designs is the ability to estimate time-varying hazards from treatment initiation.

The inclusion of prevalent users may be needed if the effects of interventions are cumulative, there are too few incident users in the data, or follow up is limited ([Vandenbroucke and Pearce 2015](https://pubmed.ncbi.nlm.nih.gov/26507305/), [Suissa et al. 2016](https://pubmed.ncbi.nlm.nih.gov/27610604/)).

Data on comparators would ideally come from the same period as the intervention as well as from the same healthcare system and settings. This is to minimise any differences between treatment groups resulting from differences in care access, pathways (including diagnostic tests) or time-based trends in outcomes.

#### Assignment procedure

In randomised controlled trials, individuals (or groups) are randomly assigned to interventions. If possible, providers, patients and analysts are blinded to this assignment. Neither random assignment nor blinding are possible in observational studies. With sufficient information on confounders, random assignment can, however, be approximated through various analytical approaches (see [analysis](#_Analysis)).

In some applications, individuals will meet eligibility criteria at multiple time points. For example, they may initiate treatment more than once after a sufficient washout period. There are several approaches to deal with this including using only the first eligible time point, a random eligible time or all eligible time points ([Hernan and Robins 2016](https://pubmed.ncbi.nlm.nih.gov/26994063/)).

#### Follow-up period

The start and end of follow up must be defined. The start of follow up should ideally begin at the same time at which all eligibility criteria are met and the intervention is assigned (or just after). If a substantial latency period is expected between treatment initiation and outcomes, it may be necessary to define an induction period before which outcomes are not counted. This can reduce the risk of reverse causation, in which the outcome influences the exposure.

The follow-up period should be long enough to capture the outcomes of interest but should not exceed the period beyond which outcomes could be reasonably impacted by the intervention (known as the exposure-effect window). Censoring events should be clearly defined and will depend on the causal effect of interest.

#### Outcomes

Primary and secondary outcomes should be defined and can include both patient and health system outcomes (such as resource use or costs). Patient outcomes should reflect how a patient feels, functions, or how long a patient lives. Objective clinical outcomes (such as survival) are typically subject to a lower risk of bias than subjective outcomes if outcome detection or reporting could be influenced by known treatment history.

For a surrogate outcome there should be good evidence that changes in the surrogate outcome are causally associated with changes in the final patient outcomes of interest ([Ciani et al. 2017](https://pubmed.ncbi.nlm.nih.gov/28292495/)).

While outcome ascertainment is not blinded in observational data, analysts can be blinded to outcomes before finalising the analysis plan (see [analysis](#_Analysis)).

#### Causal effect of interest

Researchers should describe the causal effect of interest. This will depend on the [estimand](#_Analysis_plan).

There are 2 main causal effects estimated in trials: the effect of assignment to an intervention ([intention-to-treat](https://www.nice.org.uk/Glossary?letter=I#Intention-to-treat%20analysis)) and the effect of adhering to treatment protocols ([per-protocol](https://www.nice.org.uk/Glossary?letter=P#Per-protocol%20analysis)). It is not usually possible to estimate the effect of treatment assignment using observational data because this is not typically recorded. However, it can be proxied using treatment initiation (the as-started effect). The equivalent of the per-protocol effect is sometimes called the on-treatment effect.

The as-started effect is usually of primary interest to NICE. However, if treatment discontinuation (or switching) is substantial or is not expected to reflect routine practice or outcomes in the NHS, it is important to present results from the on-treatment analysis. On-treatment analyses may also be most appropriate for the analysis of safety and adverse events. Some research questions involve dynamic treatment strategies including treatment sequences.

#### Analysis plan

The analysis plan should describe how the causal effect of interest is to be estimated, taking into account post-randomisation events. These are events (such as treatment switching or non-adherence) occurring after treatment initiation that affect the interpretation of the outcome of interest. This is supported using the estimand framework ([ICH E9 [R1]](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf)).

The relevance of post-randomisation events will depend on the causal effect of interest. In an as-started analysis, treatment discontinuation, switching or augmentation are usually ignored. However, if these changes are substantial there is a risk of increasing exposure misclassification over time. If this is similar between treatment groups, it would tend to bias estimates of effect towards the null.

In an on-treatment analysis or when modelling dynamic treatment strategies the follow up is often censored once the patient stops adhering to the treatment plan plus some biologically informed effect window. For medicines (and some devices) continued exposure is proxied by dates of prescriptions and number of days’ supply, with some grace period between observations permitted. Particular attention needs to be given to the possibility of informative censoring, which causes bias if censoring depends on outcomes and differs across interventions, and time-varying confounding.

Further content on statistical analysis including addressing confounding, informative censoring, missing data and measurement error is presented in the [analysis section](#_Analysis).

Panel 1 shows examples of using the target trial approach:

#### Panel 1

Examples of the target trial approach

|  |
| --- |
| Example 1: What is the effect of initiating HRT on coronary heart disease in postmenopausal women?  The Women’s Health Initiative randomised controlled trial showed that initiating treatment with hormone replacement therapy increased the risk of coronary heart disease in postmenopausal women. This contradicted earlier observational studies that found a reduction in the risk of coronary heart disease. [Hernan et al. 2008](https://pubmed.ncbi.nlm.nih.gov/18854702/) followed a target trial approach, replicating as far as possible the Women’s Health Initiative trial using data from the Nurses’ Health Study. They were able to show that the difference in results between the trial and observational studies resulted from the inclusion of prevalent users of hormone replacement therapy in the observational cohort. These women had already survived a period of time on-treatment without experiencing the outcome. Following a new-user design (as well as other principles of the target trial approach) they were able to produce effect estimates consistent with the trial.  Example 2: What is the optimal estimated glomerular filtration rate (eGFR) at which to initiate dialysis treatment in people with advanced chronic kidney disease?  The IDEAL randomised controlled trial showed a modest reduction in mortality and cardiovascular events for early versus late initiation of dialysis. The average eGFR scores in the early and late treatment arms were 9.0 and 7.2 mL/min/1.73 m2, respectively. There therefore remains considerable uncertainty about the optimal time to initiate dialysis. [Fu et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34844936/) emulated the IDEAL trial using data from the National Swedish Renal Registry and were ability to produce similar results over the narrow eGFR separation achieved in the trial. They were then able to extend the analysis to a wider range of eGFR values to identify the optimal point at which to initiate dialysis therapy.  Example 3: What is the effect of initiating treatment with fluticasone propionate plus salmeterol (FP-SAL) versus 1) no FP-SAL or 2) salmeterol only on COPD exacerbations in people with COPD?  The TORCH trial found that treatment with FP-SAL was associated with a reduction in the risk of COPD exacerbations compared with no FP-SAL or salmeterol only. However, the trial excluded adults aged above 80 years and those with asthma or mild COPD. There is uncertainty about the extent to which results from the TORCH trial apply to these patients. [Wing et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34463610/) were able to replicate the findings of the TORCH trial for COPD exacerbations using primary care data from Clinical Practice Research Datalink in England for the comparison with salmeterol only but not with no FP-SAL. This reflects the challenge in emulating a trial with placebo control. By extending their analysis to a wider target population they were able to demonstrate evidence of treatment effect heterogeneity by COPD severity but not by age or asthma diagnosis. |

## Analysis

### Addressing risk of confounding bias

#### Selection of confounders

Potential confounders should be identified prospectively (that is, before analysis), based on a transparent, systematic and reproducible process. Key sources of evidence are published literature and expert opinion. Consideration should be given to the presence of time-varying confounders. These affect the outcome and future levels of the exposure, and can be affected by previous levels of the exposure. They are especially relevant when modelling time-varying interventions or dynamic treatment strategies or addressing informative censoring.

Developers should outline their assumptions about the causal relationships between interventions, covariates and outcomes of interest. Ideally, this would be done using causal diagrams known as directed acyclic graphs ([Shier and Platt 2008](https://pubmed.ncbi.nlm.nih.gov/18973665/)).

Inappropriate adjustment for covariates should be avoided. This may result from controlling for variables on the causal pathway between exposure and outcomes (overadjustment), colliders or instruments. Confounders that may change value over time should be recorded before the index date, except when using statistical methods that appropriately address time-varying confounding.

The selection of covariates may use advanced computational approaches such as machine learning to identify a sufficient set of covariates ([Ali et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31619986/)). The use of these methods should be clearly justified and their consistency with causal assumptions examined. Choosing covariates based on statistical significance should be avoided.

#### Selecting methods for addressing confounding

Adjusted comparisons based on clear causal assumptions are preferred to naive (or unadjusted) comparisons. Statistical approaches should be used to address confounding and approximate randomisation (see [treatment assignment](#_Assignment_procedure)).

Methods to address baseline confounding can be divided into 2 broad categories ([Matthay et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31890846/)):

* Confounding-control methods – these control for observed confounders and include various approaches such as stratification, matching, multivariable regression and propensity score methods, or combinations of these.
* Instrument-based (or quasi-experimental) methods – these try to address measured and unmeasured (or poorly measured) confounders based on an instrument that makes use of exogeneous variation in intervention status. Common methods include instrumental variable regression, regression discontinuity and interrupted time series methods.

Confounding-control methods assume no unmeasured confounding. They include simple adjustment methods such as stratification, restriction and matching. These may be appropriate for research questions in which confounding is well understood and there are only a small number of confounders that are well recorded.

If there are many potential confounders, more complex methods such as multivariable regression and propensity score (or disease risk score) methods are preferred. Propensity scores give the probability of receiving an intervention based on observed covariates. Several methods use propensity scores including matching, stratification, weighting and regression (or combinations of these). General discussions of the strengths and weaknesses of these different approaches can be found in [Ali et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31619986/). The choice of method should be justified.

There is mixed evidence on the relative performance of regression and propensity score methods for addressing confounding bias ([Stürmer et al. 2006](https://pubmed.ncbi.nlm.nih.gov/16632131/)). However, using propensity score methods may have advantages in terms of the transparency of study conduct:

* Propensity scores are developed without reference to outcome data, which can reduce the risk of selective reporting of results when combined with strong research governance processes.
* With certain propensity score methods it is possible to examine the similarity of intervention groups in terms of observed covariates, providing evidence on the extent to which comparability was achieved. Standardised differences of less than 0.1 are generally considered to indicate good balance.

Some methods including propensity scores involve restriction of the population. This is to ensure good balance in propensity scores between treatment arms. When using restriction methods, trade-offs between internal validity, power and generalisability should be considered. For studies of comparative effects, internal validity should generally be prioritised.

Further consideration should be given to the effect estimate of interest since some propensity score methods permit the estimation of average treatment effects (ATE - the average intervention or policy effect across the target population) or average treatment effect in the treated population (ATT - the average effect for subgroups of individuals who have actually had the intervention).

Time-varying confounders should typically not be adjusted for using the above methods. It may be acceptable for on-treatment analyses if confounders that vary over time are not affected by previous levels of the intervention but this is uncommon. G-methods including marginal structural models with weighting are preferred ([Pazzagli et al. 2017](https://pubmed.ncbi.nlm.nih.gov/29285840/), [Mansournia et al. 2017](https://pubmed.ncbi.nlm.nih.gov/29038130/)[).](https://pubmed.ncbi.nlm.nih.gov/28039382/) Adjustment for time-varying confounders requires high-quality data over the whole follow-up period.

Various sensitivity and bias analyses can be used to adjust for bias because of residual confounding or to explore its likely impact (see [assessing robustness of studies](#_Assessing_robustness_of)). This may be informed by external data on confounder-outcome relationships or data from a data-rich subsample of the analytical database, if available ([Ali et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31619986/)). Negative controls (that is, outcomes that are not expected to be related to the intervention) may also be useful ([Lipsitch et al. 2010](https://pubmed.ncbi.nlm.nih.gov/20335814/)).

If there are multiple potential sources of suitable real-world data to provide external control to trial data, developers should consider whether to estimate effects separately for each data source or to increase power by pooling data sources. Data sources should only be pooled when there is limited heterogeneity between sources in terms of coverage and data quality. Individual estimates of effects for each data source should always be provided.

External controls can also be used to supplement internal (or concurrent) controls in randomised controlled trials. There are several methods available to combine internal and external controls, which place different weight on the external data ([NICE Decision Support Unit 2020](https://nicedsu.sites.sheffield.ac.uk/methods-development/chte2020-sources-and-synthesis-of-evidence)).

Instrument-based approaches may be preferred if:

* there is expected to be confounding because of unknown or poorly measured confounders
* an appropriate instrument is available that is strongly correlated with the exposure of interest and does not affect the outcome except through the exposure.

Common instruments in healthcare applications include variation in physician treatment preferences, healthcare providers, geography, arbitrary thresholds for treatment or access, or time (for example, time of changes to clinical guidelines that have immediate and substantial impacts on care patterns).

It can be challenging to identify strong instruments in healthcare. The assumptions needed for valid estimation include that the instrument affects the outcome only through the intervention and is not influenced by other variables that also influence the outcome. Also, these methods typically estimate the treatment effect in a subgroup of people for whom the intervention status is determined by the instrument (known as local average treatment effect) and this may not always be the effect of interest ([Matthay et al. 2019](https://pubmed.ncbi.nlm.nih.gov/31890846/)). Further technical guidance on methods for addressing baseline confounding using individual patient level data is given in [NICE's technical support document 17](http://nicedsu.org.uk/technical-support-documents/observational-data-tsd/).

### Addressing information bias

Limitations in data quality including missing data, measurement error or misclassification can cause bias and loss of precision. Here we describe analytical approaches to address information bias. The information needed to understand data suitability will provide an insight into the likely importance of information bias (see [data suitability assessment](#_Chapter__)).

#### Informative censoring

Censoring occurs in longitudinal studies if follow up ends before the outcome is fully observed. It can happen because the data collection period ends (administrative censoring), loss to follow up, occurrence of events such as treatment switching or non-adherence, or an adverse event. It may be induced by analytical strategies such as cloning to avoid time-related biases in studies without active comparators ([Hernan and Robins 2016](https://pubmed.ncbi.nlm.nih.gov/26994063/)). Censoring can create bias if it is informative (that is, it is related to the outcomes and treatment assignment). For example, in on-treatment analyses, if people on an experimental drug are less likely to adhere to the treatment protocol because of a perceived lack of benefit this could lead to informative censoring. When modelling effects on-treatment or dynamic treatment strategies, censoring because of treatment switching is likely to be informative. Methods to address informative censoring are similar to those for time-varying confounding such as marginal structural models with weighting or other G-methods ([Pazzagli et al. 2017](https://pubmed.ncbi.nlm.nih.gov/29285840/)). Methods for dealing with [missing data](#_Missing_data) may also be used.

#### Missing data

The impact of missing data depends on the amount of missing data, the variables that have missing data, and the missing data mechanism. Developers should compare patterns of missingness across exposure groups and over time, if relevant, considering causes of missingness and whether these are related to outcomes of interest. Missing data on outcomes may arise for a number of reasons including non-response to questionnaires or censoring.

If the amount of missing data is low and likely to be missing completely at random, complete records analysis will be sufficient. Advanced methods for handling missing data include imputation, inverse probability weighting and maximum likelihood estimation. Most of these methods assume the missing data mechanism can be adequately modelled using available data (that is, missing at random). If this is not the case, sensitivity or bias analysis may be preferred (see [assessing robustness](#_Assessing_robustness_of)). A framework for handling missing data is provided in [Carpenter and Smuk 2021](https://pubmed.ncbi.nlm.nih.gov/33624862/).

#### Measurement error and misclassification

Measurement error describes the extent to which measurements of study variables deviate from the truth. For categorical variables, this is known as misclassification. The impact of measurement error depends on the size and direction of the error, the variables measured with error, and whether error varies across intervention groups. Measurement error can induce bias or reduce the precision of estimates.

Random measurement error in exposures tends to (but does not always) bias estimates of treatment effects towards the null ([van Smeden et al. 2020](https://pubmed.ncbi.nlm.nih.gov/31821469/)). Random measurement error in continuous outcomes reduces the precision of estimates but provides unbiased estimates of comparative effects. Non-differential misclassification of a categorical outcome provides unbiased estimates of comparative effects when specificity is 100%, even if sensitivity is low. So it is often recommended to define outcome variables to achieve high specificity.

Differential measurement error in exposures, covariates or outcomes generally produces biased estimates of comparative effects but the direction of bias can be hard to predict. If data is available on the likely structure and magnitude of measurement error (for example, through an internal or external validation study), this information can be incorporated into analyses using calibration or other advanced methods ([van Smeden et al. 2020](https://pubmed.ncbi.nlm.nih.gov/31821469/)).

## Assessing robustness of studies

The complexity of studies of comparative effects using real-world data means developers must make many uncertain decisions and assumptions during data curation and analysis. These decisions can have a large impact, individually or collectively, on estimates of comparative effects. It is therefore essential that the robustness of results to deviations in these assumptions is demonstrated. We describe key sensitivity analyses across several domains in [table 4](#Table4). Which sensitivity analyses to focus on will vary across use cases depending on the strengths and weaknesses of the data as well as the areas in which the impact of bias, study assumptions and uncertainty are greatest. These approaches can be applied directly to measures of clinical effectiveness or propagated through to cost-effectiveness analyses.

For some important risks of bias (for example, those arising because of unmeasured confounding, missing data or measurement error in key variables), quantitative bias analysis may be valuable. Quantitative bias analysis describes a set of techniques that can be used to:

* estimate the direction, magnitude and uncertainty of bias associated with measures of effect, or
* examine the extent to which bias would have to be present to change results or affect a threshold for decision making.

There are a range of complex methods that incorporate external information of bias into studies ([Lash et al. 2016](https://pubmed.ncbi.nlm.nih.gov/25080530/)). If used, the identification and validation of the external evidence should be clearly described and its use justified.

Methods that examine the extent to which bias would have to be present to change study conclusions tend to be simpler and include the e-value approach. Developers should consider and pre-specify a plausible level of bias in the parameter before application of these methods.

Bias analysis may be particularly valuable in studies using real-world data external controls if differences between data collection, settings and time may reduce comparability of data. [Panel 2](#Panel2) shows an example of bias analysis in practice, and [table 4](#Table4) shows examples of sensitivity analysis.

#### Panel 2

Example of bias analysis

|  |
| --- |
| What is the effectiveness of the ALK-inhibitor alectinib compared with ceritinib in crizotinib-refractory, ALK-positive non-small-cell lung cancer?  The comparative effectiveness of alectinib versus ceritinib on overall survival in patients with ALK-positive non-small-cell lung cancer is uncertain because of a lack of head-to-head trials. [Wilkinson et al. 2021](https://pubmed.ncbi.nlm.nih.gov/34618040/) used real-world data on ceritinib from the Flatiron Health database (derived from US electronic health records) to form an external control to the alectinib arm of a phase 2 trial. The authors found a significant improvement in survival for those initiating alectinib. However, the study was at risk of residual bias from unmeasured confounding and missing baseline data on Eastern Cooperative Group Performance Status (ECOG PS) in patients having ceritinib (47% of patients had missing data).  Bias analysis methods were used to explore these risks. The e-value approach was used to estimate the relative risk of the unknown confounder between intervention and mortality that would be needed to remove the treatment effect. The estimated relative risk of 2.2 was substantially higher than for any observed confounders.  For missing ECOG data they assumed the causes of missing data were non-random and missing data values in the ceritinib arm were likely to be worse than expected based on multiple imputation. They argued that no plausible assumptions about missing data could explain the observed association between intervention and mortality. |

## Table 4

Examples of sensitivity analyses to examine robustness of results to data curation, study design, and analysis decisions

| Domain | Example sensitivity or bias analysis |
| --- | --- |
| Exposure misclassification | * On-treatment analyses * Vary exposure definitions including, if relevant, days’ supply, grace period, washout period, exposure effect window and latency |
| Outcome misclassification | * Adjust for known performance metrics * Quantitative bias analysis |
| Population | * Alternative patient eligibility criteria |
| Detection bias | * Include measures of healthcare use as covariates * Restrict to those with regular contact with the health system before baseline |
| Follow-up time | * As-started and on-treatment analyses * Restrict outcome period so it is similar between groups for informative censoring * Prevalent-user and new-user analyses |
| Reverse causation | * Introduce or change lag time between exposure end and start of follow up for outcomes |
| Confounding | * Add or remove selected confounders * Extend look-back period over which covariates are identified * Use negative controls (also known as falsification endpoints or probe variables) to estimate comparative effects using the same model on outcomes, which should not be related to treatment (results from these can also be used to calibrate effect estimates) * Propensity score calibration to adjust observed effect estimates for unmeasured bias using variables observed in a validation study * Quantitative bias analysis (for example, e-value approach to estimate the influence of an unknown confounder) |
| Missing data | * Use different methods * Include missing variable indicators for covariates in statistical models * Quantitative bias analysis (for instance, assuming missing not at random mechanisms) |
| Model specification | * Vary model specifications |
| Data curation | * Alternative categorisations of continuous variable or adjust data exclusions |

## Reporting

We provide general principles for the transparent reporting and good conduct of real-world evidence studies in the [study conduct section](#_Use__) The following reporting considerations are especially important for comparative effects studies:

* Justification of the use of real-world evidence. This should cover, as relevant, the reasons for the absence of randomised evidence, the limitations of existing trials and the ability to produce meaningful real-world evidence for the specific research question.
* Publish a study protocol (including statistical analysis plan) on a publicly accessible platform before the analysis is done.
* Report studies in sufficient detail to enable the study to be reproduced by an independent researcher.
* Present study design diagrams.
* For each data source, provide the information needed to understand data provenance and fitness-for-purpose (see [assessing data suitability).](#_Chapter__)
* Justify the use of statistical method for addressing confounding and report methods clearly (see appendix 3).
* Clearly describe the exclusion of patients from the original data to the final analysis, including reasons for exclusion.
* Present characteristics of patients across treatment groups, before and after statistical adjustment if possible. For external control studies, differences in variable definitions and data collection should be clearly described.
* Present results for adjusted and unadjusted analyses and for all subgroup and sensitivity and bias analyses.

## Quality appraisal

Evidence developers should identify risks of bias at the [study planning](#_Study_planning) stage. These should be described alongside how design and analytical methods have been used to address them, and how robust the results are to deviations from assumptions in the main analysis using sensitivity or bias analysis. This can be done for specific domains of bias using the reporting methods in appendix 2. This information will help those completing (or critically appraising) risk of bias tools. The preferred risk of bias tool for non-randomised studies is the ROBINS-I ([Sterne et al. 2016](https://pubmed.ncbi.nlm.nih.gov/27733354/)) but it should be recognised that it may not cover all risks of bias ([D’Andrea et al. 2021](https://pubmed.ncbi.nlm.nih.gov/33762237/)). It should be recognised that the uncertainty in non-randomised studies will not typically be fully captured by the statistical uncertainty in the estimated intervention effect ([Deeks et al. 2003](https://pubmed.ncbi.nlm.nih.gov/14499048/)).

Developers should comment on the generalisability of study results to the target population in the NHS. This may draw on differences in patients, care settings, treatment pathways or time and is supported by information provided from [data suitability assessment](#_Chapter__). It should be noted that differences in characteristics of patients in the analytical sample and the target population does not necessarily mean results do not generalise. This will depend on the scale of estimation (for example, relative versus absolute effects), the extent of heterogeneity in the treatment effect, and whether this is adequately modelled.

# Glossary

Terms included in the [NICE glossary](https://www.nice.org.uk/glossary) are not presented here.

Code list

A list of codes (such as SNOMED-CT codes) used to define a phenotype (such as a patient characteristic, condition or clinical event).

Covariate

A variable used in statistical analysis other that is not an exposure or outcome. This includes confounders, effect modifiers and predictive variables.

Data cleaning

The process of fixing or removing incorrect, duplicate, corrupt or inaccurate records in a dataset.

Data controller

Organisation which determines when and how data can be processed. A data controller may not necessarily access or process the data themselves.

Data curation

The process of preparing data for analysis including data linkage, cleaning, transformation and selection.

Data dictionary

A description of the contents, format and structure of a dataset.

Data element

A piece of structured information in a dataset such as age.

Data linkage

Process of combining sources of data for a single individual (or other entity). Linkage can be deterministic or probabilistic.

Data model

A representation of the data elements in a data source and how they relate to each other.

Data pooling

Process of combining sources of data on different individuals. It is often used to increase sample size especially in rare diseases.

Data provenance

Information about the origin of data, such as who it was collected by and how, the original purpose of collecting the data, and what happens to it over time.

Data transformation

Process of changing the format, structure or values of data to make it ready for analysis.

Developer

The group responsible for producing evidence that is used by NICE, which can be a team at NICE, a collaborating centre, a manufacturer or third party.

Estimand

A precise description of the treatment effect investigated by a study. This includes the treatment, population, primary outcome of interest, and how post-randomisation events will be handled.

External control

A control group that comes from a dataset or study than the group that received the intervention of interest.

Information governance

Framework for handling information in a robust and transparent manner, applying confidentiality and security where appropriate and operating to high ethical and quality standards (see [Health Research Authority](https://www.hra.nhs.uk/about-us/governance/information-governance/))

Non-randomised study

A comparative study which does not involve randomisation. This can include purely observational studies, non-randomised interventional studies, and single-arm trials with external control.

Omics

Areas of biological science that study cellular molecules such as genes, proteins or small metabolites. The individual fields of study have names that end in -omics, such as genomics or proteomics.

Primary data

Data collected and used for a specific research purpose.

Propensity score

The estimated probability of being assigned to a particular intervention conditional on a set of observed covariates.

Real-world data

Data collected outside the context of a highly-controlled clinical trial. Real-world data can be routinely collected during the delivery of health or social care. It can also be collected prospectively, to address 1 or more specific research questions.

Real-world evidence

Evidence generated from the analysis of real-world data. This includes studies using real-world data to form an external control to a clinical trial.

Reverse causation

Where the direction of causal effect is from the outcome (or symptoms thereof) to the intervention or exposure.

Secondary data

Data collected routinely or collected for research when reused in other studies.

Single-arm trial

A clinical trial in which all participants receive the experimental intervention of interest. This is especially common in phase I and II trials.

Statistical analysis plan

A technical document that describes in detail the planned statistical analyses for a study. It usually forms part of a study protocol.

Structured data

Data that follows a predefined data model such as patient characteristics, clinical diagnoses, and test results.

Unstructured data

Data that is not stored in a structured data model with pre-defined formats. Common types of unstructured data include free text or imagine data.