National Institute for Health and Care Excellence

Final

Workplace health: long-term sickness absence and capability to work

NICE guideline: methods

NICE guideline NG146

Methods

November 2019

Final

Evidence reviews were developed by Public Health Internal Guideline Development 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-3597-0

Contents

Development of the guideline

What this guideline covers

Everyone aged over 16 in full or part-time employment who has had:

- Long-term sickness absence (4 or more weeks)
- Recurring short-term sickness absence (less than 4 weeks per episode)

What this guideline does not cover

People who are self-employed. Pregnant women who have taken sickness absence related to their pregnancy.

Methods

This guideline was developed in accordance with the process set out in 'Developing NICE guidelines: the manual (2018)'. A booklet, 'How NICE guidelines are developed: an overview for stakeholders, the public and the NHS' is available. In instances where the guidelines manual does not provide advice, 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

This is an update of a previous guideline (Workplace health: long-term sickness absence and incapacity to work PH19). Three of the four areas in the previous guideline were identified as requiring an update by a surveillance report carried out in 2017 The three review questions developed for this guideline were based on these and key areas identified in the guideline scope. One remaining area in PH19 was considered not to need an update (for those who are unemployed and receiving benefits) and the recommendation from this area was retained. Review questions to cover these key areas were drafted by the NICE Public Health Internal Guideline Development team, and refined and validated by the guideline committee.

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.

Reviewing research evidence

The identification of evidence for evidence review in the guideline conformed to the methods set out in chapter 5 of the "Developing NICE Guidelines: the manual" (2014). 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, cost effectiveness and qualitative research evidence. The principal database search strategy is listed in Search strategies (Appendix B). The principal strategy has been developed in MEDLINE (Ovid interface) and was be adapted, as appropriate, for use in the other sources listed in Search strategies taking into account their size, search functionality and subject coverage. As this was an update of existing guidance, evidence relevant to the new review protocols from the previous guideline was included for assessment. Review protocols were developed for each review question. The protocols are in Appendix A. No date limit was applied to the searches.

The study selection approach followed inclusion/ exclusion approach set out in the review protocols Randomised controlled trials were included if they evaluated interventions related to each specific review question. Systematic reviews of intervention studies were used as a source for primary studies. The committee

further requested the inclusion of large UK based observational studies. Qualitative studies from the UK were included wherever exploring views, preferences and/or experiences of those returning to work or aiming to return to work.

Papers were excluded if they:

- were not published in the English language or were not carried out in OECD countries
- were only available as abstracts, conference proceedings, guideline/health technology assessment reports

Expert testimony from four topic experts was provided to supplement and provide additional context to areas with limited published evidence. The experts presented at committee meetings and responded to committee questions and contributed to committee discussion. They also completed expert testimony proformas for inclusion in the review question sections. The committee used the expert testimony to consider the development of recommendation in areas where there was no, or insufficient evidence identified for the review questions. When drafting recommendations, the committee reflected on the key points emerging from the testimonies presented to them.

The committee identified areas where they agreed additional contextual information would augment the evidence identified by the reviews and could potentially inform the development of recommendations. Expert testimony from four topic experts was provided to supplement and provide additional context to areas with limited published evidence. These were;

- The role of an occupational health and wellbeing service in supporting the management of sickness absence and return to work in an NHS Trust
- Support for employees with a mental health condition to return to and stay in work
- Reducing sickness absence in the workplace – from the perspective an expert in employment research
- Support available for return to work and the use of workplace adjustments

Methods of combining evidence

Data synthesis for intervention studies

Where an outcome was reported similarly by more than one included study, data were pooled using meta-analyses performed in Cochrane Review Manager v5.3 as follows:

- Dichotomous outcomes – data were pooled on the relative risk scale (using the Mantel–Haenszel method). Absolute risks were calculated by applying the relative risk to the pooled risk in the control arm of the meta-analysis
- Continuous outcomes – data were pooled on mean difference if the unit of measurement was meaningful on a practical level. If the metrics of variables being studied had no intrinsic meaning (e.g. scores on measure of depression using an arbitrary scale). Where some or all of the studies used different scales and it was considered that the outputs would be interpretable and usable by the committee to potentially make recommendations (e.g. mean number of

sickness days over differing follow-up periods), data were analysed using standardised mean differences (Hedges' g).

- For time-to-event data, unadjusted hazard ratios were pooled, where reported, using the inverse variance method.
- Adjusted ratios from multivariate models were used in pooled analyses only if unadjusted values were not reported.
- Where data from more than one study were pooled in a meta-analysis, a random effects model was used to account for the different effects anticipated across the included studies, which had differing population inclusion criteria, intervention content and duration and different comparators. A random effects modal was used as the committee considered that there was likely to be considerable heterogeneity among the studies. The studies were likely to include multi-faceted interventions applied to potentially quite different employment sectors. Where high I^2 values are found subgroup analysis, such as on employment sector, reasons for absence or differing interventions were considered.
- Where cluster RCTs have statistically adjusted for the effects of clustering and have reported the adjusted OR/RR or MD and 95% confidence intervals, this data was extracted and included.

Where studies included very different interventions and had reported outcomes in different ways it was not considered reasonable to pool the studies by outcome into a meta-analysis. Though meta-analysis was not considered reasonable, some of the data from the studies have been organised within forest plots to enable committee discussion, summary statistics have not been calculated or presented due to the differences in the included studies. Where pooling of evidence was not possible evidence statements have been presented on an individual study-by-study basis.

For evidence synthesis, interventions have been categorised into 3 types, as follows:

- Individual employee-focused interventions: these interventions are primarily aimed at increasing the personal physical and / or mental resources of the absent employee to enable them to return to work. Such interventions may include an educational element and behavioural strategies such as goal-setting, graded activity, cognitive restructuring, behavioural activation, stress management and problem-solving skills.
- Workplace-focused interventions: typically these interventions address the structure and environment of the workplace to identify and reduce barriers to the employee's return to work. Such interventions typically involve the agreement of a 'return to work plan' which may include modifications to the content or pattern of working, and other adaptations to meet the needs of the employee.
- Combined interventions incorporating elements of both individual- and workplace-focused interventions.

Data synthesis for qualitative reviews

A secondary thematic analysis was performed to synthesise identified qualitative evidence. Qualitative evidence was categorised into key themes identified from the included qualitative evidence and the committee expertise and knowledge of current

practice (for example use of fit notes). Data from individual studies was narratively described under these themes. Quotes representing the overall trend of the evidence within each theme were also reported.

Appraising the quality of evidence

Intervention studies

Quality of individual studies were assessed using the Cochrane Risk of Bias, ROBINS-I or the CASP qualitative research checklist for each particular group.

GRADE methodology was used to appraise effectiveness evidence across five potential sources of uncertainty: risk of bias, indirectness, inconsistency, imprecision and other issues. Overall GRADE ratings start at 'High' where the evidence comes from RCTs, and 'Low' for evidence derived from observational studies.

GRADE 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 study subject, healthcare professional or assessor) and attrition bias (due to missing data causing systematic bias in the analysis). Where there are no study limitations, evidence is assessed as having 'no serious' risk of bias. Alternatively, evidence may be downgraded one level ('serious' risk of bias) or two levels ('very serious' 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). For the purposes of this review, the committee agreed that a large amount of clinical and methodological diversity would be expected from pooled analyses of studies in this area. Heterogeneity could be explained by differences in study design, content of interventions and comparators, or differences between study populations. A decision was therefore made to downgrade pooled analyses by 1 level (indicating 'serious' inconsistency) only when the I^2 statistic was ≥75%. If the I^2 statistic for a pooled analysis was less than 75%, the evidence was not downgraded for inconsistency. Where heterogeneity of ≥75% occurred subgroup analysis was considered, if there were sufficient studies included for this to be feasible. If the committee requested the inclusion of pooled analysis where ≥75% reasons for possible heterogeneity were discussed by the committee.

GRADE domain	Description
Imprecision	Results are imprecise when studies include relatively few participants 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. No minimally important difference (MID) thresholds relevant to this guideline were identified from the COMET database or other published source. Topic experts agreed that any change in sickness absence or return-to-work (the primary outcome of interest) should be considered a minimally important difference. The committee viewed that any change in these outcomes was considered clinically important and that it would show evidence of effectiveness of the interventions. They considered any reduction in sickness absence to be meaningful. Imprecision was therefore assessed with reference to the line of no effect for all data, binary or continuous. 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 the certainty of this importance. Uncertainty is introduced where confidence intervals crossed the MID threshold (here, the line of no effect). If the confidence interval crosses the line of no effect this indicates 'serious' risk of imprecision. Where the 95%CI does not cross the 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 participants across both intervention and comparator groups). In that case, the results were downgraded one level for 'serious' imprecision to reflect uncertainty in the effect estimate.

Overall 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.

Qualitative reviews

Individual qualitative studies were initially quality assessed using the CASP qualitative checklist to support the use of CERQual. Each individual study was classified into one of the following three groups:

- Low risk of bias – The findings and themes identified in the study are likely to accurately capture the true picture.
- Moderate risk of bias – There is a possibility the findings and themes identified in the study are not a complete representation of the true picture.

- High risk of bias – It is likely the findings and themes identified in the study are not a complete representation of the true picture

CERQual methodology was used to appraise qualitative evidence across four potential sources of uncertainty: methodological limitations, coherence, adequacy and relevance. All CERQual assessment of confidence in the evidence ratings start at 'High'.

CERQual criteria	Reasons for downgrading or not downgrading confidence
Methodological limitations	The certainty of the evidence was downgraded if there were concerns about the design or execution of the study including whether the research design and methods of data collection were appropriate to address the aims of the research, researcher reflexivity, ethical consideration and the clarity of findings.
Coherence	Assesses how clear and convincing the fit is between the data from the primary studies and the review findings based on the synthesis of the data. The certainty of the evidence was downgraded if some of the data contradicts the review finding.
Adequacy	Assesses the degree of richness and the amount of data to support the review finding. The certainty of the evidence was downgraded if the data was not sufficiently rich or if there was an insufficient number of studies to support findings.
Relevance	Assesses the extent to which the data from the primary studies supporting the review finding is applicable to its context. The certainty of the evidence was downgraded if the data available was not applicable to the review question.

Each CERQual criteria is described using one of the following levels of concern:

- *No or very minor concerns* that are unlikely to reduce confidence in the review finding
- *Minor concerns* that may reduce confidence in the review finding [concerns should be described]
- *Moderate concern* that will probably reduce confidence in the review finding [concerns to be described]
- *Serious concerns* that are very likely to reduce confidence in the review finding

Reviewing economic evidence

The PHAC is required to make decisions based on the best available evidence of both general effectiveness and cost-effectiveness. Guideline recommendations should be based on the expected costs of the different options in relation to their expected benefits (that is, their 'cost-effectiveness') rather than the total implementation cost. Thus, if the evidence suggests that a strategy provides significant benefits at an acceptable cost per person treated, it should be recommended.

In order to assess the cost effectiveness of the key issues addressed in this guideline, the following actions were carried out:

- A systematic review of economic evidence in the literature was conducted
- A de novo economic model was developed, in order to provide cost effectiveness evidence on interventions to enable return to work

Inclusion and exclusion of economic studies

The study selection approach followed the inclusion/ exclusion approach set out in the review protocols. Economic studies (economic evaluations, cost-utility evaluations, cost benefit analyses, cost-effectiveness analyses, cost minimization analyses, cost-consequence analyses) were included if they evaluated interventions related to each specific review question. Systematic reviews of intervention studies were used as a source for primary studies.

Papers were excluded if they:

- were not published in the English language or were not carried out in OECD countries
- were only available as abstracts, conference proceedings, guideline/health technology assessment reports

Appraising the quality of economic evidence

The quality of individual studies was assessed using the NICE quality assessment checklist for economic evaluations. The two-part checklist assesses applicability of a study's findings to the review in one part and assesses to what degree the study's limitations are, with regard to the review, in the second part.

Applicability was assessed through an eight-item checklist with 'yes'/'partly'/'no'/'unclear'/'not applicable' answers, each judgement accompanied by a brief narrative justification. These assess the appropriateness of various facets to the review, e.g. population, interventions, evaluation methodology and study setting and context. An overall judgement of the applicability is made ('directly applicable'/'partially applicable'/'not applicable') based on these items. This helps the reader determine how generalisable the study's findings may be.

Limitations were assessed through an eleven-item checklist with 'yes'/'partly'/'no'/'unclear'/'not applicable' answers, each judgement accompanied by a brief narrative justification. These items assess the economic evaluation methodology used in the study, e.g. model structure, assumptions, data sources, outcomes measured and reported and conflicts of interest. An overall judgement of the study limitations is made ('no limitations'/'minor limitations'/'potentially serious limitations') based on these items.

Health economic modelling

As well as reviewing the published economic literature for each review question, as described above, a de novo economic analysis was undertaken for relevant research questions. The following general principles were adhered to in developing the analysis:

- Methods were consistent with the NICE reference case (https://www.nice.org.uk/process/pmg20/chapter/incorporating-economic-evaluation).

- The committee was involved in the design of the model, selection of inputs and interpretation of the results.
- Where possible, model inputs were based on the systematic review of the effectiveness literature, supplemented with other published data sources identified by the committee as required.
- When published data were not available committee expert opinion was used to populate the model.
- Model inputs and assumptions were reported fully and transparently.
- The results were subject to sensitivity analysis and limitations were discussed.

For further details on the review and modelling see separate report by York Health Economics Consortium (YHEC).