How to make sense of a Cochrane systematic review

What is the point of a systematic review?

More and more papers are published in medical journals every day, so how do you decide which ones to read and, having read a paper, how do you decide whether to change your practice as a result of what you have read? Perhaps the paper was atypical in some way. What does the other research on the topic say?

The purpose of systematic reviews is to summarise all the available, high-quality evidence that can be found on a particular topic. A narrative review, in which an expert can cite a selection of papers that support a particular viewpoint, says very little about the papers that do not. In contrast, a systematic review involves a search for all available literature, whatever the findings may be.

Systematic reviews start with a well-defined clinical question, and aim to identify, appraise, synthesise and then apply all the available good-quality evidence that can be found (published or unpublished) that is relevant to the question. In particular, Cochrane systematic reviews have to meet a defined set of quality standards and the authors and editors set out to make them the best around. They are the current gold standard in the systematic review field.

The Cochrane Collaboration

The Cochrane Collaboration is an international group which is now 20 years old. The collaboration depends upon the voluntary contribution of thousands of authors and is supported by editorial bases and methodologists. In the UK, these bases are supported by funding from the National Institute of Health Research (NIHR), and any funding for editorial bases must be free from commercial interests. Review groups are divided up into areas of clinical interest and, within the respiratory field, there are groups for lung cancer (based in France), acute respiratory infections (based in Australia), cystic fibrosis (based in the UK) and airways (based in the UK with a satellite in Australia).

The Cochrane Airways Group has 875 authors and 18 editors who volunteer their services from all over the world and have prepared and maintain 278 systematic reviews published electronically on the Cochrane Database of Systematic Reviews. We have addressed a wide variety of clinical interventions for asthma, chronic obstructive
pulmonary disease (COPD) and other respiratory diseases.

Reading a systematic review

Let’s consider a real-life clinical problem (as set out in the box below), to see how a Cochrane review [1] is constructed and how you can find your way around it.

A mum brings her 5-year-old son back to the clinic for review following a recent exacerbation of his asthma, for which he was treated in the emergency department with nebulised salbutamol and a short course of oral steroids. He made a full recovery and is back on his usual maintenance treatment. The mum asks if she should buy a nebuliser so that she can treat her son at home next time he has an exacerbation.

In order to find the evidence to answer her question, we need to start by refining the question using a PICO framework:

- **Population**
- **Intervention**
- **Comparison**
- **Outcomes**
- **Study design**

The PICO framework is the structure used in all Cochrane systematic reviews and, once the question that we want to answer is decided, we can use this to define search terms to find the evidence that is available (not just papers published in the English language). In this case the question might run something as follows.

In randomised trials of children with acute asthma, how does a nebuliser compare with a metered-dose inhaler (pMDI) and spacer for the delivery of \( \beta_2 \)-agonists in an acute asthma exacerbation, in terms of duration spent in the emergency department and the risk of being admitted to hospital?

In practice, you can use this question to search for a systematic review which might give you the answer, and a screenshot of a PubMed search on “spacers and acute asthma” is shown in figure 1. If you click on the hyperlink for the review in PubMed it will take you to the abstract of the Cochrane review which is held in PubMed. The first part of the abstract is shown in figure 2.

In the top right corner of the abstract screen there is a link to the full text version, which is held on the Cochrane Database of Systematic reviews (www.thecochranelibrary.org). The Cochrane Database is freely available to anyone in Denmark, Finland, Ireland, Norway, Spain, Sweden and the UK through funded national provision. The Cochrane Collaboration and Wiley also provide free one-click access in Albania, Bosnia, Kosovo, Macedonia, Moldova and Ukraine. In the rest of Europe subscriptions are to institutions or individuals.

Protocol for systematic review

If a systematic review to answer the question is found, you might want to check whether it was prepared using methods that were defined in a pre-published protocol. Protocols specify full details of the methods that will be used to carry out the review and also set the review in context. The PICO framework is fleshed out with enough details to define which trials will be included and which will not.

Not all systematic reviews have a protocol that is fully published in advance, but this is a hallmark of Cochrane reviews and is a good way of trying to reduce bias in the process of carrying out the review itself. Otherwise, the
reviewers might find a study with particularly good results and adjust their inclusion criteria or their methods in order to be able to get these promising results into their review! Moreover the Cochrane protocols are externally peer reviewed to try to ensure that the question to be answered and the proposed methods are appropriate and of a high standard.

Check for
Did the systematic review follow methods that were clearly set out in a pre-published protocol, including a clear description of the search strategy to be used in the methods section? For a Cochrane review the authors should document any changes in methods between the protocol and the final review. This can be found at the end of the review just before the references.

A comprehensive search of the electronic literature databases
One of the biggest headaches for systematic reviewers is publication bias. This is when trials with statistically significant results are published more quickly (and in more prominent journals), than those which fail to achieve a significant result. There are a variety of reasons why this occurs and these include the fact that trials are more interesting to readers if they show a significant treatment effect. However, if you consider the trials that had a more positive effect, then the treatment effect derived from these trials will give an over-optimistic estimate of how well the treatment works. Trial registries (such as clinicaltrials.gov) have been introduced to try to reduce the risk that trials are carried out without their results ever being published.

For this reason, the search for studies for inclusion in a Cochrane review on asthma starts with the Airways Group register of controlled clinical trials. This is regularly updated from the results of systematic searches of multiple databases including CENTRAL, MEDLINE, Embase and CINAHL by the trials search co-ordinator, and is supplemented by manual searching of conference abstracts. The trials search co-ordinator helps reviewers to design the search strategy for the review and will also conduct the search.

The search should not be limited by language, type of publication or by date, unless the intervention of interest did not exist prior to a particular date. We also search trial registries, such as clinicaltrials.gov, and manufacturers’ websites for reports of trials that have not been published as journal papers.

Check for
Was there a search for results from unpublished as well as published trials? This minimises the risk of publication bias in which trials with the most promising results appear more quickly in the most accessible journals.

How was the decision made about which trials to include?
The literature search results are sifted by two separate systematic reviewers. On the basis of the title and abstract, the reviewers usually identify three groups of references: clearly not relevant; likely to be included in the review; and unclear. Full text articles of the first and second groups are needed in order to make a final decision on inclusion for each paper. The reviewers do their best to obtain translations of any possibly relevant papers published in languages other than English.
Check for

Were the included trials selected by two authors independently?

How was the risk of bias assessed for the included studies?

Cochrane systematic reviews of interventions are usually focussed primarily on randomised trials in order to reduce the risks of biases that are present in observational study designs, but even so there are still risks of bias that may be present in each trial (see the box below).

The results of the bias assessment for each outcome in each trial are reported in a risk of bias table in Cochrane systematic reviews, and these may be summarised as a figure to give an overview of the risks of bias across the whole body of evidence (see figure 3 for an example). The assessed risk of bias for each outcome is also used to make the GRADE recommendations about how confident the reviewers are in the results found for each outcome.

Check for

Were the included trials assessed for bias and were the risks of bias assessed by two authors independently?

What were the characteristics of the included studies?

Some information is needed about the participants in the trials and the way the treatments were delivered, to judge how well they apply to the patients we see in our own practice. We also need to ask whether or not it makes clinical sense to calculate an average treatment effect for outcomes across this group of trials.

The characteristics of the participants, interventions, comparisons and outcomes are extracted and tabulated in the review. In the current example, the majority of trials were carried out in emergency departments where salbutamol was given by spacer to some adults or children with acute asthma and compared with salbutamol delivered by a jet nebuliser. However, one of the problems was a wide variation in the dose of salbutamol that was delivered in each trial. Moreover, the different spacers and nebulisers also impact on the amount of salbutamol that reaches the lungs, as does the severity of the asthma exacerbation. All the studies excluded people with life-threatening asthma, so the results cannot be applied to this group of people with very severe exacerbations.

To a certain extent, this problem was overcome in some of the trials by using multiple treatments titrated against the clinical response of the patient, so the main focus of the Cochrane review is on trials that compare multiple treatments with salbutamol via a spacer or nebuliser, not just a single treatment. For the pMDI and spacer, the most common dose was 4–6 puffs (delivered one at a time and then each one inhaled separately from the spacer). The set of 4–6 puffs through the spacer or nebulised dose of salbutamol was repeated two or three times at intervals of 20–30 minutes, until there was a satisfactory response to treatment or the patient was escalated to more intensive treatment.

Check for

Were the study methods well enough described to know how to apply the treatments in the trials to our own patients? Who was excluded from the trials? In this case, patients with life-threatening asthma were excluded, so we cannot apply the results of the review to such patients.
Extract the outcome data that was pre-specified in the protocol

The next job for the reviewers is to go through each trial report and extract the data from that trial on the outcomes specified in their protocol. However, trials may measure dozens of different outcomes, and there will not be room to report all of these in the journal paper. In particular, the reporting of serious adverse events is often patchy in papers, so it is important to check the trial registry sites for reports of adverse events, and to look for reports on manufacturers’ websites as well.

If there are no reports found in any of these sources for outcomes specified in the review protocol, the next stage is to try to find out from the trial authors (or the trial protocol) whether the outcome was measured, and what the result was. This is not always possible, especially for older trials. However, if there is a substantial proportion of trials that do not report an outcome, this will reduce our confidence in the results for this outcome.

Check for

Did the authors assess how much of the trial data that was measured was actually reported in the trials and available to include in the systematic review?

Combine the data in a meta-analysis (using risk ratios or odds ratios)

Meta-analysis is a statistical technique that is used to calculate the weighted average of the results for each outcome in the included trials. This may be presented as the mean difference between the trial arms (for example, for peak flow, forced expiratory volume in 1 s or quality of life, which can be measured in a variety of different ways using percentage predicted values or absolute values, and might be measured as change from baseline or final scores). More recently, the results of adjusted analyses are more commonly presented using ANOVA or ANCOVA analyses, which are presented as a difference between groups with an associated confidence interval. Since these adjusted analyses take into account the difference in specified covariates between the trial arms, they usually produce a narrower confidence interval than either final readings or changes from baseline.

Figure 3

Risk of bias summary for Cochrane review comparing pMDI and spacer to nebulisers to deliver β₂-agonists to adults and children with acute asthma. The red circle indicates high risk of bias, the green circle low risk and the yellow circle indicates that the risk is unclear as judged by the authors of the Cochrane Review. Figure reproduced and modified from [1] with the publisher’s permission.

There are some outcomes that are measured as binary or dichotomous data; this might include the number of participants in
each arm who died, or who suffered a serious adverse event, or who were admitted to hospital (as shown in a forest plot in figure 4).

Forest plots are graphical displays of data from each trial that contributes data to a particular outcome, with each trial in a separate row. The given example examines trials which contribute data on hospital admissions. The intervention in this case was pMDI and spacer; so, after the first column which gives an identifier for each trial, the next column shows the number of people who were admitted using pMDI and spacer (labelled as “Events”). This is followed by the total number of people randomised to using a spacer in that trial (labelled as “Total”). The fourth and fifth columns give the same information for the nebuliser group. Then there is a column showing the weight given to that trial and the risk ratio is reported as numbers and then shown as a box (the point estimate) and horizontal whiskers that show the width of the 95% confidence interval for each trial. Finally the diamond under the adult trials is the weighted average of all the trials, and the whole width of the diamond represents its 95% CI, which is where we are 95% sure that the true average of the set of trials is to be found.

**Figure 4**
Forest plot showing the number of adults (at the top) and children (lower down) who were admitted to hospital after treatment in the emergency department for an asthma exacerbation. Figure reproduced and modified from [1] with the publisher’s permission.

Report the weighted average of the treatment effects from the trials

This type of event data can be analysed as risk ratios (as shown in figure 4), or odds ratios which can both then be converted into absolute treatment differences or numbers needed to treat (NNT). In the case of adults in the forest plot, the risk ratio is very close to 1 (which would mean the same number of admissions on pMDI with a spacer as with a nebuliser), so a significant difference between using a spacer or a nebuliser in adults has not been shown.

However, it cannot be said that the two delivery methods are identical, since the 95% CI stretches from a lower risk ratio of 0.61 to an upper risk ratio of 1.43. So we can only say that we are 95% sure that the average risk ratio from these trials is somewhere between 0.61 and 1.43 adults being admitted on pMDI and nebuliser for every 100 admitted on nebuliser. This is potentially a large difference either way, so the bottom line is that we do not know which the better delivery method in adults is.

In the children, the risk ratio is 0.71, so for every 100 children admitted on nebuliser we would expect 71 to be admitted using pMDI and spacer. In this case, the 95% CI just crosses the vertical line where the risk ratio is one. We have therefore not disproved the null hypothesis, and there may be no difference between spacers and nebulisers in children. However it is still important to note that the average risk ratio in children has a 95% CI that runs from 0.47 to 1.08, so for every 100 children admitted on nebuliser we expect somewhere between 47 and 108 to be admitted using a spacer. Therefore in the worst case a spacer might increase the risk of admission by 8% (relative risk increase), whereas in the best-case scenario, a spacer might halve the risk of admission (a 53% relative risk reduction).

**Check for**
Look at the labels at the bottom of the forest plot to confirm that the children with acute asthma treated with pMDI and spacer were less likely to be admitted to hospital.

**Absolute differences and numbers needed to treat**
The possibility of a 29% reduction in the risk of admission to hospital by using spacers.
rather than nebulisers in children with acute asthma sounds quite impressive but, in order to make sense of this, we also need to know that this is a 29% reduction in comparison to the risk of admission on nebuliser. In other words, we need to check how many children were admitted to hospital in the nebuliser arms of these trials. If we look at the subtotal line under the trials for children in figure 4, we see that 40 children were admitted out of 363 in these trials. This is an average risk of admission of 11% using nebulisers.

The Cates plot (www.nntonline.net/visualrx/cates_plot/) shown in figure 5 demonstrates this average rate using a nebuliser to deliver \( \beta_2 \)-agonists in the trials on children. The 100 faces represent children with acute asthma who were all treated with nebuliser; 11 were admitted to hospital (the red faces) and 89 did not need admission (the green faces).

In contrast, the Cates plot in figure 6 shows what we would expect to happen if all 100 children had been treated with pMDI and spacer. There are still 89 children who would not be admitted to hospital, whichever way they were treated. However, instead of 11 red faces there are now eight red faces representing expected admissions if all 100 children were treated with pMDI and spacer, and the three yellow faces are three children who would have been admitted using nebuliser, but we would not expect them to need admission using a spacer.

There is uncertainty around these results due to chance, so if all 100 children were treated with nebuliser there would be 11 requiring admission to hospital, whereas if all 100 children had been treated with pMDI and spacer we expect three fewer to need admission, because we can be 95% sure that the number requiring admission lies between six fewer and one more.

**Check for**

Are the results of the meta-analysis presented as both a relative measure (risk ratio) and an absolute difference (for example number needed to treat or absolute risk reduction).

**Duration of treatment in the emergency department**

Does using a spacer versus a nebuliser make any difference to how long children spend in the emergency department? This is a different type of outcome that is measured on a continuous scale (in this case, minutes); so the randomised trials compared the mean duration on each treatment and reported a mean difference between the children given pMDI and spacer in comparison with nebuliser. This is calculated from the mean and standard deviation of the time in the

**Figure 5**

For every 100 children treated with nebuliser for acute asthma, there were 11 children admitted to hospital.

**Figure 6**

If all 100 children were treated with pMDI and spacer we expect that 8 would be admitted to hospital (95% CI 5–12).
emergency department for each trial arm, as shown in figure 7. There were only three trials in children that reported this outcome but, even so, each showed a statistically significant reduction in the time spent in the emergency department when using a pMDI and spacer rather than a nebuliser. This is in contrast with the adult trials that did not show any significant difference.

The overall average difference was a reduction of 33 minutes using pMDI and spacer, with a 95% CI of 24–43 minutes reduction. However, there was variation (heterogeneity) between the results from the three trials in children. The mean reduction in each trial varied from 26 to 40 minutes. On closer inspection, it appears that the three trials all showed one-third less time in the emergency department for the children treated with pMDI and spacer and all the trials showed a statistically significant difference. For this reason, the statistical variation between the results of the trials may not be of much clinical importance.

It should however be noted that these three trials in children did not use a double dummy design, whereas the trials in adults did. The children were randomised to either pMDI and spacer or nebuliser, whereas the adults were treated with both devices and were randomised to receive salbutamol through one and a placebo through the other. Since it generally takes longer to deliver treatment through a nebuliser, and since the trials used multiple treatments that were titrated to the patient response, this may have a bearing on the difference between the results in adults and children. Although the test for subgroup differences was statistically significant there are many possible reasons for the difference, other than the difference in age of the participants.

Check for
Comparison between subgroups is indirect, so was it interpreted cautiously? Subgroup comparisons are not protected by randomisation and are subject to differences between the population and the way that the interventions are delivered in each subgroup.

Putting it all together
One of the recent developments in Cochrane reviews is the addition of a summary of findings table (table 1).

Table 1 summarises the evidence available for each outcome in a separate row. The treatment effect is shown both as an absolute difference and as a relative measure (for the dichotomous outcomes such as hospital admission). Then the number of trials and participants that contributed to the outcome is shown and followed by a grading of the confidence that the reviewers judged to be appropriate for each outcome.

Table 1 summarises the evidence available for each outcome in a separate row. The treatment effect is shown both as an absolute difference and as a relative measure (for the dichotomous outcomes such as hospital admission). Then the number of trials and participants that contributed to the outcome is shown and followed by a grading of the confidence that the reviewers judged to be appropriate for each outcome.

Figure 7
Forest plot showing duration in the emergency department (in minutes). Figure reproduced and modified from [1] with the publisher’s permission.
Table 1 Summary of findings for children: multiple treatment of β2-agonist via spacer (chamber) compared with nebuliser for children with acute asthma

| Outcomes | Illustrative comparative risks\(^{a}\) (95% CI) | Relative effect (95% CI) | No of Participants (studies) | Quality of the evidence (GRADE) | Comments |
|----------|-----------------------------------------------|--------------------------|-----------------------------|-------------------------------|----------|
| Hospital admission | 110 per 1000 | 78 per 1000 (52 to 119) | RR 0.71 (0.47 to 1.08) | 757 (9 studies) | ++, Low\(^{b}\) | Large increases in the proportion of children admitted to hospital on spacer in comparison to nebuliser are ruled out by this 95% confidence interval. |
| Duration in emergency department min | The mean duration in emergency department (minutes) in the control groups was 103 minutes | The mean duration in emergency department (minutes) in the intervention groups was 33 minutes shorter (43 minutes shorter to 24 minutes shorter) | 396 (3 studies) | +++ Moderate\(^{c}\) | There was a consistent direction of shortening of time in the emergency department in all three studies and, although the size of this effect varied between studies (I\(^{2}\) = 66%), we felt that the mean difference was important in all studies. |
| Final rise in FEV\(_{1}\) % pred | Control groups: 27% predicted at baseline | Intervention groups: 0.92% higher (4.96% lower to 6.79% higher) | 48 (2 studies) | ++, Low\(^{c}\) |
| Mean rise in pulse rate % baseline | Control groups: 7% rise from baseline | Intervention groups: 5.62% lower (7.52% to 3.72% lower) | 670 (9 studies) | +++ Moderate\(^{c}\) |
| Number of participants developing tremor | 142 per 1000 | 91 per 1000 (62 to 135) | RR 0.64 (0.44 to 0.95) | 254 (4 studies) | +++ Moderate\(^{c}\) |

\(^{a}\): The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes, the corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI); \(^{b}\): Mostly open label studies; \(^{c}\): Wide confidence intervals. Patient or population: children with acute asthma; settings: community or emergency department; intervention: multiple treatments with β2-agonist via spacer (chamber); comparison: multiple treatments with β2-agonist via nebuliser. GRADE Working Group grades of evidence are as follows: high quality: further research is very unlikely to change our confidence in the estimate of effect; moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low quality: 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 quality: we are very uncertain about the estimate.
Summary of the points to check when reading a systematic review

- **Protocol**
  - Was there a pre-published protocol which defined the methods to be used in advance? Did the reviewers document whether there were any changes from the protocol?

- **Search strategy**
  - Did the searches include published and unpublished articles? Were articles in languages other than English included in the review?

- **Trial selection**
  - Was this carried out independently by two review authors?

- **Characteristics of included studies**
  - Were the characteristics of the trial participants and methods well enough described to assess how well the findings of the trials would apply to our patients? Were the interventions well enough described to apply them in our practice?

- **Assessment of risks of bias**
  - Was this carried out independently by two review authors? Were the risks of bias reflected in the results and conclusions of the review abstract? Was it clear how many of the available trials and participants contributed data for each outcome?

- **Sub-group analysis**
  - Was any sub-group analysis reported cautiously, acknowledging that any differences found between sub-groups are not protected by randomisation?

- **Reporting of results**
  - Were the results reported as relative measures (such as risk ratio) and also as absolute differences (such as risk difference or NNT)?

- **Confidence in the results**
  - Was there a GRADE assessment for the main outcomes of the review (for example as presented in a summary of findings table)?

The evidence from trials in the emergency department has not shown important superiority of nebulisers over spacers in children with acute asthma. There is no reason for the mother to spend her money on a portable nebuliser, when a pMDI and spacer could do just as well. However, she will need careful instruction with an agreed management plan for how to treat her son if his asthma flares up (how many puffs of salbutamol to give him and when the treatment can be repeated). A course of oral prednisolone to take at home might be part of the management plan, and clear agreement about when to call for further help.

The review used as the example in this article was published as a Cochrane Review in the *Cochrane Database of Systematic Reviews* 2013, Issue 9. Cochrane Reviews are regularly updated as new evidence emerges and in response to comments and criticisms, and the *Cochrane Database of Systematic Reviews* should be consulted for the most recent version of the review.

### Conclusion

We started with a problem from clinical practice, and have seen the importance of turning this into a well-focused question. A Cochrane systematic review takes the question and aims to pull together all the available evidence in an unbiased way to provide a clear answer to the question. At the same time, the quality of the evidence is summarised and a GRADE assessment made in relation to our confidence in the findings of the review in a summary of findings table (table 1).

### References

1. Cates CJ, Welsh EJ, Rowe BH. Holding chambers (spacers) versus nebulisers for β-agonist treatment of acute asthma. Cochrane Database of Syst Rev 2013; 9: CD000052.