Major Limitations of Randomized Controlled Trials Assessing Antidepressant Efficacy

Anton J.M. Loonen, MD, PharmD, PhD

In a review article in this issue of the journal, Kevin P. Kennedy and colleagues argue that there is insufficient evidence from randomized clinical trials (RCTs) to support the various steps in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial. They caution against applying the results of pragmatic studies to clinical practice and advocate for conducting RCTs with alternative medications.

It has now been more than 40 years since evidence-based medicine (EBM) began its rise. "Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."[1] This approach replaced the 2-and-a-half-millennium-old (Hippocratic) strategy of training physicians in a guild-like tradition. Over the past few centuries, this was transformed into the medical university model, in which leading professors developed their own school of thought. In retrospect, much of what was taught there by these now-extinct teachers has proven invalid; it is good that a different approach has been widely adopted in medicine. However, as with any movement, some EBM proponents take things too far. Generally, strong evidence stems from randomized controlled trials (RCTs), preferably double-blind and placebo-controlled. Many epidemiologists classify evidence into hierarchical levels, with systematic reviews and meta-analyses of multiple high-quality RCTs representing the highest standard. Kennedy and his colleagues (2025) are among them.[2] In a comprehensive review article, they summarize the results of RCTs—with or without inclusion in meta-analyses—that, in their view, should provide evidence for the various steps in the pragmatic STAR*D trial, which has shown value in treating patients with depression. They conclude that the current evidence does not support the appropriateness of the STAR*D trial's treatment schedule, urge caution in interpreting the results of pragmatic trials, and advocate for more placebo-controlled studies of new or alternative drugs. The present author draws a diametrically opposite conclusion from the same material. Apparently, the prevailing model evaluating antidepressant efficacy is not suited for determining the best treatment for individual patients. This model, adapted from drug regulatory procedures, involves selecting patients who respond to a reference compound (eg, imipramine, as measured by a reduction in HAMD score) and testing whether the investigational drug elicits a similar effect in such a cohort and setting. These RCTs aim not for remission, but only for symptom improvement. They may also diminish confidence in treatment, leading to early dropout among participants. Thus, these RCTs may lack relevance for the clinical treatment of depression, as the STAR*D trial demonstrated that patients can benefit from treatments that do not produce statistically significant results in conventional RCTs.

It is well known that RCTs have many limitations. In general, the most important limitations stem from the reciprocal relationship between a study's internal and external validity. The more tightly research methods are standardized and aligned with a specific hypothesis, the more generalizability is lost. You may be able to show very precisely that a particular drug is more effective than a placebo in a specific patient population, but this does not reveal how it performs in typical clinical practice. It's like losing your glasses during an evening walk. You may determine with certainty that they are not under the lamppost, but they could still be elsewhere along the path, just like a drug might be effective in practice, though not in a narrowly defined trial.

Fortunately, humans also rely on life experience. For example, we do not need an RCT to prove that drinking sweet water quenches thirst; humans and animals already know this intuitively. Kennedy and colleagues point out that remarkably few RCTs have examined the effects of lithium salts as an adjunctive therapy for depressive disorders; a 2019 meta-analysis identified only 12 such placebo-controlled trials. Despite known challenges with blinding and monitoring serum levels, lithium has been extensively studied and used for more than 75 years.

Received for publication June 5, 2025; accepted July 13, 2025.
From the Unit of Pharmacotherapy, - Epidemiology & -Economics, Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands.
Address correspondence to: Anton J.M. Loonen, MD, PharmD, PhD, HPC: XB45, Antonius Deusinglaan 1, Groningen 9713AV, The Netherlands (e-mail: a.j.m.loonen@rug.nl).

ISSN: 0271-0749
DOI: 10.1097/JCP.0000000000002075

Thus, repeated, large-scale RCTs may be unnecessary for this indication. Clinical observations and prior data from other conditions may suffice, especially when corroborated by a physician's experience.

Apparently, Kennedy and colleagues believe that each step in the STAR*D protocol should be supported by RCT evidence. By this logic, even the initial step of increasing the dose in nonresponders must be validated by RCTs. However, published RCTs have not demonstrated that antidepressants are more effective at higher doses than at lower doses, leading these authors to conclude that dose escalation is unjustified. This assumption is flawed. Abandoning a treatment before properly completing it does not serve the patient's best interest. While some patients may benefit from a dose increase, the primary rationale is to ensure a rational, complete trial of therapy before deeming it ineffective. Whether RCTs demonstrate a dose-response relationship is irrelevant here for individualized patient care. The clinical decision to adjust dosage must be based on the therapeutic context, not population-level outcomes. These authors appear to assume that a treatment should not be used unless it has been proven to be effective through RCTs. They explicitly argue that the use of an ineffective drug could expose patients to unnecessary side effects. In the opinion of the present author, that position is incorrect in this context, as chronic depression itself inflicts substantial harm on the patient, their loved ones (eg, children), and even their role in society. A similar argument applies when comparing the side effects of higher versus lower doses. Kennedy and colleagues state that, according to meta-analyses, adverse effects increase strongly with dosage. This may be true for composite measures that aggregate all adverse effects. However, it does not necessarily apply to specific side effects. For example, the occurrence of orthostatic hypotension with imipramine is not correlated with plasma levels as a proxy for drug exposure.[3]

There is another key issue: standard RCTs may not produce accurate findings for antidepressants due to changes in the nature of depressive disorders over the past half-century. Loonen et al[4] recently argued that while some cases of depression still resemble sudden-onset, delusion-like illness, most modern presentations are more reactive and contextual, once even considered a "sin." Depressive reactions are completely normal and occur repeatedly in everyone's life. In mentally healthy individuals, natural resilience mechanisms promote rapid recovery. In depressive disorder, these mechanisms are impaired or ineffective. Paradoxically, standard RCTs deliberately suppress these natural recovery mechanisms by minimizing the likelihood of spontaneous improvement and placebo responses, while excluding other therapies. This design evaluates whether treatment with the active compound produces outcomes that differ significantly from placebo. However, it remains unknown, and is even unlikely, whether such differences reflect the pharmacologic mechanism by which antidepressants exert their actual therapeutic or prophylactic effects. Thus, this model is poorly suited to guiding clinical treatment decisions in depressive disorder. This type of RCT should, therefore, not be used as a model for treating depression. Instead, it would be more productive to investigate optimal treatment strategies, such as add-on therapies, via small proof-of-concept studies that rely on well-established interventions.

Why am I critical of the work of Kennedy and his colleagues? It is not because I dismiss the value of EBM or oppose research and scientific inquiry. Rather, I believe that standard RCTs involving antidepressants do not represent the best available evidence, and that the way these trials are applied in their framework is not sufficiently rigorous. Patients and practitioners are frequently exposed to public claims that antidepressants are ineffective or dangerous. These claims are unfounded and may lead patients to discontinue treatment, potentially causing harm. Although EBM was initially intended to reduce costs by eliminating ineffective treatments, this goal has not consistently been achieved. Newer medications are typically evaluated in well-designed RCTs, whereas older drugs that have demonstrated practical effectiveness are often underrepresented. In my view, replacing affordable, established drugs with more expensive new ones, solely because more RCTs exist for the latter, is questionable, especially when the quality of that evidence is itself uncertain.

AUTHOR DISCLOSURE INFORMATION

The author declares no conflicts of interest.

REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
2. Kennedy KP, Heldt JP, Oslin DW. What if STAR*D had been placebo-controlled? A critical re-examination of a foundational study in depression treatment. *J Clin Psychopharmacol*. 2025;45:648–661. doi:10.1097/JCP.0000000000002025
3. Glassman AH, Bigger JT Jr, Giardina EV, et al. Clinical characteristics of imipramine-induced orthostatic hypotension. *Lancet*. 1979;1:468–472.
4. Loonen AJM, Ochi T, Geers LM, et al. A new paradigm to indicate antidepressant treatments. *Pharmaceuticals (Basel)*. 2021;14:1288.