

Supported by sanofi-aventis

**Pippa Hemingway**

PhD BSc (Hons) RGN
RSCN Research Fellow
in Systematic
Reviewing, School of
Health and Related
Research (SchARR),
University of Sheffield
Nic Brereton PhD BSc
(Hons) Health
Economist, NB
Consulting Services,
Sheffield

What is a systematic review?

- **Systematic reviews have increasingly replaced traditional narrative reviews and expert commentaries** as a way of summarising research evidence.
- Systematic reviews attempt to bring the **same level of rigour to reviewing research evidence** as should be used in **producing that research evidence** in the first place.
- Systematic reviews should be based on a peer-reviewed **protocol** so that they can be **replicated** if necessary.
- **High quality systematic reviews seek to:**
 - **Identify** all relevant published and unpublished **evidence**
 - **Select studies** or reports **for inclusion**
 - **Assess the quality of each study** or report
 - **Synthesise the findings** from individual studies or reports **in an unbiased way**
 - **Interpret the findings and present a balanced and impartial summary** of the findings with due consideration of any flaws in the evidence.
- Many high quality peer-reviewed systematic reviews are available in journals as well as from databases and other electronic sources.
- Systematic reviews may examine **quantitative or qualitative evidence**; put simply, when the two or more types of evidence are examined within one review it is called a **mixed-method** systematic review.
- Systematic reviewing techniques are in a period of rapid development. Many systematic reviews still look at **clinical effectiveness**, but methods now exist to enable reviewers to examine issues of **appropriateness, feasibility and meaningfulness**.
- Not all published systematic reviews have been produced with meticulous care; therefore, the **findings may sometimes mislead**. Interrogating published reports by asking **a series of questions can uncover deficiencies**.

For further titles in the series, visit:
www.whatisseries.co.uk

What is a systematic review?

Why systematic reviews are needed

The explosion in medical, nursing and allied healthcare professional publishing within the latter half of the 20th century (perhaps 20,000 journals and upwards of two million articles per year), which continues well into the new millennium, makes keeping up with primary research evidence an impossible feat. There has also been an explosion in internet access to articles, creating sometimes an awe-inspiring number of hits to explore. In addition, there is the challenge to build and maintain the skills to use the wide variety of electronic media that allow access to large amounts of information.

Moreover, clinicians, nurses, therapists, healthcare managers, policy makers and consumers have wide-ranging information needs; that is, they need good quality information on the effectiveness, meaningfulness, feasibility and appropriateness of a large number of healthcare interventions; not just one or two. For many, this need conflicts with their busy clinical or professional workload. For consumers, the amount of information can be overwhelming, and a lack of expert knowledge can potentially lead to false belief in unreliable information, which in turn may raise health professional workload and patient safety issues.

Even in a single area, it is not unusual for the number of published studies to run into hundreds or even thousands (before they are sifted for inclusion in a review). Some of these studies, once read in full text, may give unclear, confusing or contradictory results; sometimes they may not be published in our own language or there may be lack of clarity whether the findings can be generalised to our own country. Looked at individually, each article may offer little insight into the problem at hand; the hope is that, when taken together within a systematic review, a clearer (and more consistent) picture will emerge.

If the need for information is to be fulfilled, there must be an evidence translation stage. This is 'the act of transferring knowledge to individual health professionals, health facilities and health systems (and consumers) by means of publications, electronic media, education, training and decision support systems. Evidence transfer is seen to involve careful development of strategies that identify target audiences – such as clinicians, managers, policy makers and consumers – and designing methods to package and transfer information that is understood and used in decision-making'.¹

Failings in traditional reviews

Reviews have always been a part of the healthcare literature. Experts in their field have sought to collate existing knowledge and publish summaries on specific topics. Traditional reviews may, for instance, be called literature reviews, narrative reviews, critical reviews or commentaries within the literature. Although often very useful background reading, they differ from a systematic review in that they are not led via a peer-reviewed protocol and so it is not often possible to replicate the findings. In addition, such attempts at synthesis have not always been as rigorous as might have been hoped. In the worst case, reviewers may not have begun with an open mind as to the likely recommendations, and they may then build a case in support of their personal beliefs, selectively citing appropriate studies along the way. Indeed, those involved in developing a review may well have started a review (or have been commissioned to write one) precisely because of their accumulated experience and professional opinions. Even if the reviewer does begin with an open mind, traditional reviews are rarely explicit about how studies are selected, assessed and integrated. Thus, the reader is generally unable to assess the

likelihood of prior beliefs or of selection or publication biases clouding the review process. Despite all this, such narrative reviews were and are widespread and influential.

The lack of rigour in the creation of traditional reviews went largely unremarked until the late 1980s when several commentators exposed the inadequacies of the process and the consequent bias in recommendations.^{2,3} Not least of the problems was that small but important effects were being missed, different reviewers were reaching different conclusions from the same research base and, often, the findings reported had more to do with the specialty of the reviewer than with the underlying evidence.⁴

The inadequacy of traditional reviews and the need for a rigorous systematic approach were emphasised in 1992 with the publication of two landmark papers.^{5,6} In these papers, Elliot Antman, Joseph Lau and colleagues reported two devastating findings.

- First, **if original studies of the effects of clot busters after heart attacks had been systematically reviewed, the benefits of therapy would have been apparent as early as the mid-1970s.**
- Second, **narrative reviews were woefully inadequate in summarising the current state of knowledge.** These reviews either omitted mention of effective therapies or suggested that the treatments should be used only as part of an ongoing investigation – when in fact the evidence (if it had been collated) was near incontrovertible.

These papers showed that there was much knowledge to be gained from collating existing research but that traditional approaches had largely failed to extract this knowledge. What was needed was the same rigour in secondary research (research where the objects of study are other research studies) as is expected from primary research (original study).

When systematic reviews are needed

Conventionally, systematic reviews are needed to establish clinical and cost-

effectiveness of an intervention or drug. Increasingly, however, they are required to establish if an intervention or activity is feasible, if it is appropriate (ethically or culturally) or if it relates to evidence of experiences, values, thoughts or beliefs of clients and their relatives.¹

Systematic reviews are also:

- Needed to propose a future research agenda⁷ when the way forward may be unclear or existing agendas have failed to address a clinical problem
- Increasingly required by authors who wish to secure substantial grant funding for primary healthcare research
- Increasingly part of student dissertations or postgraduate theses
- Central to the National Institute for Health and Clinical Excellence health technology assessment process for multiple technology appraisals and single technology appraisals.

However, systematic reviews are most needed whenever there is a substantive question, several primary studies – perhaps with disparate findings – and substantial uncertainty. One famous case is described by The Cochrane Library:⁸ a single research paper, published in 1998 and based on 12 children, cast doubt on the safety of the mumps, measles and rubella (MMR) vaccine by implying that the MMR vaccine might cause the development of problems such as Crohn's disease and autism. The paper by Wakefield *et al*⁹ has since been retracted by most of the original authors because of potential bias, but before that it had triggered a worldwide scare, which in turn resulted in reduced uptake of the vaccine.¹⁰ A definitive systematic review by Demicheli *et al* on MMR vaccines in children concluded that exposure to MMR was unlikely to be associated with Crohn's disease, autism or other conditions.¹¹

Here, then, is an area where a systematic review helped clarify a vital issue to the public and to healthcare professionals; preparing such a review, however, is not a trivial exercise.

What is a systematic review?

The process of systematic review

The need for rigour in the production of systematic reviews has led to the development of a formal scientific process for their conduct. Understanding the approach taken and the attempts to minimise bias can help in the appraisal of published systematic reviews, which should help to assess if their findings should be applied to practice. The overall process should, ideally, be directed by a peer-reviewed protocol.

Briefly, developing a systematic review requires the following steps.

1. Defining an appropriate healthcare question. This requires a clear statement of the objectives of the review, intervention or phenomena of interest, relevant patient groups and subpopulations (and sometimes the settings where the intervention is administered), the types of evidence or studies that will help answer the question, as well as appropriate outcomes. These details are rigorously used to select studies for inclusion in the review.

2. Searching the literature. The published and unpublished literature is carefully searched for the required studies relating to an intervention or activity (on the right patients, reporting the right outcomes and so on). For an **unbiased** assessment, this search must seek to cover all the literature (not just MEDLINE where, for example, typically less than half of all trials will be found), including non-English sources. In reality, a designated number of databases are searched using a standardised or customised search filter. Furthermore, the grey literature (material that is not formally published, such as institutional or technical reports, working papers, conference proceedings, or other documents not normally subject to editorial control or peer review) is searched using specialised search engines, databases or websites. Expert opinion on where appropriate data may be located is sought and key authors are contacted for clarification. Selected journals are hand-searched when necessary and the references of full-text papers are also searched. Potential biases

within this search are publication bias,¹² selection bias and language bias.¹³

3. Assessing the studies. Once all possible studies have been identified, they should be assessed in the following ways.

- Each study needs to be assessed for **eligibility** against inclusion criteria and full text papers are retrieved for those that meet the inclusion criteria.
- Following a full-text selection stage, the remaining studies are assessed for **methodological quality** using a **critical appraisal framework**. Poor quality studies are excluded but are usually discussed in the review report.
- Of the remaining studies, **reported findings** are extracted onto a **data extraction form**. Some studies will be excluded even at this late stage. A list of **included studies** is then created.
- Assessment should ideally be conducted by two independent reviewers.

4. Combining the results. The findings from the individual studies must then be aggregated to produce a 'bottom line' on the clinical effectiveness, feasibility, appropriateness and meaningfulness of the intervention or activity. This aggregation of findings is called evidence synthesis. The type of **evidence synthesis** is chosen to fit the types(s) of data within the review. For example, if a systematic review inspects qualitative data, then a **meta-synthesis** is conducted.¹⁴ Alternatively, a technique known as **meta-analysis** (see *What is meta-analysis?*¹⁵ in this series) is used if homogenous quantitative evidence is assessed for clinical effectiveness. Narrative summaries are used if quantitative data are not homogenous.

5. Placing the findings in context. The findings from this aggregation of an unbiased selection of studies then need to be discussed to put them into context. This will address issues such as the quality and heterogeneity of the included studies, the likely impact of bias, as well as the chance and the applicability of the findings. Thus, judgement and balance are not obviated by the rigour of systematic reviews – they are just reduced in impact and made more explicit.

A word of caution, however. Performing a rigorous systematic review is far from easy. It requires careful scientific consideration at inception, meticulous and laborious searching, as well as considerable attention to methodological detail and analysis before it truly deserves the badge 'systematic'. The quality of a systematic review can be assessed by using a standard checklist. Example checklists are available from the NHS Public Health Resource Unit via the Critical Appraisal Skills Programme (CASP)¹⁶ or from the Centre for Evidence-Based Medicine at the University of Oxford.¹⁷ It is useful to have experience of primary and secondary research, or to collaborate with those that do, prior to undertaking a systematic review and to ensure that an academic and practice partnership directs the review.

The above has been an overview of the systematic review process. Clear guidance on the process of developing systematic reviews is available electronically^{18,19} from key texts such as the one by Khan *et al*²⁰ or via courses run at centres of excellence such as the NHS Centre for Reviews and Dissemination at the University of York or the Centre for Evidence-Based Medicine at the University of Oxford.

Some trends in systematic reviewing

Rapid evidence assessment reviews

Increasingly, health policy makers, clinicians and clients cannot wait the year or so required for a full systematic review to deliver its findings. Rapid evidence assessments (REAs) can provide quick summaries of what is already known about a topic or intervention. REAs use systematic review methods to search and evaluate the literature, but the comprehensiveness of the search and other review stages may be limited. The Government Social Research Unit has produced an REA toolkit which is recommended as a minimum standard for rapid evidence reviews.²¹ The toolkit states that an REA takes two to six months to complete and 'is a quick overview of existing research on a constrained topic and a synthesis of the evidence provided by these studies to answer the REA question'. Examples

of when an REA can be undertaken according to the REA toolkit include:

- 'When there is uncertainty about the effectiveness of a policy or service and there has been some previous research
- 'When a decision is required within months and policy makers/researchers want to make decisions based on the best available evidence within that time
- 'When a map of evidence in a topic area is required to determine whether there is any existing evidence and to direct future research needs.'²¹

An example of an REA to allow examination of the methods is a report by Underwood *et al* (2007), who evaluated the effectiveness of interventions for people with common mental health problems on employment outcomes.²²

User involvement

User involvement is well established as a prerequisite within primary research and is now increasingly expected within a systematic review. The Campbell Collaboration Users Group proposes 'a spectrum of user involvement in the systematic review process, ranging from determining the scope of the review and the outcomes of relevance, to determining the need for a review and involvement throughout all stages of production and dissemination.'²³ The definition of user involvement within the systematic review protocol is recommended; thus, what is expected from a user or user group and at which stages of the review should be clearly defined. For guidance on public involvement in research, access INVOLVE at www.invo.org.uk

Mixed methods

Increasingly, qualitative methods are used together with a randomised controlled trial to obtain a fuller picture of an intervention and the way it works.²⁴ It is also possible to mix methods within a systematic review as the methods to systematically review qualitative evidence, such as from grounded theory, phenomenology and other qualitative research designs, are now developed. This is particularly useful when different types of data such as qualitative data and quantitative data are available to inform a review topic. For

What is a systematic review?

example, the issues of a mixed-method synthesis have been described by Harden and Thomas (2005) on the basis of their review of the barriers to, and facilitators of, fruit and vegetable intake among children aged four to ten years.²⁵ The following issues arose from the merger of two simultaneous meta-syntheses of trial data (quantitative) and studies of experiences (qualitative).

Strengths of mixed methods

- They preserve the integrity of the findings of different types of studies by using the appropriate type of analysis that is specific to each type of finding.
- The use of categorical codes as a 'halfway' house to mediate between two forms of data was unproblematic.²⁵

Limitation of mixed methods

- There is potential researcher bias when categorical subgroups are not created *a priori* and are created later on in the review.²⁵

Finding existing reviews

High quality systematic reviews are published in many of the leading journals and electronic databases. In addition, electronic publication by the Cochrane Collaboration, the NHS Centre for Reviews and Dissemination and other organisations offers speedy access to regularly updated summaries (Box 1).

Drawbacks of systematic reviews

Systematic reviews appear at the top of the 'hierarchy of evidence' that informs evidence-based practice (practice supported by research

findings) when assessing clinical effectiveness (Box 2).²⁶ This reflects the fact that, when well conducted, they should give us the best possible estimate of any true effect. As noted previously, such confidence can sometimes be unwarranted, however, and caution must be exercised before accepting the veracity of any systematic review. A number of problems may arise within reviews of clinical effectiveness.

- Like any piece of research, **a systematic review may be done badly**. Attention to the questions listed in the section 'Appraising a systematic review' can help separate a rigorous review from one of poor quality.

- **Inappropriate aggregation of studies** that differ in terms of intervention used, patients included or types of data can lead to the **drowning of important effects**. For example, the effects seen in some subgroups may be concealed by a lack of effect (or even reverse effects) in other subgroups.

The **findings from systematic reviews are not always in harmony with the findings from large-scale high quality single trials**.^{27,28} Thus, findings from systematic reviews need to be weighed against perhaps conflicting evidence from other sources. Ideally, an updated review would deal with such anomalies.

Hierarchies of evidence for feasibility or appropriateness reviews are available²⁹ when most of the above applies.

Appraising a systematic review

Not all systematic reviews are rigorous and

Box 1. Useful websites for systematic reviews

- **The Cochrane Library** www.cochrane.org
- **The Joanna Briggs Institute** www.joannabriggs.edu.au/pubs/systematic_reviews.php
- **The Campbell Collaboration** www.campbellcollaboration.org
- **The Centre for Evidence-Based Medicine** www.cebm.net
- **The NHS Centre for Reviews and Dissemination** www.york.ac.uk/inst/crd
- **Bandolier** www.medicine.ox.ac.uk/bandolier
- **PubMed Clinical Queries: Find Systematic Reviews**
www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml

Box 2. Hierarchies of evidence for questions of therapy, prevention, aetiology or harm²⁶

- Level 1a** Systematic review (with homogeneity) of randomised controlled trials (RCTs)
- Level 1b** Individual RCT (with narrow confidence interval)
- Level 1c** All-or-none studies
- Level 2a** Systematic review (with homogeneity) of cohort studies
- Level 2b** Individual cohort study (including low quality RCT; eg <80% follow-up)
- Level 2c** 'Outcomes' research; ecological studies
- Level 3a** Systematic reviews (with homogeneity) of case-control studies
- Level 3b** Individual case-control study
- Level 4** Case series (and poor quality cohort and case-control studies)
- Level 5** Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

unbiased. The reader will want to interrogate any review that purports to be systematic to assess its limitations and to help decide if the recommendations should be applied to practice. Further guidance on appraising the quality of a systematic review can be found in several useful publications.^{16,30,31} Guidance focuses on the critical appraisal for reviews of clinical effectiveness. To reflect this, the following questions provide a framework.

- **Is the topic well defined** in terms of the intervention under scrutiny, the patients receiving the intervention (plus the settings in which it was received) and the outcomes that were assessed?
- **Was the search for papers thorough?** Was the search strategy described? Was manual searching used as well as electronic databases? Were non-English sources searched? Was the 'grey literature' covered – for example, non-refereed journals, conference proceedings or unpublished company reports? What conclusions were drawn about the possible impact of publication bias?
- **Were the criteria for inclusion of studies clearly described and fairly applied?** For example, were blinded or independent reviewers used?
- **Was study quality assessed by blinded or independent reviewers?** Were the findings related to study quality?
- **Was missing information sought from the original study investigators?** Was

the impact of missing information assessed for its possible impact on the findings?

- **Do the included studies seem to indicate similar effects?** If not, in the case of clinical effectiveness, was the heterogeneity of effect investigated, assessed and discussed?
- **Were the overall findings assessed for their robustness** in terms of the selective inclusion or exclusion of doubtful studies and the possibility of publication bias?
- **Was the play of chance assessed?** In particular, was the range of likely effect sizes presented and were null findings interpreted carefully? For example, a review that finds no evidence of effect may simply be an expression of our lack of knowledge rather than an assertion that the intervention is worthless.
- **Are the recommendations based firmly on the quality of the evidence presented?** In their enthusiasm, reviewers can sometimes go beyond the evidence in drawing conclusions and making their recommendations.
All studies have flaws. It is not the mere presence of flaws that vitiates the findings. Even flawed studies may carry important information. The reader must exercise judgement in assessing whether individual flaws undermine the findings to such an extent that the conclusions are no longer adequately supported.

What is a systematic review?

References

- Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. *Int J Evid Based Healthc* 2005; **3**: 207–215.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987; **106**: 485–488.
- Teagarden JR. Meta-analysis: whither narrative review? *Pharmacotherapy* 1989; **9**: 274–281.
- Spector TD, Thompson SG. The potential and limitations of meta-analysis. *J Epidemiol Community Health* 1991; **45**: 89–92.
- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; **268**: 240–248.
- Lau J, Antman EM, Jimenez-Silva J *et al*. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; **327**: 248–254.
- Torgerson C. *Systematic Reviews*. London: Continuum, 2003.
- The Cochrane Library. *The Cochrane Library publishes the most thorough survey of MMR vaccination data which strongly supports its use*. www.cochrane.org/press/MMR_final.pdf (last accessed 19 November 2008)
- Wakefield AJ, Murch SH, Anthony A *et al*. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 1998; **351**: 637–641.
- Murch SH, Anthony A, Casson DH *et al*. Retraction of an interpretation. *Lancet* 2004; **363**: 750.
- Demicheli V, Jefferson T, Rivetti A, Price D. Vaccines for measles, mumps and rubella in children. *Cochrane Database Syst Rev* 2005; CD004407.
- Dubben HH, Beck-Bornholdt HP. Systematic review of publication bias in studies on publication bias. *BMJ* 2005; **331**: 433–434.
- Egger M, Zellweger-Zahner T, Schneider M *et al*. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; **350**: 326–329.
- Sandelowski M, Barroso J. Toward a metasynthesis of qualitative findings on motherhood in HIV-positive women. *Res Nurs Health* 2003; **26**: 153–170.
- Crombie IK, Davies HTO. *What is meta-analysis?* London: Hayward Medical Communications, 2009.
- Public Health Resource Unit. Critical Appraisal Skills Programme (CASP). *Making sense of evidence: 10 questions to help you make sense of reviews*. www.phru.nhs.uk/Doc_Links/S.Reviews%20Appraisal%20Tool.pdf (last accessed 19 November 2008)
- Centre for Evidence-Based Medicine. *Critical Appraisal*. www.cebm.net/index.aspx?o=1157 (last accessed 23 January 2009)
- Higgins J, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.0.1 [updated September 2008]. www.cochrane-handbook.org/ (last accessed 19 November 2008)
- NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*. CRD Report 4. York: University of York, 2001.
- Khan KS, Kunz R, Kleijnen J (eds). *Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research*. London: Royal Society of Medicine Press, 2003.
- Government Social Research. *Rapid Evidence Assessment Toolkit*. www.gsr.gov.uk/professional_guidance/rea_toolkit/index.asp (last accessed 23 January 2009)
- Underwood L, Thomas J, Williams T, Thieba A. *The effectiveness of interventions for people with common mental health problems on employment outcomes: a systematic rapid evidence assessment*. <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=Wgr%2bPGyPMD0%3d&tabid=2315&mid=4279&language=en-US> (last accessed 23 January 2009)
- Campbell Collaboration Users Group. *User Involvement in the systematic review process*. *Campbell Collaboration Policy Brief*. http://camp.ostfold.net/artman2/uploads/1/Minutes_March_2008_Oslo.pdf (last accessed 19 November 2008)
- Evans D, Pearson A. Systematic reviews: gatekeepers of nursing knowledge. *J Clin Nurs* 2001; **10**: 593–599.
- Harden A, Thomas J. Methodological Issues in Combining Diverse Study Types in Systematic Reviews. *International Journal of Social Research Methodology* 2005; **8**: 257–271.
- Phillips B, Ball C, Sackett D *et al*; Centre for Evidence-Based Medicine. *Levels of Evidence*. www.cebm.net/index.aspx?o=1025 (last accessed 19 November 2008)
- LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; **337**: 536–542.
- Egger M, Smith GD. Misleading meta-analysis. *BMJ* 1995; **310**: 752–754.
- Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003; **12**: 77–84.
- Crombie IK. *Pocket Guide to Critical Appraisal*. London: BMJ Publishing Group, 1996.
- Moher D, Cook DJ, Eastwood S *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; **354**: 1896–1900.

First edition published 2001
Authors: Huw TO Davies and
Iain K Crombie

This publication, along with the others in the series, is available on the internet at www.whatisseries.co.uk
The data, opinions and statements appearing in the article(s) herein are those of the contributor(s) concerned. Accordingly, the sponsor and publisher, and their respective employees, officers and agents, accept no liability for the consequences of any such inaccurate or misleading data, opinion or statement.

Published by Hayward Medical Communications, a division of Hayward Group Ltd.
Copyright © 2009 Hayward Group Ltd.
All rights reserved.



Supported by *sanofi-aventis*