

## Research

## Developing the evidence for evidence-based practice

A. John Rush MD

∞ See related article page 1293

In this issue of *CMAJ*, Deshauer and colleagues<sup>1</sup> note that most (93%) trials of drugs for the treatment of depression last less than 6 months (typically 6–8 weeks); indeed, most are conducted for registration purposes by industry. Despite the widely accepted view that depression is chronic or recurrent, longer-term efficacy and effectiveness studies are few and far between. Thus, when asked by patients about the pros and cons of longer-term treatment, we have little evidence with which to respond.

Clinical trials are complex, costly and time consuming. They can be roughly divided into “efficacy” trials and “effectiveness” trials. The former are usually designed to have the highest internal validity, thereby ensuring that differences between treatment groups are entirely attributable to the study treatments (e.g., drug v. placebo). These trials typically look for a signal that the treatment is better than placebo and establish safety and tolerability. Effectiveness trials (sometimes called practical or management trials) are more inclusive of patient groups and often use treatment conducted in routine practices rather than research-guided treatment methods. They often aim to define how the treatment performs in usual practice conditions, but these trials may also define how, for whom, when or in what setting a treatment is to be recommended. Effectiveness trials include a broad assessment of effectiveness, including outcomes such as daily function, quality of life and health care utilization, whereas efficacy trials typically focus on symptoms.

Furthermore, short-term efficacy trials focus on a narrowly defined group of outpatients, typically recruited as volunteers with depressive symptoms and few or no concurrent psychiatric or general medical comorbidities (e.g., anxiety disorders, substance abuse, heart disease).<sup>2</sup> These trials also usually exclude patients who have experienced the index major depressive episode for more than 2 years and patients who have not benefited adequately from more than 1 prior treatment attempt. In addition, the procedures used to deliver treatment in typical practices often differ greatly from those used in efficacy trials (e.g., in efficacy trials, symptoms and side effects are routinely measured at each visit, which informs clinical decisions about dosing, and the visits are often longer and more frequent). Finally, short-term efficacy trials often choose outcomes to detect an efficacy “signal” (e.g., a change in symptom severity over 8 weeks), although recently these trials have begun to also report clinically relevant out-

## Key points

- Results of short-term efficacy trials of depression have minimal relevance to most patients with depression, who have chronic or recurring episodes of depression.
- Most trials establish drug safety and efficacy only in a narrowly defined group of patients under atypical treatment conditions.
- Information on how, when and for whom a treatment is to be recommended in usual practice is unknown.
- Clinically relevant evidence to inform decision-making is needed.

comes such as response or remission. Response refers to a clinically meaningful degree of improvement (e.g., a 50% reduction in severity of depressive symptoms), and remission refers to the virtual absence of such symptoms. Remission is the preferred, most valid and most clinically relevant outcome, because it is associated with a return to normal function<sup>3</sup> and a better prognosis compared with response.<sup>4,5</sup>

Although these first-step, short-term efficacy trials may be essential, they provide little information to clinicians or patients to inform clinical decisions. For example, we lack information about which patients will benefit most from a particular treatment, or how long to try a medication to determine if it will work or not, or whether the maximum benefit has been achieved, or whether longer-term treatment is called for and, if so, for whom.

A third type of trial — so-called hybrid trials — attempt to control some of the parameters that affect outcome while allowing others to vary. For example, the Sequenced Treatment Alternatives to Relieve Depression (STAR\*D) trial<sup>6–8</sup> involved a wide range of patients with depression recruited from diverse practices to enhance generalizability. This trial included multiple outcomes (e.g., symptoms, function, health care utilization), and it included commonly used treatments. In these respects, it was an effectiveness trial. On the other hand, the

John Rush is with the Departments of Clinical Sciences and Psychiatry, the University of Texas Southwestern Medical Center at Dallas, Dallas, Tex.

Une version française de cet article est disponible à l'adresse [www.cmaj.ca/cgi/content/full/178/10/1313/DC1](http://www.cmaj.ca/cgi/content/full/178/10/1313/DC1)

All editorial matter in *CMAJ* represents the opinions of the authors and not necessarily those of the Canadian Medical Association.

researchers controlled the quality of treatment (i.e., how treatment was delivered) to ensure that, if a treatment failed, the failure was because of the treatment itself and not because of poor delivery of an otherwise effective treatment. In these respects, the trial had elements of an efficacy trial, although treatment was conducted in actual practices, not in research clinics. Randomization was used to determine which of several next-step treatments would be best if the first or subsequent treatments failed.<sup>5,7,9-12</sup>

Results of the STAR\*D trial underscore the conclusions drawn by Deshauer and colleagues. Namely, short-term trials (6-8 weeks) are too brief to gauge the full proportion of patients whose symptoms will respond or remit.<sup>6-8</sup> Fully half of the patients who reached remission did so after 6 weeks of a single treatment; a full third did not even respond until after 6 weeks.<sup>8</sup> Consonant with the review by Deshauer and colleagues, longer-term outcomes were less robust than the short-term results, especially among patients who had more prior failed treatment attempts<sup>5</sup> or selected psychiatric or general medical comorbidities (unpublished data).

Effectiveness or hybrid studies are often encompassed within the concept of "T<sub>2</sub> translational research." This research focuses on how safe and effective the treatment is in representative practice as well as on questions about the optimal application and use of the treatments: for whom (e.g., which patients, defined by comorbidities or depressive "subtypes," age, sex), when (e.g., in the course of illness or in a sequence of treatment steps), how (e.g., dosing, duration, dose adjustment) and how long should a treatment be used, and to what ends in terms of benefits and costs?<sup>13</sup> In contrast, "T<sub>1</sub> translational research" refers to the development and application of basic science findings with the aim of expanding our understanding of the pathobiology of the disorder and developing candidate treatments to address the newly defined pathobiological pathways.<sup>13</sup>

The review by Deshauer and colleagues underlines the fact that, following drug-registration trials, many critical clinical questions remain unanswered, which indicates a scarcity of evidence from T<sub>2</sub> translational research. These questions include: For which patients is the treatment preferred? Is the treatment safe and effective for patients with various comorbidities? Can the drug be combined with other antidepressants for partial responders? How long should it be tried before one knows whether it will work fully or whether other treatments are needed? Is the drug effective as a second or third treatment step? And, perhaps most importantly, is the drug effective in the longer term (beyond 6 months) in terms of clinically important outcomes such as function, quality of life, work performance and health care utilization.

To answer these T<sub>2</sub> research questions, one needs large numbers of representative patients given representative or feasible but high-quality care ("measurement-based care"<sup>8,14</sup>) for the longer term (at least 6 months) to provide sufficient power and generalizability as well as to allow for moderator analyses that can potentially inform clinical decisions.<sup>15-17</sup> Moderator analyses can identify which patients are best served by longer-term treatment on the basis of outcomes that matter to clinicians, patients and other stakeholders

(e.g., sustained remission, function, quality of life and work performance). To accomplish these ends, practical clinical trials offer a cost-efficient approach, can engage sufficient numbers of patients and, because these trials are broadly inclusive, can determine through moderator analyses whether specific subgroups benefit from the treatment.<sup>18</sup> Systems of care with common electronic medical records provide efficient venues for practical clinical trials. Such studies also provide opportunities for genome-wide scans or other biomarker assessments that may better assist us in matching patients and treatments.

It is especially important to identify patient groups for which the treatment is not effective or safe. The review by Deshauer and colleagues suggests that patients with depression who have some types of general medical comorbidities may not benefit from treatment over the longer term. If this is established as a fact, not only could such patients avoid an ineffective, burdensome and expensive treatment, but researchers could focus on why this occurs and attempt to develop better treatments for these patients.

Defining how, when, for whom and in what settings available treatments are best, and how safe and effective new treatments are in representative practice (T<sub>2</sub> research) cannot be the sole responsibility of industry. Why not? Because such research will potentially limit the use of a new drug or create data for counter-marketing by competitors. Companies logically want to control study designs to protect their products. Who can blame them? Indeed, even if these studies are well designed and executed, doubt may remain about the findings, given the source of funding.

Independent sources of funding (e.g., government and foundations) are needed if we are serious about putting real evidence into evidence-based medicine. Definitive, generalizable studies that involve "real world" patients given treatment under clinically feasible conditions and that use both registries and randomization are needed to better inform us about patient safety, provide evidence for clinical decision-making and improve outcomes. Without such studies, clinical decision-making will remain the art of medicine rather than the science it should be.

**Competing interests:** John Rush has received research support from the National Institute of Mental Health, the Robert Wood Johnson Foundation and the Stanley Medical Research Institute. He has been on the advisory boards of, and a consultant to, Advanced Neuromodulation Systems, Inc., AstraZeneca, Best Practice Project Management, Inc., Bristol-Myers Squibb Company, Cyberonics, Inc., Eli Lilly and Company, Gerson Lehman Group, GlaxoSmithKline, Jazz Pharmaceuticals, Magellan Health Services, Merck & Co., Inc., Neuronetics, Ono Pharmaceutical, Organon USA Inc., PamLab, Personality Disorder Research Corp., Pfizer Inc., The Urban Institute and Wyeth-Ayerst Laboratories Inc. He has been on the speaker's bureau for Cyberonics, Inc., Forest Pharmaceuticals, Inc., and GlaxoSmithKline. He has had equity holdings (excluding mutual funds and blinded trusts) in Pfizer Inc. and has royalty income affiliations with Guilford Publications and Healthcare Technology Systems, Inc.

## REFERENCES

1. Deshauer D, Moher D, Fergusson D, et al. Selective serotonin reuptake inhibitors for unipolar depression: a systematic review of classic long-term randomized controlled trials. *CMAJ* 2008;178:1293-301.
2. Depression Guideline Panel. *Clinical practice guideline, number 5: depression in*

- primary care. Volume 2. Treatment of major depression. Rockville (MD): US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993.
3. Miller IW, Keitner GI, Schatzberg AF, et al. The treatment of chronic depression, part 3: psychosocial functioning before and after treatment with sertraline or imipramine. *J Clin Psychiatry* 1998;59:608-19.
  4. Rush AJ, Kraemer HC, Sackeim HA, et al. Report by the ACNP Task Force on Response and Remission in Major Depressive Disorder. *Neuropsychopharmacology* 2006;31:1841-53.
  5. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry* 2006;163:1905-17.
  6. Koran LM, Gelenberg AJ, Kornstein SG, et al. Sertraline versus imipramine to prevent relapse in chronic depression. *J Affect Disord* 2001;65:27-36.
  7. Rush AJ, Trivedi MH, Wisniewski SR, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med* 2006;354:1231-42.
  8. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *Am J Psychiatry* 2006;163:28-40.
  9. Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med* 2006;354:1243-52.
  10. Fava M, Rush AJ, Wisniewski SR, et al. A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR\*D report. *Am J Psychiatry* 2006;163:1161-72.
  11. McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR\*D report. *Am J Psychiatry* 2006;163:1531-41.
  12. Nierenberg AA, Fava M, Trivedi MH, et al. A comparison of lithium and T<sub>3</sub> augmentation following two failed medication treatments for depression: a STAR\*D report. *Am J Psychiatry* 2006;163:1519-30.
  13. Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008;299:211-3.
  14. Trivedi MH, Rush AJ, Gaynes BN, et al. Maximizing the adequacy of medication treatment in controlled trials and clinical practice: STAR\*D measurement-based care. *Neuropsychopharmacology* 2007;32:2479-89.
  15. Kraemer HC, Stice E, Kazdin A, et al. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry* 2001;158:848-56.
  16. Kraemer HC, Wilson GT, Fairburn CG, et al. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 2002;59:877-83.
  17. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. *JAMA* 2006;296:1286-9.
  18. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003;290:1624-32.

---

**Correspondence to:** Dr. A. John Rush, Department of Clinical Sciences, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas TX 75390-9066, USA; fax 214 648-6863; john.rush@utsouthwestern.edu