8 Assessing all the relevant, reliable evidence

IS ONE STUDY EVER ENOUGH?

The simple answer is 'hardly ever'. Very seldom will one fair treatment comparison yield sufficiently reliable evidence on which to base a decision about treatment choices. However, this does sometimes happen. Such rare single studies include one showing that taking aspirin during a heart attack reduces the risk of premature death;[1] another making clear that giving steroids to people with acute traumatic brain injury is lethal (see below and Chapter 7, p89-90); and a third identifying caffeine as the only drug known to prevent cerebral palsy in children born prematurely (see Chapter 5, p57-58). Usually, however, a single study is but one of several comparisons addressing the same or similar questions. So evidence from individual studies should be assessed alongside evidence from other, similar studies.

One of the pioneers of fair tests of treatments, the British statistician Austin Bradford Hill, said in the 1960s that reports of research should answer four questions:

- Why did you start?
- What did you do?
- What did you find?
- And what does it mean anyway?

WHY DID YOU START?

'Few principles are more fundamental to the scientific and ethical validity of clinical research than that studies should address questions needing to be answered, and that they are designed in a way that will produce a meaningful answer. A prerequisite for either of these goals is that relevant prior research be properly identified. . . . An incomplete picture of pre-existing evidence violates the implicit ethical contract with research participants that the information they provide is necessary and will be useful to others.'

Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Annals of Internal Medicine* 2011:154:50-55.

These key questions are equally relevant today, yet they are too often inadequately addressed or overlooked completely. The answer to the last question – what does it mean? – is especially important since this is likely to influence decisions about treatment and future research.

Take the example of a short, inexpensive course of steroid drugs given to women expected to give birth prematurely. The first fair test of this treatment, which was reported in 1972, showed a reduced likelihood of babies dying after their mothers had received a steroid. A decade later more trials had been done, but these were small and the individual results were confusing, because none of them had taken systematic account of previous, similar studies. Had they done so, it would have been apparent that very strong evidence was emerging favouring a beneficial effect of the drugs. In fact, because this was not done until 1989, most obstetricians, midwives, paediatricians and neonatal nurses had meanwhile not realized the treatment was so effective. As a result, tens of thousands of premature babies had suffered and died unnecessarily.[2]

To answer the question 'what does it mean?', the evidence from a particular fair treatment comparison must be interpreted

SYNTHESIZING INFORMATION FROM RESEARCH

More than a century ago, the president of the British Association for the Advancement of Science, Lord Rayleigh, commented on the need to set the results of new research in the context of other relevant evidence:

'If, as is sometimes supposed, science consisted in nothing but the laborious accumulation of facts, it would soon come to a standstill, crushed, as it were, under its own weight . . . Two processes are thus at work side by side, the reception of new material and the digestion and assimilation of the old; and as both are essential we may spare ourselves the discussion of their relative importance . . . The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out.'

Rayleigh, Lord. In: *Report of the fifty-fourth meeting of the British Association for the Advancement of Science; held at Montreal in August and September 1884*. London: John Murray, 1884: pp3-23.

alongside evidence from the other, similar fair comparisons. Reporting new test results without interpreting them in the light of other relevant evidence, reviewed systematically, can delay identification of both useful and harmful treatments, and lead to unnecessary research.

SYSTEMATIC REVIEWS OF ALL THE RELEVANT, RELIABLE EVIDENCE

Whilst it is easy to state that we should review the results of a particular study alongside other relevant, reliable evidence, this is a challenge in many ways. Reviews are important because people should be able to depend on them, and that means that they must be done systematically, otherwise they will be misleading.

THE IMPORTANCE OF SYSTEMATIC REVIEWS

'Systematic reviews and meta-analyses have become increasingly important in health care. Clinicians read them to keep up to date with their field, and they are often used as a starting point for developing clinical practice guidelines. Granting [funding] agencies may require a systematic review to ensure there is justification for further research, and some health care journals are moving in this direction. As with all research, the value of a systematic review depends on what was done, what was found, and the clarity of reporting. As with other publications, the reporting quality of systematic reviews varies, limiting readers' ability to assess the strengths and weaknesses of those reviews.'

Moher D, Liberati A, Tetzlaff, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (www.equator-network.org), 2009.

Systematic reviews addressing what appears to be the same question about treatments may reach different conclusions. Sometimes this is because the questions addressed are subtly different, or because the methods used by the researchers differed; and sometimes it is because the researchers have introduced 'spin' in their conclusions. So, it is important to identify reviews that address the treatment questions that match those we are interested in; which are most likely to have been prepared in ways that reduce the effects of biases and the play of chance successfully; and which reach honest conclusions, in ways that reflect the evidence presented.

Reducing biases in systematic reviews

Just as biases can distort individual tests of treatments and lead to false conclusions, so they can also distort reviews of evidence. For example, researchers can simply 'cherry pick' those studies which they know will support the treatment claims they wish to make.

To avoid these problems, plans for systematic reviews, as for

individual research studies, should be set out in research protocols. Protocols need to make clear what measures researchers will take to reduce biases and the effects of the play of chance during the process of preparing the reviews. These will include specifying which questions about treatments the review will address; the criteria that make studies eligible for inclusion in the review; the ways in which potentially eligible studies will be identified; and the steps that will be taken to minimize biases in selecting studies for inclusion in the review, and for analysing the data.

Identifying all the relevant evidence for systematic reviews

Identifying all the relevant evidence for systematic reviews – irrespective of the language or format of the relevant reports – always presents a substantial challenge, not least because some relevant evidence has not been reported in public. Under-reporting stems principally from researchers not writing up or submitting reports of their research for publication because they were disappointed with the results. And pharmaceutical companies suppress studies that do not favour their products. Journals, too, have tended to show bias when they reject submitted reports because they deem their results insufficiently 'exciting'.[3]

Biased under-reporting of research is unscientific and unethical, and there is now widespread acceptance that this is a serious problem. In particular, people trying to decide which treatments to use can be misled because studies that have yielded 'disappointing' or 'negative' results are less likely to be reported than others, whereas studies with exciting results are more likely than others to be 'over-reported'.

The extent of under-reporting is astonishing: at least half of all clinical trials are never fully reported. This under-reporting of research is biased and applies to large as well as small clinical trials. One of the measures that has been taken to tackle this problem has been to establish arrangements for registering trials at inception, and encouraging researchers to publish the protocols for their studies.[3]

Biased under-reporting of research can even be lethal. To their great credit, some British researchers decided to report in 1993 the results of a clinical trial that had been done thirteen

MARKETING-BASED MEDICINE

'Internal documents from the pharmaceutical industry suggest that the publicly available evidence base may not accurately represent the underlying data regarding its products. The industry and its associated medical communication firms state that publications in the medical literature primarily serve marketing interests. Suppression and spinning of negative data and ghostwriting [see Chapter 10, p124-5] have emerged as tools to help manage medical journal publications to best suit product sales, while disease mongering and market segmentation of physicians are also used to efficiently maximize profits. We propose that while evidence-based medicine is a noble ideal, marketing-based medicine is the current reality.'

Spielmans GI, Parry PI. *From Evidence-based Medicine to Marketing-based Medicine: Evidence from Internal Industry Documents. Journal of Bioethical Inquiry* 2010;7(1):13-29. Available online: http://tinyurl.com/Spielmans.

years earlier. It concerned a new drug for reducing heart rhythm abnormalities in patients experiencing heart attacks. Nine patients had died after taking the drug, whereas only one had died in the comparison group. 'When we carried out our study in 1980,' they wrote, 'we thought that the increased death rate in the drug group was an effect of chance… The development of the drug [lorcainide] was abandoned for commercial reasons, and this study was therefore never published; it is now a good example of "publication bias". The results described here…might have provided an early warning of trouble ahead'.[4] The 'trouble ahead' to which they were referring was that, at the peak of their use, drugs similar to the one they had tested were causing tens of thousands of premature deaths every year in the USA alone (see Chapter 2, p14-15).[5]

Reducing the play of chance in systematic reviews

In Chapter 7 (p91), we explained how the play of chance can be reduced by combining data from similar but separate studies – a process known as 'meta-analysis'. We used the example of five studies in five different countries organized and funded separately to address a 60-year-old quandary about what blood level of oxygen in prematurely born infants is needed to maximize the likelihood that they will survive with no major disabilities. That example illustrated how this process could be planned *before* the results of the studies were available, but the same process can be used *after* a group of similar studies have been completed.

For example, in 1974 a Swedish doctor conducted a systematic review of studies comparing the results of surgery for breast cancer with or without radiotherapy.[6] He found that, in all of the studies, women were more likely to die in the groups receiving radiotherapy. When all of this evidence was synthesized statistically using meta-analysis, it became clear that this excess mortality was unlikely to reflect the play of chance. Subsequent, more detailed analyses, based on evidence from individual patients, confirmed that the radiotherapy being used during that era did indeed increase mortality.[7] Recognizing this led to the development of safer practices.

Recognizing vested interests and spin in systematic reviews

What if the reviewers have other interests that might affect the conduct or interpretation of their review? Perhaps the reviewers have received money from the company that made the new treatment being tested. When assessing the evidence for an effect of evening primrose oil on eczema, reviewers who were associated with the manufacturer reached far more enthusiastic conclusions about the treatment than those with no such commercial interest (see Chapter 2, p18-20). However, commercial interests are not alone in leading to biased reviews. We all have prejudices that can do this – researchers, health professionals, and patients alike.

Disappointingly, people with vested interests sometimes exploit biases to make treatments look as if they are better than they really are (see also Chapter 10).[8] This happens when some researchers – usually but not always for commercial reasons –

deliberately ignore existing evidence. They design, analyze, and report research to paint their own results for a particular treatment in a favourable light. This is what happened in the 1990s when the manufacturer of the anti-depressant drug Seroxat (paroxetine) withheld important evidence suggesting that, in adolescents, the drug actually increased symptoms that prompted some of these young patients to contemplate suicide as a way of dealing with their depression.[9]

Over-reporting is a problem as well. In a phenomenon known as 'salami slicing', researchers take the results from a single trial (the salami) and slice the results into several reports without making clear that the individual reports are not independent studies. In this way, a single 'positive' trial can appear in several journals in different articles, thereby introducing a bias.[10] Here again, registering trials at inception with unique identifiers for every study will help to reduce the confusion that can result from this practice.

WHAT CAN HAPPEN IF ALL THE RELEVANT, RELIABLE EVIDENCE IS NOT ASSESSED?

Fair tests of treatments involve reviewing systematically all the relevant, reliable evidence, to see what is already known, whether from animal or other laboratory research, from the healthy volunteers on whom new treatments are sometimes tested, or from previous research involving patients. If this step is overlooked, or done badly, the consequences can be serious – patients in general, as well as participants in research, may suffer and sometimes die unnecessarily, and precious resources both for healthcare and for research will be squandered.

Avoidable harm to patients

Recommended treatments for heart attacks that had appeared in textbooks published over a period of 30 years were compared with evidence that could have been taken into account had the authors systematically reviewed the results of fair tests of treatment reported during that time.[11] This comparison showed that the textbook recommendations were often wrong because the authors

SCIENCE IS CUMULATIVE, BUT SCIENTISTS DON'T ACCUMULATE EVIDENCE SCIENTIFICALLY

'Academic researchers have been talking about something called "cumulative meta-analysis" for 25 years: essentially, you run a rolling meta-analysis on a given intervention, and each time a trial is completed, you plug the figures in to get your updated pooled result, to get a feel for where the results are headed, and most usefully, have a good chance of spotting a statistically significant answer as soon as it becomes apparent, without risking lives on further unnecessary research.'

Goldacre B. Bad Science: How pools of blood trials could save lives. *The Guardian*, 10 May 2008, p16.

had not reviewed the relevant evidence systematically. The impact of this was devastating. In some cases, patients with heart attacks were being deprived of life-saving therapies (for example, clot-busting drugs). In other cases, doctors continued to recommend treatments long after fair tests had shown they were lethal (for example, the use of drugs that reduce heart rhythm abnormalities in patients having heart attacks (see above and Chapter 2, p14-15).

The failure to combine the results of studies in systematic reviews as new evidence becomes available continues to harm patients. Blood substitutes that need no refrigeration or cross-matching are an obviously attractive alternative to real blood for the treatment of haemorrhage. Unfortunately these products increase the risk of heart attacks and death. Furthermore, a systematic review of the randomized trials reported since the late 1990s reveals that their dangers could and should have been recognized several years earlier than they were.[1]

Avoidable harm to people participating in research

Failure to assess all relevant, reliable evidence can also result in avoidable harm to people who participate in research. Researchers

continue to be commissioned and allowed to do studies that involve withholding treatments known to be effective. For example, long after reliable evidence was available showing that giving antibiotics to patients having bowel surgery reduced their chances of dying from complications of the operation, researchers continued to do comparison studies that involved withholding antibiotics from half the patients participating in controlled trials. The researchers' failure to review systematically what was already known deprived half the participants in their studies of a known beneficial treatment. This serious lapse was evidently overlooked by the funding bodies who financed their research, and by the research ethics committees which reviewed the protocols and failed to challenge the researchers.

It is not only patients requiring treatment who can be put at risk if researchers do not assess systematically what is already known about the effects of the treatments they will be given. Healthy volunteers can be harmed too. The first phase of testing some treatments often involves a very small number of healthy volunteers. In 2006, six young men volunteers at a private research facility in West London were given infusions of a drug that had not previously been used in people. They all suffered life-threatening complications – one of them losing fingers and toes – and their long-term health has been compromised. This tragedy could most probably have been avoided[13] if a report of a severe reaction to a similar drug had been submitted for publication,[14] and if the researchers had assessed systematically what was already known about the effects of such drugs.[15] Had they done so, they might not have proceeded with their study at all, or if they had decided to go ahead, they might have injected the volunteers one at a time rather than simultaneously; and they could and should have warned the healthy young volunteers about the possible dangers.[16]

Wasted resources in healthcare and research

Failure to do systematic reviews of relevant, reliable research evidence does harm even when it is not harming patients and people participating in research. This is because it can result in resources being wasted in healthcare and health research. During

COULD CHECKING THE EVIDENCE FIRST HAVE PREVENTED A DEATH?

'In a tragic situation that could have been averted, Ellen Roche, a healthy, 24-year-old volunteer in an asthma study at Johns Hopkins University, died in June [2001] because a chemical she had been asked to inhale led to the progressive failure of her lungs and kidneys. In the aftermath of this loss, it would appear that the researcher who conducted the experiment and the ethics panel that approved it allegedly overlooked numerous clues about the dangers of the chemical, hexamethonium, given to Roche to inhale. Adding particular poignancy to the case is that evidence of the chemical's dangers could easily have been found in the published literature. *The Baltimore Sun* concluded that while the supervising physician, Dr. Alkis Togias, made "a good-faith effort" to research the drug's adverse effects, his search apparently focused on a limited number of resources, including PubMed, which is searchable only back to 1966. Previous articles published in the 1950s, however, with citations in subsequent publications, warned of lung damage associated with hexamethonium.'

Perkins E. Johns Hopkins Tragedy. *Information Today* 2001;18:51-4.

the 1980s and 1990s, for example, a total of more than 8,000 patients participated in several tests of a proposed new drug for stroke. Dutch researchers reviewed the results of these drug studies systematically, and were unable to find any beneficial effects (see Chapter 10, p121).[17] They then decided to review the results of tests of the drug done previously in animals; again, they were unable to find any beneficial effects.[18] Had the researchers who did the tests in animals and the clinical researchers reviewed the results of the animal studies systematically, as they had emerged, it is very likely that thousands of patients would not have been invited to participate in the clinical trials. Indeed, this might have resulted in better use of resources for treating patients experiencing stroke, and studies

that were more likely to be relevant to identifying improvements in treatments for the condition. And this is far from an isolated example.[19]

REPORTS OF NEW RESEARCH SHOULD BEGIN AND END WITH SYSTEMATIC REVIEWS

The report of a study[20] to assess the effects of giving steroids to people with acute traumatic brain injury shows how to address all of Bradford Hill's four questions. The researchers explained that they had embarked on the study because their systematic review of all the existing evidence, as well as evidence of variations in clinical use of the treatment, showed that there was important uncertainty about the effects of this widely used treatment. They reported that they had registered and published the protocol for

INSTRUCTIONS TO AUTHORS TO PUT RESEARCH RESULTS IN CONTEXT BY THE EDITORS OF THE MEDICAL JOURNAL *THE LANCET*

Systematic Review

This section should include a description of how authors searched for all the evidence. Authors should also say how they assessed the quality of that evidence – ie, how they selected and how they combined the evidence.

Interpretation

Authors should state here what their study adds to the totality of evidence when their study is added to previous work.

'We ask that all research reports – randomised or not – submitted from Aug 1 . . . put the results into the context of the totality of evidence in the Discussion.'

Clark S, Horton R. Putting research in context – revisited. *Lancet* 2010;376:10-11.

their study, when it started.

They described the measures they had taken to minimize biases and to achieve adequate control of the play of chance by studying a sufficiently large number of patients. They reported that their study had shown that steroids given to patients with serious brain injury increased the likelihood that these patients would die.

Finally and importantly, they provided readers of their report with all the evidence needed for action to prevent thousands of deaths from this widely used treatment because they updated their original systematic review of previous studies by incorporating the new evidence generated by their study.

KEY POINTS

- A single study rarely provides enough evidence to guide treatment choices in healthcare
- Assessments of the relative merits of alternative treatments should be based on systematic reviews of all the relevant, reliable evidence
- As in individual studies testing treatments, steps must be taken to reduce the misleading influences of biases and the play of chance
- Failure to take account of the findings of systematic reviews has resulted in avoidable harm to patients, and wasted resources in healthcare and research