A through C

D through N

Absolute risk (AR)
The probability that an individual will experience the specified outcome during a specified period. It lies in the range 0 to 1, or is expressed as a percentage. In contrast to common usage, the word "risk" may refer to adverse events (such as myocardial infarction) or desirable events (such as cure).

Absolute risk increase (ARI)
The absolute difference in risk between the experimental and control groups in a trial. It is used when the risk in the experimental group exceeds the risk in the control group, and is calculated by subtracting the AR in the control group from the AR in the experimental group. This figure does not give any idea of the proportional increase between the two groups: for this, relative risk (RR) is needed.

Absolute risk reduction (ARR)
The absolute difference in risk between the experimental and control groups in a trial. It is used when the risk in the control group exceeds the risk in the experimental group, and is calculated by subtracting the AR in the experimental group from the AR in the control group. This figure does not give any idea of the proportional reduction between the two groups: for this, relative risk (RR) is needed.

Allocation concealment
A method used to prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups. Allocation concealment prevents researchers from (unconsciously or otherwise) influencing which intervention group each participant is assigned to.

Applicability
The application of the results from clinical trials to individual people. A randomised trial only provides direct evidence of causality within that specific trial. It takes an additional logical step to apply this result to a specific individual. Individual characteristics will affect the outcome for this person. CE guidance People involved in making decisions on health care must take relevant individual factors into consideration. To aid informed decision-making about applicability, we provide information on the characteristics of people recruited to trials.

Baseline risk
The risk of the event occurring without the active treatment. Estimated by the baseline risk in the control group. CE guidance The base line risk is important for assessing the potential
beneficial effects of treatment. People with a higher baseline risk can have a greater potential benefit.

**Best evidence**

Systematic reviews of RCTs are the best method for revealing the effects of a therapeutic intervention.

CE guidance Usually only systematic reviews of randomised controlled trials (RCTs) and RCTs will be accepted in the Benefits section. However, sometimes other evidence is sufficient to assign causality and in this case an RCT would not be ethical. In other cases RCTs are not practical. In these instances it is legitimate to include other forms of evidence within the Benefits section. RCTs are unlikely to adequately answer clinical questions in the following cases:

1. where there are good reasons to think the intervention is not likely to be beneficial or is likely to be harmful;
2. where the outcome is very rare (e.g. a 1/10000 fatal adverse reaction);
3. where the condition is very rare;
4. where very long follow up is required (e.g. does drinking milk in adolescence prevent fractures in old age?);
5. where the evidence of benefit from observational studies is overwhelming (e.g. oxygen for acute asthma attacks);
6. when applying the evidence to real clinical situations (external validity);
7. where current practice is very resistant to change and/or patients would not be willing to take the control or active treatment;
8. where the unit of randomisation would have to be too large (e.g. a nationwide public health campaign); and
9. where the condition is acute and requires immediate treatment. Of these, only the first case is categorical. For the rest the cut off point when an RCT is not appropriate is not precisely defined. If RCTs would not be appropriate we search and include the best appropriate form of evidence.

**Bias**

Systematic deviation of study results from the true results, because of the way(s) in which the study is conducted. CE guidance In the Comment section we aim to include any likely sources of bias within a trial/review.

**Blinding/blinded**

A trial is fully blinded if all the people involved are unaware of the treatment group to which trial participants are allocated until after the interpretation of results. This includes trial participants and everyone involved in administering treatment or recording trial results. CE guidance Ideally, a trial should test whether people are aware of which group they have been allocated to. This is particularly important if, for example, one of the treatments has a distinctive taste or adverse effects. Unfortunately such testing is rare. The terms single and
double blind are common in the literature but are not used consistently. Therefore, we attempt to report specifically who is unaware of treatment allocation.

Block randomisation
Randomisation by a pattern to produce the required number of people in each group.

C

Case control study
A study design that examines a group of people who have experienced an event (usually an adverse event) and a group of people who have not experienced the same event, and looks at how exposure to suspect (usually noxious) agents differed between the two groups. This type of study design is most useful for trying to ascertain the cause of rare events, such as rare cancers. CE guidance Case control studies can only generate odds ratios (OR) and not relative risk (RR). Case control studies provide weaker evidence than cohort studies but are more reliable than case series. We do not include case control studies within the Benefits section, unless it is not reasonable to expect higher levels of evidence.

Case series
Analysis of series of people with the disease (there is no comparison group in case series). CE guidance Case series provide weaker evidence than case control studies. We try not to include case series within the Benefits section.

Cluster randomisation
A cluster randomised study is one in which a group of participants are randomised to the same intervention together. Examples of cluster randomisation include allocating together people in the same village, hospital, or school. If the results are then analysed by individuals rather than the group as a whole bias can occur. CE guidance The unit of randomisation should be the same as the unit of analysis. Often a cluster randomised trial answers a different question from one randomised by individuals. An intervention at the level of the village or primary care practice may well have a different effect from one at the level of an individual patient. Therefore, trying to compensate by allowing for intra class correlation coefficients or some other method may not be appropriate. Clinical Evidence style is to include only results analysed according to the unit of randomisation; otherwise the trial is included only in the Comment.

Cohort study
A non-experimental study design that follows a group of people (a cohort), and then looks at how events differ among people within the group. A study that examines a cohort, which differs in respect to exposure to some suspected risk factor (e.g. smoking), is useful for trying to ascertain whether exposure is likely to cause specified events (e.g. lung cancer). Prospective cohort studies (which track participants forward in time) are more reliable than retrospective cohort studies. CE guidance Cohort studies should not be included within the Benefits section, unless it is not reasonable to expect higher levels of evidence.
Completer analysis
Analysis of data from only those participants who remained at the end of the study. Compare with intention to treat analysis, which uses data from all participants who enrolled.

Conference proceedings CE guidance
See unpublished evidence.

Confidence interval (CI)
The 95% confidence interval (or 95% confidence limits) would include 95% of results from studies of the same size and design in the same population. This is close but not identical to saying that the true size of the effect (never exactly known) has a 95% chance of falling within the confidence interval. If the 95% confidence interval for a relative risk (RR) or an odds ratio (OR) crosses 1, then this is taken as no evidence of an effect. The practical advantages of a confidence interval (rather than a P value) is that they present the range of likely effects. CE guidance We always try to provide 95% confidence intervals for results.

Consistency CE guidance
If two sections in Clinical Evidence address the same question then we attempt to avoid repetition of the evidence, but aim instead to provide a cross reference.

Controlled clinical trial (CCT) CE guidance
A trial in which participants are assigned to two or more different treatment groups. In Clinical Evidence, we use the term to refer to controlled trials in which treatment is assigned by a method other than random allocation. When the method of allocation is by random selection, the study is referred to as a randomised controlled trial (RCT; see below). Non-randomised controlled trials are more likely to suffer from bias than RCTs.

Controls
In a randomised controlled trial (RCT), controls refer to the participants in its comparison group. They are allocated either to placebo, no treatment, or a standard treatment.

Correlation coefficient
A measure of association that indicates the degree to which two variables change together in a linear relationship. It is represented by r, and varies between −1 and +1. When r is +1, there is a prefect positive relationship (when one variable increases, so does the other, and the proportionate difference remains constant). When r is −1 there is a perfect negative relationship (when one variable increases the other decreases, or vice versa, and the proportionate difference remains constant). This, however, does not rule out a relationship — it just excludes a linear relationship.

Crossover randomised trial
A trial in which participants receive one treatment and have outcomes measured, and then receive an alternative treatment and have outcomes measured again. The order of treatments is randomly assigned. Sometimes a period of no treatment is used before the trial starts and in
between the treatments (washout periods) to minimise interference between the treatments (carry over effects). Interpretation of the results from crossover randomised controlled trials (RCTs) can be complex.

CE guidance Crossover studies have the risk that the intervention may have an effect after it has been withdrawn, either because the washout period is not long enough or because of path dependency. A test for evidence of statistically significant heterogeneity is not sufficient to exclude clinically important heterogeneity. An effect may be important enough to affect the outcome but not large enough to be significant. Therefore, we try to only include results from crossover studies before the cross over.

Cross sectional study
A study design that involves surveying a population about an exposure, or condition, or both, at one point in time. It can be used for assessing prevalence of a condition in the population.

CE guidance Cross sectional studies should never be used for assessing causality of a treatment.

D through N

Glossary

D through N

Data pooling CE guidance
Crude summation of the raw data with no weighting, to be distinguished from meta-analysis.

Decimal places CE guidance
We always precede decimal points with an integer. Numbers needing treatment to obtain one additional beneficial outcome (NNTs) are rounded up to whole numbers e.g. an NNT of 2.6 would become 3. Numbers needing treatment to obtain one additional harmful outcome (NNHs) are rounded down to whole numbers e.g an NNH of 2.3 would become 2. For P values, we use a maximum of three noughts after the decimal: P < 0.0001. We try to report the number of decimal places up to the number of noughts in the trial population e.g 247 people, with RR 4.837 would be rounded up to 4.84. We avoid use of more than three significant figures.

Disability Adjusted Life Year (DALY) CE guidance
A method for measuring disease burden, which aims to quantify in a single figure both the quantity and quality of life lost or gained by a disease, risk factor, or treatment. The DALYs lost or gained are a function of the expected number of years spent in a particular state of health, multiplied by a coefficient determined by the disability experienced in that state (ranging from
0 [optimal health] to 1 [deaths]). Later years are discounted at a rate of 3% per year, and childhood and old age are weighted to count for less.

**Drillability CE guidance**
Refers to the ability to trace a statement from its most condensed form through to the original evidence that supports it. This requires not only the data but also all the methods used in the generation of the condensed form to be explicit and reproducible. We see it as an important component of the quality of evidence-based publications.

**E**

**Eclipsing CE guidance**
In Clinical Evidence a systematic review should be excluded (eclipsed) if, and only if, there is a review with a later search date with either identical methods, clearly superior methods, or similar methods including the same primary sources.

**Effect size (standardised mean differences)**
In the medical literature, effect size is used to refer to a variety of measures of treatment effect. In Clinical Evidence it refers to a standardised mean difference: a statistic for combining continuous variables (such as pain scores or height), from different scales, by dividing the difference between two means by an estimate of the within group standard deviation. CE guidance We avoid if possible. Standardised mean differences are very difficult for non-statisticians to interpret and combining heterogenous scales provides statistical accuracy at the expense of clinical intelligibility. We prefer results reported qualitatively to reliance on effect sizes.

**English language papers CE guidance**
See language.

**Event**
The occurrence of a dichotomous outcome that is being sought in the study (such as myocardial infarction, death, or a four-point improvement in pain score).

**Event rates CE guidance**
In determining the power of a trial the event rate is more important than the number of participants. Therefore, we provide the number of events as well as the number of participants when this is available.

**Experimental study**
A study in which the investigator studies the effect of intentionally altering one or more factors under controlled conditions.

**External validity (generalisabilty) CE guidance**
The validity of the results of a trial beyond that trial.
CE guidance A randomised controlled trial (RCT) only provides direct evidence of causality within that trial. It takes an additional logical step to apply this result more generally. However, practically it is necessary to assume that results are generalisable unless there is evidence to the contrary. If evidence is consistent across different settings and in different populations (e.g. across ages and countries) then there is evidence in favour of external validity. If there is only evidence from atypical setting (e.g. teaching hospital when most cases are seen in primary care) then one should be more sceptical about generalising the results. The Comment section should address questions of generalisability. Generalisability is not just a consequence of the entry requirements for the trial, but also depends on the population from which the trial population was drawn (see applicability).

F

Factorial design
A factorial design attempts to evaluate more than one intervention compared with control in a single trial, by means of multiple randomisations.

False negative
A person with the target condition (defined by the gold standard) who has a negative test result.

False positive
A person without the target condition (defined by the gold standard) who has a positive test result.

Fixed effects
The "fixed effects" model of meta-analysis assumes, often unreasonably, that the variability between the studies is exclusively because of a random sampling variation around a fixed effect (see random effects below).

H

Harms CE guidance
Evidence-based healthcare resources often have great difficulty in providing good quality evidence on harms. Most RCTs are not designed to assess harms adequately: the sample size is too small, the trial too short, and often information on harms is not systematically collected. Often a lot of the harms data are in the form of uncontrolled case reports. Comparing data from these series is fraught with difficulties because of different numbers receiving the intervention, different baseline risks and differential reporting. We aim to search systematically for evidence on what are considered the most important harms of an intervention. The best evidence is from a systematic review of harms data that attempts to integrate data from different sources. However, because of these difficulties and following the maxim "first one must not do harm" we accept weaker evidence in the Harms than in the Benefits section. This
can include information on whether the intervention has been either banned or withdrawn because of the risk of harms.

**Hazard ratio (HR)**

Broadly equivalent to relative risk (RR); useful when the risk is not constant with respect to time. It uses information collected at different times. The term is typically used in the context of survival over time. If the HR is 0.5 then the relative risk of dying in one group is half the risk of dying in the other group.

CE guidance If HRs are recorded in the original paper then we report these rather than calculating RR, because HRs take account of more data.

**Heterogeneity**

In the context of meta-analysis, heterogeneity means dissimilarity between studies. It can be because of the use of different statistical methods (statistical heterogeneity), or evaluation of people with different characteristics, treatments or outcomes (clinical heterogeneity). Heterogeneity may render pooling of data in meta-analysis unreliable or inappropriate.

CE guidance Finding no significant evidence of heterogeneity is not the same as finding evidence of no heterogeneity. If there are a small number of studies, heterogeneity may affect results but not be statistically significant.

**Homogeneity**

Similarity (see heterogeneity).

**Incidence**

The number of new cases of a condition occurring in a population over a specified period of time.

**Inclusion/exclusions CE guidance**

We use validated search and appraisal criteria to exclude unsuitable papers. Authors are then sent exclusion forms to provide reasons why further papers are excluded (see Literature searches).

**Intention to treat (ITT) analysis**

Analysis of data for all participants based on the group to which they were randomised and not based on the actual treatment they received.

CE guidance Where possible we report ITT results. However, different methods go under the name ITT. Therefore, it is important to state how withdrawals were handled and any potential biases, e.g. the implication of carrying last result recorded forward will depend on the natural history of the condition.
Jadad scale CE guidance
See Literature searches

Language CE guidance
We aim to include all identified relevant papers irrespective of language. If we have not been able to translate a paper in time for publication of the topic then we state this in the Comment section.

Likelihood ratio
The ratio of the probability that an individual with the target condition has a specified test result to the probability that an individual without the target condition has the same specified test result.

Meta-analysis
A statistical technique that summarises the results of several studies in a single weighted estimate, in which more weight is given to results of studies with more events and sometimes to studies of higher quality.
CE guidance We use meta-analysis to refer to the quantitative methods (usually involving weighting) used to integrate data from trials. This is logically distinct from a systematic review, which is defined by an explicitly systematic search and appraisal of the literature. It is also distinct from data pooling, which is based purely on the raw data. If an unpublished meta-analysis is included in Clinical Evidence then the methods should be made explicit, which we do through publication on the Clinical Evidence website. It should be noted that the statistical package RevMan assumes that all outcomes are adverse and therefore if RevMan states that the results for a beneficial outcome favour control this means the beneficial outcome is more likely with the experimental intervention.

Morbidity
Rate of illness but not death.

Mortality
Rate of death.

Negative likelihood ratio (NLR)
The ratio of the probability that an individual with the target condition has a negative test result to the probability that an individual without the target condition has a negative test result. This is the same as the ratio (1-sensitivity/specificity).
**Negative predictive value (NPV)**
The chance of not having a disease given a negative test result (not to be confused with specificity, which is the other way round.)

**Glossary**

**Negative statements CE guidance**
At what stage does no evidence of an effect become evidence of no effect? If confidence intervals are available then we should aim to indicate in words the potential size of effect they encompass. If a result is not significant we try and state if the confidence intervals include the possibility of a large effect (e.g. "The RCT found no significant effect but included the possibility of a large harm/benefit/harm or benefit"). The exact wording depends on the mean result and the width of the confidence intervals.

**Non-systematic review**
A review or meta-analysis that either did not perform a comprehensive search of the literature and contains only a selection of studies on a clinical question, or did not state its methods for searching and appraising the studies it contains.

**Not significant/non-significant (NS)**
In *Clinical Evidence*, not significant means that the observed difference, or a larger difference, could have arisen by chance with a probability of more than 1/20 (i.e. 5%), assuming that there is no underlying difference. This is not the same as saying there is no effect, just that this experiment does not provide convincing evidence of an effect. This could be because the trial was not powered to detect an effect that does exist, because there was no effect, or because of the play of chance. If there is a potentially clinically important difference that is not statistically significant then do not say there was a non-significant trend. Alternative phrases to describe this type of uncertainty include, "Fewer people died after taking treatment x but the difference was not significant" or "The difference was not significant but the confidence intervals covered the possibility of a large beneficial effect" or even, "The difference did not quite reach significance."
**Number needed to harm (NNH)**
One measure of treatment harm. It is the average number of people from a defined population you would need to treat with a specific intervention for a given period of time to cause one additional adverse outcome. NNH can be calculated as 1/ARI. In Clinical Evidence, these are usually rounded downwards.

**Number needed to treat (NNT)**
One measure of treatment effectiveness. It is the average number of people who need to be treated with a specific intervention for a given period of time to prevent one additional adverse outcome or achieve one additional beneficial outcome. NNT can be calculated as 1/ARR. In Clinical Evidence, NNTs are usually rounded upwards.

CE guidance

1. NNTs are easy to interpret, but they can only be applied at a given level of baseline risk
2. How do we calculate NNTs from meta-analysis data? The odds ratio (OR) (and 95% CI) with the AR in the control group can be used to generate absolute risk (AR) in the intervention group and from there to the NNT. This is a better measure than using the pooled data, which only uses trial size (not variance) and does not weight results (e.g. by trial quality). As people can not be treated as fractions, we round NNTs up and numbers needed to harm (NNHs) down to the largest absolute figure. This provides a conservative estimate of effect (it is most inaccurate for small numbers)
3. NNTs should only be provided for significant effects because of the difficulty of interpreting the confidence intervals for non-significant results. Non-significant confidence intervals go from an NNT to an NNH by crossing infinity rather than zero

**NNT for a meta-analysis**
Absolute measures are useful at describing the effort required to obtain a benefit, but are limited because they are influenced by both the treatment and also by the baseline risk of the individual. If a meta-analysis includes individuals with a range of baseline risks, then no single NNT will be applicable to the people in that meta-analysis, but a single relative measure (odds ratio or relative risk) may be applicable if there is no heterogeneity. In Clinical Evidence, an NNT is provided for meta-analysis, based on a combination of the summary odds ratio (OR) and the mean baseline risk observed in average of the control groups.

0

**Observational studies CE guidance**
We do not include observational studies in the Benefits section unless good RCTs are unavailable. Observational studies may be included in the Harms section or in the Comment. Observational studies are the most appropriate form of evidence for the Prognosis, Aetiology, and Incidence/Prevalence sections. The minimum data set and methods requirements for observational studies have not been finalised. However, we always indicate what kind of observational study, whether case series, case control, prospective or retrospective cohort study.
Odds
The odds of an event happening is defined as the probability that an event will occur, expressed as a proportion of the probability that the event will not occur.

Odds ratio (OR)
One measure of treatment effectiveness. It is the odds of an event happening in the experimental group expressed as a proportion of the odds of an event happening in the control group. The closer the OR is to one, the smaller the difference in effect between the experimental intervention and the control intervention. If the OR is greater (or less) than one, then the effects of the treatment are more (or less) than those of the control treatment. Note that the effects being measured may be adverse (e.g. death or disability) or desirable (e.g. survival). When events are rare the OR is analogous to the relative risk (RR), but as event rates increase the OR and RR diverge.
CE guidance The ratio of events to non-events in the intervention group over the ratio of events to non-events in the control group. In Clinical Evidence we try to provide relative risks in preference to odds ratios.

Odds reduction
The complement of odds ratio (1-OR), similar to the relative risk reduction (RRR) when events are rare.

Open label trial
A trial in which both participant and assessor are aware of the intervention allocated.

Outcomes CE guidance
In Clinical Evidence we always aim to use outcomes that matter to patients and their carers. This generally means mortality, morbidity, quality of life, ability to work, pain, etc. Laboratory outcomes are avoided if possible. Even if there is a strong relationship between a laboratory outcome marker and a clinical outcome it is not automatic that it will hold under new conditions. Outcomes that are markers for clinically important patient centred outcomes are often called surrogate outcomes (e.g. ALT concentrations are a proxy for liver damage following paracetamol overdose). We only use surrogate outcomes in Clinical Evidence if patient centred outcomes are not available and a strong and consistent relationship between the surrogate outcome and patient centred outcomes has been established.

P

Personal communication CE guidance
In the Comments section of Clinical Evidence we include evidence from personal communication if it is sufficiently important. In the Benefits section, we aim to include only the evidence that has been published in peer reviewed journals.

PICOt CE guidance
Population, intervention, comparison, and outcome, all with a time element (PICOt). The
current reporting requirements of systematic reviews are: how many RCTs, how many participants in each, comparing what with what, in what type of people, with what results. Each variable needs a temporal element, (how old are the participants, how long is the treatment given for, when is the outcome measured). In the future, we hoping to have a brief description in the text with full details accessible from the website.

**Placebo**
A substance given in the control group of a clinical trial, which is ideally identical in appearance and taste or feel to the experimental treatment and believed to lack any disease specific effects. In the context of non-pharmacological interventions, placebo is usually referred to as sham treatments.
CE guidance Placebo is not the same as giving no treatment and can induce real physiological changes. Whether it is appropriate to compare the experimental with placebo or no treatment depends on the question being asked. Where possible we report on the specific intervention given as a placebo. We include, if available, information is available on whether participants or clinicians could distinguish between placebo and the intervention.

**Positive likelihood ratio (LR+)**
The ratio of the probability that an individual with the target condition has a positive test result to the probability that an individual without the target condition has a positive test result. This is the same as the ratio (sensitivity/1-specificity).

**Positive predictive value (PPV)**
The chance of having a disease given a positive test result (not to be confused with sensitivity, which is the other way round.

**Power**
A study has adequate power if it can reliably detect a clinically important difference (i.e. between two treatments) if one actually exists. The power of a study is increased when it includes more events or when its measurement of outcomes is more precise.
CE guidance We do not generally include power calculations, but prefer to provide confidence intervals (CIs) and leave it to readers to say if this covers a clinically significant difference. If no CIs are available a power calculation can be included assuming it is adequately explained.

**Pragmatic study**
An RCT designed to provide results that are directly applicable to normal practice (compared with explanatory trials that are intended to clarify efficacy under ideal conditions). Pragmatic RCTs recruit a population that is representative of those who are normally treated, allow normal compliance with instructions (by avoiding incentives and by using oral instructions with advice to follow manufacturers instructions), and analyse results by "intention to treat" rather than by "on treatment" methods.

**Prevalence**
The proportion of people with a finding or disease in a given population at a given time.
Protocols CE guidance
If there is no recent systematic review (search date within the last 3 years) we report recent protocols (last 2 years) identified by our search. The information specialists send to Clinical Evidence authors all York and Cochrane protocols identified by our search.

Proxy outcomes CE guidance
See surrogate outcomes.

Publication bias
Occurs when the likelihood of a study being published varies with the results it finds. Usually, this occurs when studies that find a significant effect are more likely to be published than studies that do not find a significant effect, so making it appear from surveys of the published literature that treatments are more effective than is truly the case.

CE guidance Can occur through both preference for significant (positive) results by journals and selective releasing of results by interested parties. A systematic review can try and detect publication bias by a forest plot of size of trial against results. This assumes that larger trials are more likely to be published irrespective of the result. If a systematic review finds evidence of publication bias this should be reported. Often publication bias takes the form of slower or less prominent publication of trials with less interesting results.

P value
The probability that an observed or greater difference occurred by chance, if it is assumed that there is in fact no real difference between the effects of the interventions. If this probability is less than 1/20 (which is when the P value is less than 0.05), then the result is conventionally regarded as being "statistically significant".

Q

Quality Adjusted Life Year (QALY) CE guidance
A method for comparing health outcomes, which assigns to each year of life a weight from 1 (perfect health) to 0 (state judged equivalent to death) dependent on the individual's health related quality of life during that year. A total score of years multiplied by weight can then be compared across different interventions. There is disagreement about the best methods for measuring health-related quality of life.

Quality Control CE guidance
At Clinical Evidence we aim to have explicit and transparent methods to formulate the most clinically relevant questions, selecting the most relevant outcomes, and searching, appraising and synthesising the medical literature.

Quality of evidence CE guidance
See best evidence
Quasi randomised
A trial using a method of allocating participants to different forms of care that is not truly random; for example, allocation by date of birth, day of the week, medical record number, month of the year, or the order in which participants are included in the study (e.g. alternation).

Random effects
The "random effects" model assumes a different underlying effect for each study and takes this into consideration as an additional source of variation, which leads to somewhat wider confidence intervals than the fixed effects model. Effects are assumed to be randomly distributed, and the central point of this distribution is the focus of the combined effect estimate.

Clinical evidence
We prefer the random effects model because the fixed effects model is appropriate only when there is no heterogeneity—in which case results will be very similar. A random effects model does not remove the effects of heterogeneity, which should be explained by differences in trial methods and populations.

Randomised controlled trial (RCT)
A trial in which participants are randomly assigned to two or more groups: at least one (the experimental group) receiving an intervention that is being tested and an other (the comparison or control group) receiving an alternative treatment or placebo. This design allows assessment of the relative effects of interventions.

Clinical evidence
Clinical evidence is built upon RCTs and systematic reviews of RCTs.
Regression analysis
Given data on a dependent variable and one or more independent variables, regression analysis involves finding the "best" mathematical model to describe or predict the dependent variable as a function of the independent variable(s). There are several regression models that suit different needs. Common forms are linear, logistic, and proportional hazards.

Relative risk (RR)
The number of times more likely (RR > 1) or less likely (RR < 1) an event is to happen in one group compared with another. It is the ratio of the absolute risk (AR) for each group. It is analogous to the odds ratio (OR) when events are rare.
CE guidance We define relative risk as the absolute risk (AR) in the intervention group divided by the AR in the control group. It is to be distinguished from odds ratio (OR) which is the ratio of events over non-events in the intervention group over the ratio of events over non-events in the control group. In the USA, odds ratios are sometimes known as rate ratios or relative risks.

Relative risk increase (RRI)
The proportional increase in risk between experimental and control participants in a trial.

Relative risk reduction (RRR)
The proportional reduction in risk between experimental and control participants in a trial. It is the complement of the relative risk (1-RR).

S

Searches CE guidance
See Literature searches

Sensitivity
The chance of having a positive test result given that you have a disease (not to be confused with positive predictive value [PPV], which is the other way around).

Sensitivity analysis
Analysis to test if results from meta-analysis are sensitive to restrictions on the data included. Common examples are large trials only, higher quality trials only, and more recent trials only. If results are consistent this provides stronger evidence of an effect and of generalisability.

Sham treatment
An intervention given in the control group of a clinical trial, which is ideally identical in appearance and feel to the experimental treatment and believed to lack any disease specific effects (e.g. detuned ultrasound or random biofeedback).
CE guidance Placebo is used for pills, whereas sham treatment is used for devices, psychological, and physical treatments. We always try and provide information on the specific sham treatment regimen.
**Significant**
By convention, taken to mean statistically significant at the 5% level. This is the same as a 95% confidence interval not including the value corresponding to no effect.

**Specificity**
The chance of having a negative test result given that you do not have a disease (not to be confused with negative predictive value [NPV], which is the other way around.

**Standardised mean difference (SMD)**
A measure of effect size used when outcomes are continuous (such as height, weight, or symptom scores) rather than dichotomous (such as death or myocardial infarction). The mean differences in outcome between the groups being studied are standardised to account for differences in scoring methods (such as pain scores). The measure is a ratio; therefore, it has no units.

CE guidance We avoid using SMDs if possible. SMD are very difficult for non-statisticians to interpret and combining heterogenous scales provides statistical accuracy at the expense of clinical intelligibility. We prefer results reported qualitatively to reliance on effect sizes, although we recognise that this may not always be practical.

**Statistically significant**
Means that the findings of a study are unlikely to have arisen because of chance. Significance at the commonly cited 5% level (P < 0.05) means that the observed difference or greater difference would occur by chance in only 1/20 similar cases. Where the word "significant" or "significance" is used without qualification in the text, it is being used in this statistical sense.

**Subgroup analysis**
Analysis of a part of the trial/meta-analysis population in which it is thought the effect may differ from the mean effect.

CE guidance Subgroup analysis should always be listed as such and generally only prespecified subgroup analysis should be included. Otherwise, they provide weak evidence and are more suited for hypothesis generation. If many tests are done on the same data this increases the chance of spurious correlation and some kind of correction is needed (e.g. Bonferroni). Given independent data, and no underlying effect, 1 time in 20 a significant result would be expected by chance.

**Surrogate outcomes CE guidance**
Outcomes not directly of importance to patients and their carers but predictive of patient centred outcomes).

**Systematic review**
A review in which specified and appropriate methods have been used to identify, appraise, and summarise studies addressing a defined question. It can, but need not, involve meta-analysis). In Clinical Evidence, the term systematic review refers to a systematic review of RCTs unless specified otherwise.
CE guidance The present requirements for reporting systematic reviews are search date, number of trials of the relevant option, number of trials that perform the appropriate comparisons, comparisons, details on the type of people, follow up period, and quantified results if available.

T

Trend CE guidance
In Clinical Evidence, we aim to avoid saying there was a non-significant trend. Alternatives include, "fewer people died after taking treatment x but the difference was not significant" or "The difference was not significant but the confidence intervals covered the possibility of a large beneficial effect" or even, "The difference did not quite reach significance."

True negative
A person without the target condition (defined by a gold standard) who has a negative test result.

True positive
A person with the target condition (defined by a gold standard) who also has a positive test result.

U

Unpublished evidence CE guidance
Clinical Evidence is based on published peer reviewed evidence. Unpublished conference proceedings will not be included in the Benefits section (except as part of a published systematic review), but may be included in the Comment section. Sometimes Clinical Evidence includes unpublished meta-analysis or, more often, data pooling performed by Clinical Evidence authors or editors. We clearly indicate as such and full details of workings will be available on the website. Results from unpublished meta-analysis will always be taken as subsequent to revision by proper published analysis.

V

Validity
The soundness or rigour of a study. A study is internally valid if the way it is designed and carried out means that the results are unbiased and it gives you an accurate estimate of the effect that is being measured. A study is externally valid if its results are applicable to people encountered in regular clinical practice.

W

Weighted mean difference (WMD)
A measure of effect size used when outcomes are continuous (such as symptom scores or
height) rather than dichotomous (such as death or myocardial infarction). The mean differences in outcome between the groups being studied are weighted to account for different sample sizes and differing precision between studies. The WMD is an absolute figure and so takes the units of the original outcome measure.

CE guidance A continuous outcome measure, similar to standardised mean differences but based on one scale so in the real units of that scale. Ideally should be replaced by a discrete outcome and a relative risk; however, we use WMD if this is not possible.