Chapter 14
Assessment and evaluation of research

14.1 Introduction

Researchers need to have the skill to assess and evaluate the research papers they read, particularly those related to the research topic they are doing. This should be done before the research is planned, during the implementation of the project, and before discussing the results and preparing to communicate them. Researchers may also want to critically assess all accessible published papers on a particular topic in order to write a systematic review. They should bear in mind that science should not be admired; science should be questioned. The words “author” and “authority” come from a common English stock and run the danger of becoming synonyms in the minds of some. A good scientist should develop a sceptical attitude when reading scientific papers. Scepticism is an inherent part of the scientific approach. What defines any statement as being scientific is that it is verifiable in principle, or, as it is sometimes put, it should be “falsifiable” in principle. There is hardly any theory in science that ever achieves a degree of certainty beyond the reach of criticism or the possibility of modification. In science, there will always be more beyond.

Researchers may also be requested to peer-review a scientific paper submitted for publication by other researchers, or to assess the scientific output of candidates for academic posts.

The need to assess and evaluate research is not limited to researchers. Learning to evaluate and use research findings is an important and lifelong part of professional development for health professionals. They need to critically assess the value of new published research before considering its practical implications for their work. Health professionals need to be aware of the fact that there are different levels for scientific evidence. Health researchers should help in outlining these different levels of evidence.

Policy-makers should have the ability to assess research results and their implications for policy. In particular, they need to assess new technologies and also currently used technologies, to introduce what is new and cost-effective, discard what is not effective or potentially harmful, promote what is effective but under-utilized, and postpone a
decision where evidence is still lacking. Health researchers need to be aware of these considerations.

Research is an investment, and is becoming more and more expensive. Those who fund the research need to evaluate the return on their investment. Researchers need to be aware about how the investment in health research is evaluated by funding agencies, particularly governments, their public paymasters.

This chapter addresses the assessment and evaluation of research by researchers, health professionals, policymakers, and investors in health research. For additional information on the subject, the sources listed in the references and additional sources for the chapter can be consulted.

14.2 Assessment and evaluation by researchers

14.2.1 Reading a research paper

The title of the paper and the abstract give an indication of the novelty and relevance of the paper.

For the critical reader, the methods section should be the first part of the paper to assess. It will tell whether it is good science or bad science. It has been rightly said that a paper will sink or swim on the strength of its methods section (Greenhalgh, 1997). A good methods section should provide sufficient detail to allow other investigators to replicate the study and confirm the results. If it does not, the study results cannot be easily accepted.

In most papers, the two most important methodological issues relate to how the sample was selected and what measurements were made. The sample must be representative of the population studied. If two samples are compared, they must be selected to be identical for every relevant variable, except the one to be studied. The critical reader must question whether the measurements used have been assessed for their validity and their reliability. As discussed in Chapter 4, validity is an index of how well a test or procedure measures what it is intended to measure. Reliability assesses consistency of measurement. It relates to the reproducibility of measurements. When reliability is high, a test that is repeated on the same patient and under the same conditions will yield the same result, whether by different investigators (Inter-rater reliability), or by the same investigator (Intra-rater reliability). Where appropriate, the investigators should provide assurance about the quality control of their data. As an example of the importance of inter-rater reliability, one study looked at the agreement among four pathologists on the classification of cervical intra-epithelial neoplasia, compared with the index pathologist. Of 101 cases of carcinoma in situ (CIS), 6 were reported as mild dysplasia, 19 as moderate dysplasia, 54 as severe dysplasia, and 22 as CIS (deVet et al., 1990).
The critical reader of a scientific paper takes a close look at the results and their interpretation. Pitfalls in the interpretation of research results are discussed in detail in Chapter 9.

Statistical jargon should not put off the critical reader. Use and interpretation of statistics can be misleading. Disraeli is quoted as saying “There are three types of lies: lies, damn lies and statistics”. One does not need to be a statistician to make some judgement about the statistical analysis of the research. Statistics is about common sense, before it is about mathematics. The first question to ask is whether the authors have used any statistical methods at all. If they have not, there is no reason to accept that the results are not being caused by chance alone. The second question is whether the authors have selected the right statistical methods to analyse their data. The third question is whether they have drawn the right conclusions from the statistical analysis. It is tempting to make wrong conclusions on the basis of statistical analysis. There is a limit to what statistics can tell us.

14.2.2 Peer review

Peer review is the critical assessment of manuscripts submitted to scientific journals by experts who are not part of the editorial process. The process of peer review helps editors to decide which manuscripts are suitable for publication, and helps authors to improve the quality of their papers. A peer-reviewed journal is a journal that submits most of its published research articles for outside review.

In the peer review process, editors generally provide reviewers with a format for the assessment of all components of the paper, from the title to the references. There is a common misconception that finding flaws is key to the high quality of peer review. The objective of the peer review process is not to find something to criticize. Finding flaws is certainly important, and scepticism is revered in scientific tradition. Authors can benefit from constructive criticism of good reviewers. However, responding to misguided comments may waste time and effort.

There are ethical considerations in the peer review process. Reviewers must disclose to editors any conflicts that could bias their opinions of the manuscript, and they should disqualify themselves from reviewing specific manuscripts if appropriate. Editors should avoid selecting external peer reviewers with obvious potential conflict of interest, for example those who work in the same department or institution. Reviewers must not use knowledge of the work before its publication to further their own scientific interests.
14.3 Assessment and evaluation by health professionals

14.3.1 Levels of evidence

Health professionals reading scientific papers for possible clinical application should recognize that there is a hierarchy of the level of evidence obtained from different study designs. In assessing the effectiveness of 169 interventions, the U.S. Preventive Services Task Force (1989), including a 20-member panel of scientific and medical experts, proposed the following guide for rating the quality of evidence for clinical effectiveness.

- **Level I evidence**: Evidence obtained from at least one properly designed randomized controlled trial
- **Level II-1 evidence**: Evidence obtained from well-designed controlled trials without randomization
- **Level II-2 evidence**: Evidence obtained from well-designed cohort or case-control studies. In these observational studies, the investigator has no role in assignment of study exposure but, rather, observes the natural course of events of exposure and outcome.
- **Level II-3 evidence**: This category includes cross-sectional studies, which are observational studies that assess the status of individuals with respect to the presence or absence of both exposure and outcome, at a particular time. The category also includes uncontrolled intervention studies. They may demonstrate impressive results, but in the absence of a control group the results may be attributable to factors other than the intervention or treatment. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) may, however, be difficult to dismiss.
- **Level III evidence**: This category includes descriptive studies, such as case reports and case series. It also includes expert opinion, often based on clinical experience.

14.3.2 Systematic reviews and meta-analyses

Results of scientific studies are often not uniform. To try to draw conclusions from these studies, systematic reviews are undertaken by researchers. A systematic review, as outlined in Chapter 11, is an overview of primary studies that contains an explicit statement of objectives, materials and methods, and has been conducted according to explicit and reproducible methodology. It is different from a narrative review, which is an overview of primary studies that have not been identified or analysed in a systematic (standardized and objective) way.
The quality of systematic reviews should generally be judged by the following two criteria:

- Have the authors performed a thorough literature review or presented only selected research findings?
- Have they accepted the primary researchers’ interpretation of study data uncritically, or do they include methodological commentary along with their content review?

A meta-analysis, as discussed in Chapter 11, is a special type of systematic review that combines results from more than one investigation to obtain a weighted average of the effect of a variable or intervention on a defined outcome. Combining data from a number of studies increases the sample size and the power of the study to provide statistically significant conclusions. A meticulously conducted meta-analysis, in which all the primary studies on a particular subject have been hunted out and critically appraised according to rigorous criteria, has a very high place in the hierarchy of evidence.

In reading a meta-analysis study, it should be recognized that a meta-analysis can only be as good as the quality of its individual components. Assessment of quality of a meta-analysis has to address the following questions:

- Is the pooling done only among studies where there is reasonable assurance that subjects and treatments are similar? Misleading conclusions can be drawn from pooling together heterogeneous data.
- Has care been taken to exclude publication bias toward positive results? Studies with positive results are more likely to be published, leading to problems with meta-analysis interpretation; many researchers are reluctant to pursue and publish negative results.

### 14.3.3 Cochrane Collaboration

The Cochrane Collaboration focuses on identifying reliable evidence and preparing systematic reviews of therapeutic interventions using randomized controlled trials (RCTs) (Bero and Rennie, 1995). Archie Cochrane was a Scottish epidemiologist who worked in Wales for most of his life. In 1972, he wrote a book in which he highlighted the absence of an adequate knowledge base for much of the health care provided. He made a strong case for the evaluation of new and current forms of care in controlled trials, which use randomization to generate unbiased comparison groups. Cochrane first challenged the profession of obstetrics to seek good evidence for its practice. The challenge was taken up, and the database of perinatal trials was the first to come out. Having demonstrated that the approach was possible with one specialty, the work was extended to other areas of health care. In 1992, the first Cochrane Centre was opened in Oxford, and the Cochrane Collaboration was launched internationally one year later. The Cochrane Library (http:
Assessment and evaluation of research

//www.update-software.com/cochrane/) is currently considered one of the best single sources of critical evidence for health care interventions. The library publishes a database solely of RCTs. It is published on a quarterly basis and made available both on CD-ROM and on the internet. It is easily accessible in a user-friendly format. It is the result of collaborative hand-searching efforts and electronic searching from many of the different review groups and centres of the Cochrane Collaboration. Collaborative review groups have evolved, which cover most areas of health care.

14.4 Assessment and evaluation by policy-makers

There has been an explosion of technologies in the past few decades as an outcome of the expansion in health research. These technologies provide great opportunities in health care. The assessment of these technologies presents major challenges to health policy-makers. A major challenge is how these technologies can be assessed to determine their appropriateness. Assessment should not be limited to newly introduced technologies. There is a need also to assess technologies currently in use, which may not be effective or even potentially harmful. There are also beneficial technologies which may be underutilized. Technology can be defined as the implementation of scientific knowledge in order to satisfy human needs. Health technologies include the drugs, devices, equipment and medical and surgical procedures used in the prevention, detection, diagnosis and treatment and rehabilitation of disease.

The responsibility for assessment of health technologies is ill defined. Drug regulatory authorities have responsibility for the approval of drugs for human use. Based on pre-clinical and clinical studies, the authority decides whether the drug is safe and effective to do what it is claimed to do. But it is not the business of the drug regulatory authority to compare the drug with other available drugs. It only ensures that the manufacturer makes no unjustified claims. This is the status of drug regulation, but health technologies include also devices, equipment and procedures. Devices are only regulated if they are used inside the human body. Medical equipment and medical and surgical procedures are not, in general, subject to regulation by authorities; not that such regulation is desirable in a rapidly advancing field.

The following four questions need to be carefully examined before any new technology is considered appropriate:

• Is the technology evidence-based?
• Is it good value for money?
• Is it culturally and ethically acceptable?
• Are the system requirements for its introduction available?
Is the technology evidence-based?

There is a need to critically assess the evidence before adopting any new technology. This is particularly important when there are strong commercial interests involved. The practice of medicine has been rapidly evolving from being authority-based to being evidence-based. The history of our medical practice is not short of examples of technologies which were widely used and subsequently proved not useful or even harmful.

There are ongoing efforts to assess currently available health technologies. In an ongoing assessment of reproductive health technologies, WHO classified these technologies into the following six categories: beneficial, likely to be beneficial, with a trade-off, of unknown effectiveness, likely to be ineffective, and likely to be harmful (WHO, 2002). In the UK, the National Institute for Clinical Excellence (NICE) was set up as a special health authority for England and Wales in 1999. Its role is to provide patients, health professionals, and the public with authoritative, robust and reliable guidance on current “best practice” (www.nice.org.uk).

Is the technology good value for money?

If the technology is evidence-based, the next question is whether it is good value for money. This is a different question from the issue of affordability. Economists have shown an increasing interest in what health professionals are doing, contributing a new discipline of health economics. With the increasing introduction of health technologies, health care has become too costly to be left to health care providers alone. Research on health economics is discussed in Chapter 4.

Health economists introduced two important concepts to consider in deciding whether a new technology is good value for money: cost-effectiveness and opportunity cost. Cost-effectiveness measures the net cost of providing a service as well as the effectiveness of the service. The result of cost-effectiveness analysis is expressed as the monetary cost per unit of effectiveness. To illustrate this concept, let us take the example of an assisted reproduction technology procedure. The cost is measured against the desired outcome, “a take home baby”, not simply by the cost of the procedure. If the success rate is, say, 25%, then the cost per take home baby is four times the cost of the procedure. If a new technology is claimed to raise the success rate by 10%, but the procedure also has an additional cost, we need to bear in mind that, for each one additional “take home baby”, ten patients must receive this new procedure. The additional cost of one “take home baby” will be ten times the additional cost of the new procedure.

The second economic concept in judging whether a technology is good value for money is the opportunity cost. The concept implies that if resources are used in one way, an opportunity to provide some other benefit has to be renounced. To illustrate
this concept, take the example of a health policy-maker deciding on whether to provide infertility patients with free assisted reproduction services. The issue is not simply about having enough budget. There are other health services which can be “bought” with the same level of resources. The issue is what opportunities the policy-maker will miss if resources are allocated to this service.

**Is the technology culturally and ethically acceptable?**

The next question to address is whether the health technology is culturally and ethically acceptable. The assessment has to be done in the context of each country and religion. This question is particularly important in reproductive health technologies. Assisted reproduction technologies and fertility control technologies are such examples.

**Are the system requirements available?**

“System requirements” have to be carefully checked before any new technology is considered. This term is used in computer jargon. If we want to install a new software program on the computer, we are asked to check that the system requirements are available, in terms of operating system, free memory, etc. If we do not have the system requirements and we still try to install the software, the attempt will be rejected. New health technologies have system requirements, in terms of facilities, qualified and trained personnel, maintenance and supply logistics. If we try to install a new technology where the system requirements are lacking, it will not be rejected by the system, but it will not perform as desired, and may even do more harm than good, wasting resources in the process.

Social concerns are often expressed about the proliferation of new health technologies. Health professionals need to be socially conscious and fully aware of these concerns. There is concern that the proliferation of health technologies is getting out of hand, contributing to escalating and soaring costs of health care. There is concern that the health divide between rich and poor may widen, if the new technologies are more responsive to the needs of the rich and are available only to those who can afford their high cost. Then, there is the concern that medicine may be moving too far away from its social roots, and that health professionals are becoming technicians rather than humane physicians. Hippocrates wrote in about 400 BC: “Whoever wishes to investigate medicine properly should proceed thus: in the first place to consider the seasons of the year. Then the winds ... In the same manner, when one comes into a city in which he is a stranger, he should consider its situation, the water which the inhabitants use ... and the mode in which the inhabitants live, and what are their pursuits.” Now medical teachers advise whoever wants to investigate medicine properly to study molecular biology, perhaps forgetting in the process that these molecules and cells make up a human being
with a social life of her or his own. Machines now stand between doctors and patients. With the obsession with the “technology fix”, the humane physician may be in danger of becoming one day an endangered species (Fathalla, 2000).

14.5 Assessment and evaluation by investors in research

Investors in health research expect a return on their investment. It is inevitable that the unpredictable nature of much scientific research should invite questions about value for money. A commitment to evaluation and accountability on the part of the scientific community is fundamental if science is not to be marginalized in the public and political agendas. Research is an investment.

Three approaches can be pursued and are being used to evaluate the return on the investment in research: impact on advancement of science, impact on health promotion, and impact on wealth creation.

Impact on the advancement of science

Investment in research may be evaluated on the basis of the quantity and quality of the scientific output. These are the criteria commonly used for the evaluation of researchers and scientific institutions. Governments, on the basis of such measures, may allocate funding. Computers now allow bibliometric analysis to provide measurement of publication outputs. Scientific quality is generally based on originality of the subject, thought and method. Quantitatively, it may be measured as the contribution to the advancement of science, reflected on the number of times a paper has been cited as a reference by subsequent authors. This information is readily available from the Science Citation Index (SCI), produced by the Institute of Scientific Information (ISI) (www.isinet.com/isi/products/citation/sci/). The journal in which the paper has been published also matters. Journals are assigned “impact factors”. The impact factor measures the frequency with which the “average article” in a journal has been cited in a particular year or period. It provides a way to judge the prestige and influence of a particular journal.

One of the primary objectives of research is to advance science. Science is advanced step by step, through the research efforts of successive investigators. From this perspective in the scientific community, the impact of research is not only about how widely it is disseminated and read; the impact is also about how much it contributes to the advancement of science by being used in subsequent work of other researchers.

Scientific journals are not ranked by scientists according to their circulation but by their impact factors. The impact factor for a journal is calculated by the Science Citation Index (Institute of Scientific Information www.isinet.com). Journal Citation Reports
calculate the number of times that articles from the journal have been cited during the previous two years divided by the total number of articles published by the journal during this period. The impact factor gives a clue to its relative intellectual influence. Some Journals with high impact factors have relatively small circulation. For example, the journal Nature has a circulation of about 30,000 and an estimated impact factor of 25; the Journal of the American Medical Association has a circulation of about 370,000 and an impact factor less than 7 (Byrne, 1998). The contribution of a scientist to the advancement of science is measured not by the number of publications, but by the impact of these publications. The impact of the publications is assessed indirectly by the impact factors of the journals in which they were published and by subsequent citation of the articles by other authors. Citation analysis tells us that between a third and one half of published papers are never cited even once in subsequent reference lists (Lock, 1984). Many articles are hardly read at all.

Too much emphasis has been put on impact factors, and this emphasis has several drawbacks (Seglen, 1997). The impact may be technically unrelated to the scientific quality of the publication. It should also be noted that citation impact increases as one moves from clinical to basic research (Dawson et al, 1998). Assessment of the impact factor does not do justice to areas of research directly applicable to improvement of health.

Impact on health promotion

The main aim of health research is to improve the health of the people. Scientific quality and impact on health do not always go together. Much research that scientists may judge to be of high quality has no measurable impact on health, often because there may be decades before it has an impact. In contrast, research that may not be judged as high quality by scientists, because of its lack of glamour, may have immediate health benefits, if it has important health policy implications. Evaluation of the investment in research, in terms of impact on health promotion, is not easy. However, this is not a reason for not doing it, with the application of qualitative as well as quantitative methodologies. It is needed and it is necessary for public and not-for-profit private investors in research.

In the evaluation of the impact of research on health promotion, there is an economic return, which should not be undervalued. Human lives are saved, and a human life has monetary worth, in its impact on economic productivity. Health is wealth. What may not be generally appreciated is that there are savings for the health service by using appropriate technologies and discarding ineffective procedures or interventions, and rational allocation of resources. Expenditure on research by the UK National Health Service (NHS) has been estimated to be more than 400 million pounds sterling every year (Wellcome Trust, 2000). In justifying a relatively high level of expenditure on health research, the NHS affirmed the truism that publicly funded research is as important in
the NHS to enable managers to save money, as it is in industry for making money on new products and services (Royal College of Pathologists, 1996).

Mary Lasker, a well-known philanthropist who played a central role in the rapid expansion of medical research and public health in the USA, has been quoted as saying “If you think research is expensive, try disease”. In 1999, the Lasker Foundation, through its Funding First initiative, asked nine academic economists from the universities of Chicago, Columbia, Harvard, Stanford and Yale to focus on the economic value of the increase in life expectancy and the impressive decline in mortality. The report “Exceptional returns: the economic value of America’s investment in medical research” (http://www.laskerfoundation.org/reports/pdf/exceptional.pdf) estimated the increase in life expectancy in the United States between 1970 and 1990 to be worth roughly US$ 2.8 trillion a year. Reduced mortality from cardiovascular disease alone was estimated to be worth US$ 1.5 trillion a year. Even if only a small percentage of this gain is attributed to advances in research, the return on the research investment would be enormous.

There are also cost savings to the health service, as a result of properly conducted health research. Cost savings include money saved from hospitalization avoided, and from production work gained, from medical procedures not required. For example, preventing hip fractures in postmenopausal women at risk of osteoporosis can save hundreds of millions of dollars annually in treatment costs, apart from loss of productivity. One study in the USA indicated that for every dollar invested throughout the public and private sectors, there was a return of at least three to one from cost savings alone (Rosenberg, 2002).

**Impact on wealth creation**

Health research may be viewed as an engine for economic growth in developed and also recently in some developing countries. The health industry is one of the fastest growing industries, and one of the most profitable. It has been estimated that companies in the health care market place contribute about 5% of the gross development product in the UK, and generate a trade surplus of some 2 billion pounds sterling (Royal College of Pathologists, 1996). Job creation in the private sector is another parameter. It has been estimated that there are more than 500 000 people employed in the US biopharmaceutical industry because of commitments to research and development (Rosenberg, 2002). These high-paying employment opportunities would not have existed if government was not priming the scientific pump by supporting research.

Governments encourage and support basic research that can provide promising leads for discovery, innovation and wealth creation. For impacts on wealth creation, patent citation indicators have been used to evaluate the investment in research. US patents cite papers as “prior art”, that is, the research that has formed the basis for the development of a new and novel product. The Wellcome trust, for example, maintains TechTrac, an
in-house database to link publications in the UK with the US patent prior art information (Dawson et al., 1998).

The importance of health research for development has received increasing international attention over the past 10–20 years. In October 2000, an International Conference on Health Research for Development was convened in Bangkok, co-sponsored by the Council on Health Research for Development, the Global Forum for Health Research, the World Bank and the World Health Organization. The Conference issued a declaration (Annex 5). A ministerial summit on health research is planned by WHO for November 2004 in Mexico.

References and additional sources of information


