National Institute for Health and Care Excellence

Draft for consultation

Behaviour change: digital and mobile health interventions

NICE guideline: methods

NICE guideline <number>

Methods

[January 2020]

Draft for Consultation

Evidence reviews were developed by Public Health Guidelines Team

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

Development of the guideline

What this guideline covers

This guideline considers evidence on digital and mobile health interventions to initiate behaviour change. That is, interventions that deliver behaviour change techniques or components through a digital platform. This includes those delivered by text message, apps, wearable devices or the internet. These interventions will focus on changing any of the following established unhealthy behaviours to improve health:

- tobacco dependence
- hazardous or binge drinking
- unhealthy eating patterns, a lack of physical activity or sedentary behaviour
- unsafe sexual behaviour.

What this guideline does not cover

This guideline will not cover the following areas:

- National policy, fiscal and legislative measures.
- Clinical or pharmacological methods of achieving behaviour change with no public health or health promotion element. For example, appointment reminders, medication reviews or self-care solely to improve medicine adherence.
- Clinical interventions to help with the diagnosis, treatment or management of a chronic physical or long-term mental health condition.
- Psychiatric interventions delivered as part of the therapeutic process for people with a mental health problem, including digital or mobile health therapies that are used to treat depression, anxiety disorders, psychosis or other psychological conditions.
- Interventions delivered solely by a healthcare professional or practitioner (for example, counselling delivered over the telephone, video-links or by real-time live instant messaging).
- Changes to the public realm to support behaviour change (such as designing and managing public spaces in a way that encourages and helps people to be physically active).
- Digital or mobile health interventions to change the behaviour of healthcare professionals or other professionals who support people to change their unhealthy behaviours.
- Digital or mobile health interventions that aim to prevent the uptake of behaviours such as smoking, harmful drinking or unsafe sexual behaviour, and/or to help maintain healthy behaviours including relapse prevention.

Methods

This guideline was developed in accordance with the process set out in 'Developing NICE guidelines: the manual (2018)'. Additional methods are described below.

Declarations of interest were recorded according to the 2018 NICE conflicts of interest policy.

Developing the review questions and outcomes

The 4 review questions developed for this guideline were based on the key areas identified in the guideline scope.

The review questions were based on the following framework:

- population, intervention, comparator and outcome (PICO) for reviews of interventions

Full literature searches, evidence tables including critical appraisal for all included studies, tables of excluded studies with reasons for exclusion and evidence reviews were completed for all review questions.

Priority screening

As the diet and physical activity and sexual health search results returned a large (≥ 5000) number of results, priority screening was used to sift on title and abstract in EPPI-reviewer systematic reviewing software. The following approach was used:

- At least 50% of the total identified records were screened
- After this point, if no study was included after another 10% of the total identified records had been sifted, no further screening was conducted.

To ensure that no potential eligible studies have been missed using priority screening, the included studies and the reference list of the eligible systematic reviews were searched to identify any studies not identified through the primary search.

Reviewing research evidence

Evidence was identified for evidence reviews according to the methods in chapter 5 of "Developing NICE Guidelines: the manual" (2018). The purpose of the search was to identify the best available evidence to address review questions without producing an unmanageable volume of results.

Relevant databases and websites, (see Search strategies) were searched systematically to identify effectiveness and cost effectiveness research evidence. The principal database search strategy is listed in Search strategies. The principal strategy was developed in MEDLINE (Ovid interface) and was adapted for use in the other databases listed in Search strategies taking into account their size, search functionality and subject coverage.

Papers were included if they met the review protocol:

- Randomised controlled trials. Before and after studies and interrupted time series were also eligible for the unsafe sexual behaviour review.
- Systematic reviews of randomised controlled trials, if the majority of included studies met the PICO. If the majority of studies did not meet the PICO, individual studies included in the systematic review were considered separately for inclusion in this evidence review.
- Published from 2000 onwards.
- Published in English language.
- Had a follow up outcome measure from baseline of at least 6 months. Any follow up was eligible for the unsafe sexual behaviour review
- Full published studies (not protocols or summaries)

The searches were limited to studies from 2000 onwards. The committee decided that results before 2000 would not need to be considered because technology before this time would be outdated and not relevant to current technology.

Data extraction

Key data elements for each study were extracted as follows: study dates, country and setting, number of participants and attrition, population of interest, participant characteristics, inclusion and exclusion criteria, intervention, comparison group, data collection methods, data analysis methods and outcomes of interest. Information regarding behaviour change techniques (BCTs), intensity, tailoring and engagement were also extracted from each study. BCTs were categorised into the 16 clusters of techniques identified by Michie et al (2013).

The reported components and characteristics of interventions were extracted. These were extracted using the 12 item TiDieR checklist, which is a guide for extracting the elements that make up the intervention and comparator arms of a study.

Data synthesis for intervention studies

Randomised controlled trials were included in all reviews. In the unsafe sexual behaviour review before and after studies and interrupted time series were also included. Where an outcome was reported similarly by more than one study, a meta-analysis was conducted in order to pool the data from the included studies. Meta-analysis was undertaken in Cochrane Review Manager (version 5.3) and the data were pooled using either the Mantel–Haenszel method or the inverse variance method depending on how data were reported. Separate meta-analyses were conducted for dichotomous and continuous outcomes. A random effects model was used in order to take into account the variability of the studies (heterogeneity). Heterogeneity between studies was quantified using I^2 statistics. When $I^2 \geq 50\%$, subgroup analyses were carried out to explain the identified heterogeneity, except when there were an insufficient number of studies to do so. Subgroup analyses were used to determine the impact of population of interest (such as those with specific conditions), mode of delivery, and the effect of comparator group on the pooled result. Studies were grouped by mode of delivery according to the intervention types specified for inclusion in the review protocol under the following headings:

- Those delivered by the internet: such as by websites, emails, videos and multi-media)
- Text message-based services (including picture messages and audio messages)
- Wearable devices
- Apps
- Social media, networking and chat forums
- Digital gaming
- Virtual or augmented reality
- Interactive voice response interventions (IVR)

Interventions and studies were included based on the review protocol. If a study used more than one digital platform (such as text messages delivered alongside an app, or internet plus text messages) the study was grouped under the intervention which was predominant and a note that it was a mixed intervention was made in the data extraction tables. In the smoking review many of the interventions used more than one mode of delivery with no predominant intervention. Therefore, in this review the study was grouped as a mixed intervention.

A meta-analysis was not conducted:

- When the evidence from the outcome was only presented in a single study, a narrative summary description of the findings of the study was provided in order to enable committee discussion.

- Where studies reported outcomes in very different ways, it was not considered reasonable to pool these studies in the meta-analysis.

Data synthesis

For dichotomous outcomes, which used two response categories, risk ratio (RR) was the preferred effect measure for pooling the results for this guideline. Results presented as odds ratios (OR), were converted to RR. The event rate in the control arm was used as the prevalence in the calculation. Where confidence intervals were not reported for effect estimates on an ordinal scale, the P-Value and point estimate were used to derive the confidence intervals using RevMan.

When raw data were available, a 2x2 table was created and the RR was calculated. When a study defined the outcome in ordinal scale, the response categories were collapsed into two to develop a composite measure, which could be pooled in the meta-analysis. When studies used incidence rate and the raw data were also available, incidence rate was converted to RR and a 2x2 table was created.

For dichotomous data, absolute risks were also presented in GRADE. Absolute risks were calculated by applying the relative risk (and 95% confidence interval) to the control group risk (number with the event in the control group divided by total number in the control group). Where multiple studies are combined, control groups were summed and averaged using GRADEpro and expressed per 1000.

For continuous outcomes (mean value and SD were provided for individual studies), the mean difference was used as the effect estimate when studies included in the meta-analysis used a single scale to measure the outcome. When the studies

assessed the same outcome but used different measurement scales, the results were standardised to have the same standard deviation before they were combined Therefore, a standardised mean difference was used as the summary statistic for the meta-analysis. If the standard deviation for the baseline, follow-up or mean difference was not reported in the study, it was calculated from data available in the publication. If this was not possible, the study results were not included in meta-analysis and reported separately.

Smoking

According to changes in the initial protocol, only follow up data ≥ 6 month were eligible for the review. Interventions were grouped according to mode of delivery in the following categories:

- Internet based interventions
- Text messaging interventions
- and mixed interventions, including any combination of internet and text interventions (e.g. text & video, internet and mobile phone).

Specific rules of preferences were used for the outcome (smoking abstinence) as follow:

1. Where biochemically validated measures are available, these will be preferred to self-reported measures

2. Longest follow up was used

3. Where continuous or sustained abstinence was reported, will be preferred to point abstinence

Sensitivity analyses were conducted to assess if the following had an impact on effectiveness:

- Pregnant women

Unsafe sexual behaviour

As it was anticipated that there would be less evidence available for this review , the ≥ 6-month follow-up was not applied and the study type included RCTs and controlled before and after studies. When results were reported at more than one follow-up, the longest follow-up was used. For dichotomous data, risk ratios were reported as intervention vs control groups at follow-up. For continuous outcomes (mean value and SD were provided for individual studies),

Sensitivity analyses were conducted to assess if the following had an impact on effectiveness:

- Condom use at last intercourse

Alcohol consumption

Studies with ≥ 6-month follow-up data were included. Change in alcohol consumption between baseline and follow-up was calculated for each intervention and control arm, which were then compared by mean difference and standard deviation. All data was continuous.

Sensitivity analyses were conducted to assess if the following had an impact on effectiveness:

- Weekly alcohol consumption, higher consumption was classed as ≥14 units a week
- Digital platform
- Non-students

Diet and exercise

Sensitivity analyses were conducted to assess if the following had an impact on effectiveness:

- Medical condition
- Digital platform

Publication bias

Funnel plots were used for visual assessment of the publication bias, where data for at least 10 studies were included in a single meta-analysis.

Behaviour Change Technique Taxonomy

A hierarchically structured taxonomy of behaviour change techniques (BCTs) was used. This taxonomy included 93 BCTs clustered into 16 groups (Michie et al 2013). This reliable taxonomy of 16 theoretical clusters of BCTs was used to code BCTs used in the intervention arms of the study.

The 16 clusters are;

- scheduled consequences, reward and threat, repetition and substitution, antecedents, associations, covert learning, natural consequences, feedback and monitoring, goals and planning, comparison of the behaviour, social support, self-belief, comparison of outcomes., identity, shaping knowledge regulation.

Summarising components and characteristics of the interventions

Intervention matrix tables were created in Excel in order to summarise the different components and characteristics of the interventions and identify their effectiveness for each review questions as well as to identify common effective components and characteristics across the four review questions. These tables were used to aid committee discussion due to complexity of the data.

Appraising the quality of evidence

Risk of bias

Quality assessment for all included RCTs was conducted using the Cochrane Risk of Bias 2 tool (2016) for individual RCTs and cluster RCTs. The quality of each individual study was assessed at outcome level using this tool.

The quality was interpreted as follows:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Some concerns – There is a possibility the true effect size for the study is substantially different from the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different from the estimated effect size.

GRADE for interventional evidence

GRADE was used to assess the quality of evidence for the selected outcomes as specified in 'Developing NICE guidelines: the manual (2018)'. Data from all RCT's were initially rated as high quality and the quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table 1.

Table 1: GRADE

Quality domain	Description
Risk of bias	Limitations in study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect. Examples of such limitations are selection bias (often due to poor allocation concealment), performance and detection bias (often due to a lack of blinding of the patient, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis). Where there are no study limitations (low risk of bias), evidence is assessed as having 'no serious' risk of bias. Alternatively, evidence may be downgraded one level to 'serious' risk of bias (some concerns of bias or two levels to 'very serious' risk of bias (high risk of bias).
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question. Where the evidence is directly applicable to the PICO, it is assessed as having 'no serious' risk of indirectness. Alternatively, evidence may be downgraded one level ('serious' risk of indirectness) or two levels ('very serious' risk of indirectness).
Inconsistency	Inconsistency refers to an unexplained heterogeneity of effect estimates between studies pooled in the same meta-analysis. The I^2 statistic describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). Heterogeneity could be explained by differences in study design, content of interventions and comparators, or differences in clinical risk factors between study populations. Subgroup analysis will be conducted to explain the reasons for the heterogeneity. A decision was made to downgrade pooled analyses

Quality domain	Description
	by 1 level (indicating 'serious' inconsistency) when the I^2 statistic was ≥50% and 2 levels (indicating very serious inconsistency) when the I^2 statistic was ≥75%.
Imprecision	Results are imprecise when studies include relatively few patients and few events (or highly variable measures) and thus have wide confidence intervals around the estimate of the effect relative to clinically important thresholds. 95% confidence intervals denote the possible range of locations of the true population effect at a 95% probability, and so wide confidence intervals may denote a result that is consistent with conflicting interpretations (for example a result may be consistent with both public health benefit AND public health harm) and thus be imprecise. Imprecision was assessed with reference to minimally important difference (MID) thresholds for individual outcomes (smallest change in an outcome that is considered important by patients or health care professionals). Established MIDs may be published in previous literature and seen and accepted in clinical community. It was decided that the point measure would be used to decide whether or not the result was clinically important, and that the 95% confidence intervals would indicate certainty of this importance. Uncertainty is introduced where confidence intervals crossed the MID threshold. If the confidence interval crosses either the lower or upper MID threshold this indicates 'serious' risk of imprecision. Crossing both MID thresholds indicates 'very serious' risk of imprecision in the effect estimate. Default MIDs were used in this. Default MIDs are used where no established MID's for individual outcomes are found (0.-8 and 1.25 for dichotomous outcomes and 0.5*SD of control group at baseline for continuous outcomes). If the MID could not be calculated (e.g. because standard deviation of outcome measure at baseline was not reported in the paper) then we downgraded by 1 level as it was 'not possible to calculate imprecision from the information reported in the study'. Where data was pooled in analyses, the study with the largest weight was used as the control group for default MID calculations. Where the 95% CI does not cross either MID threshold, the evidence is assessed as having 'no serious' risk of imprecision unless the effect estimate is derived on the basis of few events and a small study sample (that is, less than 300 events for dichotomous outcomes or total sample size less than 400 for continuous outcomes). In that case the results were downgraded one level for 'serious' imprecision to reflect uncertainty in the effect estimate.

Quality domain	Description
Other issues	Publication bias is a systematic underestimate or overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. A closely related phenomenon is where some papers fail to report an outcome that is inconclusive, thus leading to an overestimate of the effectiveness of that outcome. Sometimes randomisation may not adequately lead to group equivalence of confounders, and if so this may lead to bias, which should be taken into account. Potential conflicts of interest, often caused by excessive pharmaceutical company involvement in the publication of a study, should also be noted.

Details of how the 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) were appraised for each outcome are given below in table 2, Publication or other bias was only taken into consideration in the quality assessment if it was apparent.

Table 2: GRADE rating

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

Evidence statements

GRADE profiles provide full details of results. Evidence summaries are intended to replace evidence statements, and to provide a high-level overview to summarise GRADE profiles.

Summary statements were written as follows:

Summary statement	Meaning
There was no meaningful difference between comparators	Where the CI is confined within the two MID thresholds

Summary statement	Meaning
An effect was not detected of the intervention on the outcome	Where CIs include the line of no effect and one or both MIDs
The intervention was effective at reducing / increasing the outcome, but the change was not meaningful	Where the CI includes an MID but does not include the line of no effect, and the point estimate is not meaningful.
The intervention was effective at reducing / increasing the outcome	Where the CI does not include the line of no effect. It may include the MID, but the point estimate is meaningful.
An effect estimate could not be calculated	Narrative description of the result

Reviewing economic evidence

A literature review was conducted to identify published economic evaluations of relevance to all questions in the guideline. A single unified search for all questions (smoking, alcohol, diet and physical activity, unsafe sexual behaviour) was carried out in January 2019 retaining behaviour change, digital media and condition-specific terms from the searches for public health effectiveness evidence with economic filters added. Economic evidence profiles, including critical appraisal according to the 'Developing NICE guidelines: the manual (2014)' were completed for included studies. A re-run search was conducted in August 2019 to identify any new economic evidence that had been published during guideline development.

Economic studies identified through a systematic search of the literature are appraised using a methodology checklist designed for economic evaluations. This checklist is not intended to judge the quality of a study per se, but to determine whether an existing economic evaluation is useful to inform the decision-making of the committee for a specific topic within the guideline.

There are 2 parts in the appraisal process. The first step is to assess applicability (that is, the relevance of the study to the specific guideline topic and the NICE reference case); evaluations are categorised according to the criteria in Table 3.

Table 3: Economic evidence applicability criteria

Level	Explanation
Directly applicable	The study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness
Partially applicable	The study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness
Not applicable	The study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. These studies are excluded from further consideration

In the second step, only those studies deemed directly or partially applicable are further assessed for limitations (methodological quality); see categorisation criteria in Table 4.

Table 4: Economic evidence methodological quality criteria

Level	Explanation
No/minor limitations	Meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness
Potentially serious limitations	Fails to meet one or more quality criteria and this could change the conclusions about cost effectiveness
Very serious limitations	Fails to meet one or more quality criteria and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration

Where relevant, a summary of the main findings from the systematic search, review and appraisal of economic evidence is presented in an economic evidence table alongside the public health evidence on effectiveness.

Health economic modelling

In light of the limitations of the published economic evidence, the option to undertake original economic modelling was considered for all review questions in the guideline. Given the focus of the review questions on identifying effective components and characteristics of digital and mobile health interventions (rather than on the interventions themselves), it was felt that economic modelling around components and characteristics was unlikely to be feasible or to provide meaningful information beyond the evidence that was identified through the literature review. Therefore, no original economic modelling was undertaken for this guideline.