What does it mean to have enough evidence?

Glenda MacQueen, MD, PhD

From the Department of Psychiatry, Mathison Centre for Mental Health, Research and Education, Hotchkiss Brain Institute, University of Calgary, Calgary, Alta.

"One thing is sure. We have to do something. We have to do the best we know how at the moment..."
Franklin D. Roosevelt

"To get anywhere, or even to live a long time, a man has to guess, and guess right, over and over again, without enough data for a logical answer."
Robert Heinlein

Clinicians treating patients and scientists planning studies might agree with the words of Roosevelt: we have to do something. Most clinicians and scientists would probably also agree with the idea implied in his words: that we are often working with imperfect or incomplete evidence bases. The medical sciences have long struggled with how to ensure clinicians have the best evidence on which to make decisions in "the moment." Questions about how we make evidence-based decisions lead to a series of issues. What constitutes "evidence?" How is a decision made regarding what is the best evidence or what is sufficient evidence? How can we determine whether evidence is accurate? And, possibly most disconcerting, how inviolable is the notion of the "truth" in science? Does that not imply a view of science that has itself been disputed by philosophers of science? It is not feasible to discuss these provocative questions in detail here; however, it is possible to examine briefly some trends in our current approach to accumulating and evaluating evidence in psychiatry and neurosciences.

Practitioners of modern Western science have accepted a specific methodological framework for generating and defining what qualifies as evidence. A hierarchy of methodologies has emerged in which the generation of evidence generally involves the construction of a hypothesis that is then tested and evaluated with a formal attempt to disprove it. In clinical medicine, this often involves some form of clinical trial, and in the hierarchy of clinical trials, randomized controlled trials (RCTs) are ranked highest for their ability to reveal to clinicians whether a specific therapy is likely to be helpful for a specific condition. This is not to say that RCTs are without criticism. On the contrary, they have been criticized on many levels and for having many flaws.[1] Others have suggested that psychiatry in particular is a field in which the limitations of RCTs impact optimal patient care.[2]

Combined or pooled analyses of clinical trials have emerged as methods that generally reflect an even higher quality of evidence than a single RCT. Shortly after meta-analyses began to be widely used, the approach was even described as an "ethical imperative."[3] The meta-analytic approach was described as "more accurate, comprehensive, systematic and statistically powerful."[3] In 2012, meta-analytic reviews are certainly thriving in psychiatry and clinical neurosciences. The vast majority of meta-analyses in psychiatry have been published in the 20 years since the Cochrane Collaboration was initiated.

A nonspecific (low-quality) search in PubMed conducted in September 2012 using the search term "schizophrenia" with the limit "meta-analysis" yielded 760 results. All but 7 of these (753) were published in the last 20 years. Searching the term "depression" with the limit "meta-analysis" yielded an impressive 1638 citations, although using the term "major depression" reduced the number to 973; 1618 and 958 of these, respectively, were published in the last 20 years. Searching the term "bipolar disorder" yielded a modest 260 citations, with all but 4 published in the past 20 years, while "posttraumatic stress disorder" yielded 135, "panic disorder" yielded 64 and "generalized anxiety disorder" yielded 58. In total, there have been more than 2000 meta-analyses published on various aspects of 6 common psychiatric illnesses in the past 20 years, which translates to more than 100 per year. This quick search did not include meta-analyses related to addictions, eating disorders, neuropsychiatric conditions or any other DSM-IV psychiatric illnesses that may have been subject to the same analytic approach. An important limitation to note is that not all citations were obtained and searched to confirm that they followed meta-analytic techniques. Some studies almost certainly did not. Prior to the publication and adoption of the QUORUM and MOOSE

Correspondence to: G.M. MacQueen, Department of Psychiatry, Foothills Hospital, 29th St. NW, Calgary AB T2W 2T9; gmmacque@ucalgary.ca

J Psychiatry Neurosci 2013;38(1):3-5.

DOI: 10.1503/jpn.120240

guidelines, specifying the use of a meta-analytic technique in the article title was sporadic and made it more difficult to ascertain whether a study did, in fact, use a meta-analytic technique. On the other hand, it is apparent that the estimate of 2000 published meta-analyses is very conservative and leaves a wide margin for error in the search.

Meta-analytic techniques have been used more frequently to examine questions other than those related to clinical trials and treatment. For example, a low-quality search of PubMed in September 2012 using the term "neuroimaging" with a "meta-analysis" limit yielded 306 citations. "Neuroimaging" and "major depression" with a "meta-analysis" limit yielded 10 citations published in 2011 and 2012; a similar search that substituted "schizophrenia" yielded 9 citations published in the same time frame.

Examining biologically focused journals suggests that at least a dozen meta-analyses were published in *Biological Psychiatry* between January 2011 and September 2012. About 8 were published in each of *Molecular Psychiatry*, *Archives of General Psychiatry* and *World Journal of Biological Psychiatry* during the same period. Similarly, 10 or so were published in the more clinically focused *American Journal of Psychiatry*, and about twice as many were published in the *Journal of Clinical Psychiatry*. This suggests that even journals that are focused on more fundamental biological mechanisms are publishing a meta-analytic report every couple of months.

This proliferation of meta-analytic reviews should provoke several questions. Is there really so much evidence accumulating in psychiatric research that we need more than 100 meta-analyses each year to synthesize new research in even a few diagnostic areas? Are so many meta-analyses required because clinicians and scientists value these reviews, or paradoxically, because the meta-analysis, like the RCT before it, is not considered "sufficient" evidence on which to base a clinical decision? From an editorial perspective, when is a meta-analysis "novel?" It may be difficult to ascertain whether a sufficient amount of new information has become available to justify the publication of an updated meta-analysis. It seems unlikely that an update should be published only if the conclusions of a new meta-analysis are substantively different from those of a prior one. Conclusions may differ because there is new primary data to include or because a different methodological approach has been adopted (e.g., a different set of eligibility criteria). It can be particularly difficult to judge the importance of a new meta-analysis when the conclusions are similar to previous ones. If the new analysis has included a substantive amount of new information, it may be helpful for readers to have that new information aggregated and analyzed, but how many new studies constitute the need for an update?

Readers have become more sophisticated in reading — and consequently less accepting of — meta-analyses. This attitude has been shaped in part by a relatively extensive series of general and specific publications discussing the overall principles underlying meta-analysis and the limitations in the methodologies (for example, see Egger and colleagues,[4] Bailar,[5] Feuer,[6] Simon[7] and Walker and colleagues[8]). There are complex issues with study selection,[9,10] publication bias,[11,12] control for heterogeneity,[12] analysis method[13] and interpretation and reporting[14,15] that are well beyond the scope of this discussion. Widely discussed meta-analyses in areas such as the treatment of depression have brought these methodological questions to the forefront of psychiatry, as clinicians have recognized that conclusions from meta-analyses may be widely cited to support the utility of one treatment over another.[16]

> "It is hard to imagine a more stupid or more dangerous way of making decisions than by putting those decisions in the hands of people who pay no price for being wrong."
> *Thomas Sowell*

In possible contradiction to Sowell's provocative statement (which, it should be noted, was not aimed at those evaluating medical literature), it could be argued that the systematic review of evidence is best placed in the hands of unbiased methodologists. Indeed, the systematic evaluation of medical evidence has grown far beyond small groups of investigators conducting individual meta-analyses on topics of interest and relevance to their own research areas. There are large agencies, both publicly funded and otherwise, that specialize in reviewing medical literature. The Cochrane Collaboration has an international reputation for conducting systematic reviews, more than 5000 of which are now published in the Cochrane library. Of note, in September 2012, the Cochrane library listed 361 of these under the label "mental health," 117 under "tobacco, drugs and alcohol dependence" and 102 under "developmental, psychosocial, & learning problems."

The Agency for Healthcare Research and Quality (AHRQ) also supports large systematic reviews, including comparative effectiveness reviews. The AHRQ has a number of evidence-based practice centres that bring together content-specific and methodological experts to coordinate these large reviews. It also includes the John M. Eisenberg Center for Clinical Decisions and Communications Science, which translates results from reviews and other syntheses into tools for consumers, clinicians and policy-makers. Consistent with other groups, the AHRQ-supported reviews use the GRADE tool to evaluate the quality and strength of evidence supporting the conclusions that emerge from these reviews.[17,18]

Other agencies conduct similar large-scale, standardized reviews as part of health technology assessments, which may include systematic reviews. For example, a recent review examined the clinical effectiveness and cost-effectiveness of psychological interventions in preventing relapse after depression.[19] This review, however, highlights one of the difficulties with this approach. The report is more than 100 pages long, with superb detailing of the search strategies, included and excluded studies and data extraction methods. More than 9000 studies were identified by the initial search. Unfortunately, however, no studies met the criteria for the clinical effectiveness component of the search, and the conclusion of the review was that there was inadequate evidence to determine whether the interventions of interest are clinically effective.[19] Even the best and most systematic reviews are limited by the availability of primary literature on the question of interest. Considering the effort and investment involved in some of the larger reviews, it is easy to wonder whether a

preliminary screen could determine whether the primary evidence is extensive enough to predict the probability of a clinically relevant conclusion. The ubiquitous conclusion that more research is needed is of arguable benefit to a field in which most practitioners would likely note that they knew that already. Perhaps reasonable questions for the average practitioner are "What have any of these reviews done for me lately?" or "Are the investments in these reviews actually leading to improved quality of care?"

Meta-analyses, systematic reviews, comparative effectiveness reviews and health technology assessments can form the basis that supports the creation of clinical practice guidelines. The numbers of guidelines on any given topic have also increased notably in the past few years. For example, between 2005 and 2010 there were about 25 guidelines on the treatment of major depression published in English.[20] Using AGREE criteria to evaluate guidelines,[21] most were of moderate quality. Having 25 guidelines published on the same illness in a 5-year period raises an editorial dilemma: how to judge the novelty, impact and significance of the next guideline in the context of the extant literature. What constitutes an important update? When it is necessary or appropriate to update a guideline? These are complex questions, and models have been proposed to assess how much new data should be available to justify the process.[22]

The concept of evidence-based medicine — and more generally, evidence-based decision-making — makes sense to many physicians and scientists trained in the last few decades. In addition to the methods involved in primary data acquisition, a parallel discipline has emerged that is focused on the methodology of data identification, synthesis and reporting, necessary steps in the path to dissemination and translation of raw data into clinical practice or policy change. An increasing percentage of published work in psychiatry and neuroscience is focused on reporting aggregate or synthesized data; the editorial challenge to is to evaluate the contribution of these studies to the overall advancement of the field.

Competing interests: G. MacQueen declares that her institute has received grant funding from AstraZeneca, AllerGen NCE Inc. and the Scottish Rite Foundation and that she has received consultancy fees from Pfizer, Lundbeck, BMS, Servier, Sepracor, AstraZeneca and Lilly, lecture fees from the Norlien Foundation and AstraZeneca, and both lecture fees and payment for development of educational presentations from the Canadian Psychiatric Association.

References

1. Tharyan P. Evidence-based medicine: Can the evidence be trusted? *Indian J Med Ethics*. 2011;8:201-7.
2. Sobo S. Pursuing treatments that are not evidence based: how DSM IV clarifies, how it blinds psychiatrists to issues in need of investigation. *Med Hypotheses* 2009;72:491-8.
3. Rosenthal R. Science and ethics in conducting, analyzing and reporting psychological research. *Psychol Sci* 1994;5:127-34.
4. Egger M, Davey Smith G, Philips A. Meta-analysis. Principles and procedures. *BMJ* 1997;315:1533-7.
5. Bailar JC. The problems and promise of meta-analysis. *N Engl J Med* 1997;337:559-61.
6. Feuer DJ. Meta-analysis. *Palliat Med* 1999;13:433-7.
7. Simon R. Meta-analysis of clinical trials: opportunities and limitations. In: Stangl DK, Berry DA, editors. *Meta-analysis in medicine and health policy*. New York: Marcel Dekker; 2000. p. 273-85.
8. Walker E, Hernandez AV, Kattan MW. Meta-analysis: its strengths and limitations. *Cleve Clin J Med* 2008;75:431-9.
9. Dickersin K, Scherer R, Lefebvre C. Systematic reviews: identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286-91.
10. Steinbrook R. Searching for the right search — reaching the medical literature. *N Engl J Med* 2006;354:4-7.
11. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
12. Terrin N, Schmid CH, Lau J, et al. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22:2113-26.
13. Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare events: the rosiglitazone link with myocardial infarction and cardiac death. *Stat Med* 2007;26:4375-85.
14. 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: 1896-900.
15. Stroup DF, Berlin JA, Morton SC, et al. Meta-analyses of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283:2008-12.
16. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373:746-58.
17. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. The GRADE Working Group. *BMC Health Serv Res* 2004;4:38.
18. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6.
19. Rodgers M, Asaria M, Walker S, et al. The clinical effectiveness and cost-effectiveness of low-intensity psychological interventions for the secondary prevention of relapse after depression: a systematic review. *Health Technol Assess* 2012;16:1-130.
20. Santaguida PL, MacQueen G, Keshavarz H, et al. Treatment for depression after unsatisfactory response to SSRIs. Comparative Effectiveness Review No. 62. Prepared by McMaster University Evidence-based Practice Center under Contract No. HHSA 290-2007-10060-I. AHRQ Publication No.12-EHC050-EF. Rockville (MD): Agency for Healthcare Research and Quality; April 2012. Executive summary available: http://effectivehealthcare.ahrq.gov/ehc/products/156/1036/CER62_SSRI_execumm.pdf (accessed 2012 Nov. 26).
21. Brouwers M, Kho ME, Browman GP, et al.; AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *CMAJ* 2010;182:E839-42.
22. Shekelle P, Woolf S, Grimshaw JM, et al. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. *Implement Sci* 2012;7:62.