

The Goldilocks Problem:

Balancing Internal Validity With Generalizability in Clinical Trials

Lori L. Davis, MD

In this issue of JCP, Zimmerman and Snyder¹ report that 93% of patients seeking treatment for posttraumatic stress disorder (PTSD) in their outpatient practice would have been excluded from the brexpiprazole-sertraline industry-sponsored trials,^{2,3} raising important questions about the applicability and generalizability of these studies to their outpatient practice. They further discuss two parameters that would be additional barriers to enrollment: comorbidity (nearly two-thirds of their cohort had a concurrent major depressive episode [MDE]) and heterogeneity (roughly 60% had PTSD duration >10 years, and 50% had childhood onset of PTSD). These eligibility design decisions were not arbitrary but rather reflected a fundamental methodological imperative to establish a clear signal for PTSD treatment efficacy for a new pharmacotherapy.

Clinical trials face an inherent tension between internal validity and generalizability. While critics argue that strict inclusion/exclusion criteria limit real-world applicability, these design elements serve essential scientific and ethical purposes that cannot be dismissed. Internal validity requires us to clearly define the primary diagnosis under study. Without it, we cannot determine whether an intervention truly works for a specific condition.

Brexpiprazole is FDA-approved as adjunctive therapy to antidepressants for the treatment of major depression, a condition that is highly comorbid in PTSD populations. To demonstrate independent efficacy for PTSD, clinical trials require at least some separation between the diagnostic entities. Had the PTSD studies included

participants with a current MDE, the studies would have faced inevitable criticism of pseudospecificity, ie, a concern that any observed benefits might reflect treatment of depression rather than PTSD itself. In this case, internal validity would be sacrificed in the name of generalizability.

After defining the parameters of the primary diagnosis, investigators must next tackle the “Goldilocks problem”—a concept of defining the “just right” amount. This aim necessitates excluding patients who are too severely ill or possibly treatment refractory and those who are too mildly ill or prone to spontaneous recovery, as both obscure signal detection and inflate sample size requirements. This is not unique to psychiatry; in fact, all clinical research requires sample refinement to detect meaningful differences between treatment groups.

Two common reasons that would limit enrollment from the Zimmerman and Snyder sample in the brexpiprazole PTSD trials relate to PTSD age of onset and chronicity. These exclusion criteria were grounded in literature on treatment response variability in PTSD. Childhood-onset PTSD, particularly related to early-life sexual trauma, represents clinically distinct treatment responses compared to adult-onset PTSD.^{4,5} Similarly, chronic PTSD, particularly when present for more than a decade, has been associated with smaller effect sizes in PTSD pharmacotherapy trials.⁶ The fact that many of the trial participants had not received prior PTSD treatment does not necessarily indicate that they were less treatment-seeking or less impaired than typical clinic patients. Many individuals with PTSD avoid

mental health treatment due to PTSD-related avoidance symptoms, stigma, or lack of awareness that effective treatments exist.⁷

The clinical reality is that many PTSD patients do experience comorbid depression, have experienced childhood trauma, and have chronic symptoms of PTSD at the time of clinical presentation.⁸ Postapproval prescribing will certainly include such patients. Broader inclusion criteria increase the heterogeneity of a clinical trial sample, which demands larger samples, thereby exposing more participants to unknown drug risks or inferior treatments. The regulatory pathway for establishing a new treatment indication requires demonstrable efficacy for the target condition. The alternative, conducting trials that permit greater comorbidity and chronicity, could result in inconclusive findings that serve neither regulatory requirements nor clinical advancement. We would argue that establishing efficacy in a more defined population provides a foundation upon which real-world effectiveness can be evaluated.

Zimmerman and Snyder's¹ suggestion that product labeling should identify key exclusion criteria that limit generalizability such as the presence of comorbid MDE or childhood onset of PTSD has merit and would provide transparency for prescribers. However, we must recognize that virtually all psychiatric medications are prescribed outside the exact parameters of their registration trials. The question is not whether some extrapolation from trial data to clinical practice occurs—it always does—but rather whether the foundational evidence of efficacy and safety is sufficiently robust to justify

clinical use while acknowledging limitations.

One of the more valuable aspects of Zimmerman and Snyder's¹ analysis is their acknowledgment that PTSD pharmacotherapy trials vary substantially in their inclusion/exclusion criteria. He correctly notes that several placebo-controlled studies, such as with prazosin, have not excluded patients with comorbid depression or a more chronic PTSD duration. This variability in trial methodology reflects different research questions, development stages, and regulatory strategies across the PTSD treatment landscape. Zimmerman and Snyder's¹ call for greater attention to eligibility criteria variability across PTSD trials is well taken. The field would benefit from more systematic discussion of how methodological choices shape the evidence base and which patient populations remain understudied. However, the solution is not necessarily to design all trials identically, but rather to recognize that different trial designs serve different purposes in the treatment development pipeline.

Rather than requiring registration trials to achieve both internal validity and broad generalizability simultaneously, we need a comprehensive evidence development pathway. This includes (1) well-controlled efficacy trials establishing initial safety and efficacy in a targeted cohort; (2) replication studies in more diverse populations; (3) postmarketing pragmatic trials including previously excluded populations; and (4) systematic outcome measurement in routine care.

Another long overdue solution is to embed a simple clinical outcome measure, such as the 7-item Clinical Global Impression of Severity score, into medical record systems so that data mining of health outcomes can serve as a rich source for research into effectiveness of drug treatments.^{9,10} Systems that capture real-world outcomes for all patients receiving approved treatments would be another

possible avenue. Implementing universal mental health outcome measures in health records would reveal how interventions perform across the heterogeneous populations excluded from initial registration trials. This approach acknowledges that our highest-level evidence comes from our least generalizable populations, while creating mechanisms to rapidly expand our knowledge base once treatments reach clinical practice.

Zimmerman's previous work¹¹ on antidepressant trial generalizability has appropriately influenced how the field thinks about the research-practice gap. The present analysis extends this work to PTSD. However, several limitations merit emphasis. First, as a single-site study with a specific trauma profile (particularly high rates of childhood sexual abuse), it cannot definitively characterize "typical" PTSD patients. Zimmerman's¹¹ single-site practice, enriched for childhood sexual trauma survivors, may not represent PTSD patients nationally or internationally. Most importantly, the analysis does not address the counterfactual question: What would the results have been if the trials had been conducted with broader inclusion criteria? Would they still have detected a statistically significant and clinically meaningful treatment effect? Or would greater heterogeneity have obscured the signal, leaving us with null findings and no new treatment option for any PTSD patients? This is not merely a rhetorical question—it represents the fundamental tension in psychiatric drug development.

The brexpiprazole-sertraline trials were designed to answer a specific question: Does the combination of sertraline with brexpiprazole reduce PTSD symptoms more effectively than sertraline alone in patients with noncomorbid MDE and adult trauma within the past 10 to 15 years? Two^{2,3} out of three¹² pivotal trials concluded "yes." The totality of the brexpiprazole-sertraline data demonstrates significant efficacy and safety of the combination under conditions that allowed signal

detection, providing a foundation for clinical use and further research. This is a meaningful contribution to the limited PTSD pharmacotherapy armamentarium, particularly given that no new medications have been approved for PTSD in over 20 years.

Does this mean brexpiprazole-sertraline will benefit all PTSD patients? No. Does it mean patients with comorbid depression or childhood trauma should never receive this combination? No. Clinical judgment of the providers, individual patient characteristics and treatment history, and postmarketing experience will guide such decisions. The brexpiprazole-sertraline trials represent an important first step in understanding how this combination therapy may benefit patients with PTSD.

Zimmerman and Snyder's analysis¹ highlights the gap between research samples and clinical populations. They conclude by urging regulatory agencies to require studies that better reflect the patient populations seen in clinical practice. I would reframe this by stating that the field needs rigorous efficacy trials that establish clear treatment signals in a well-defined sample of individuals affected by a specific disease as well as studies that examine how treatments perform in more diverse populations. The former establish that treatments can work under optimal conditions for a targeted condition; the latter establish how they work in routine clinical settings for more diverse populations. Both are essential, and neither alone is sufficient. Rather than requiring registration trials to accomplish both simultaneously, let us advocate for a more comprehensive development pathway that includes systematic collection and analysis of real-world measurement-based data in the postmarketing clinical environment—an urgent call for a transdiagnostic universal outcome variable embedded in electronic medical records—a vital sign for mental health conditions.

Industry sponsors, nonprofits, and federal agencies must make

parsimonious investments in research and development. The goal is not arbitrary exclusion but rather maximizing the probability of detecting efficacy signals while minimizing participant exposure to potential harm during early-phase testing. Conversely, making eligibility criteria too strict threatens the feasibility of enrollment at a clinical research site, which can destroy a targeted timeline for recruitment and negatively impact the bottom-line budget and deliverables. This brings us back to our Goldilocks metaphor and the quest to find that “just right” sweet spot.

Article Information

Published Online: November 26, 2025.
<https://doi.org/10.4088/JCP.25com16224>

© 2025 Physicians Postgraduate Press, Inc.

J Clin Psychiatry 2026;87(1):25com16224

Submitted: November 11, 2025; accepted November 12, 2025.

To Cite: Davis LL. The Goldilocks problem: balancing internal validity with generalizability in clinical trials. *J Clin Psychiatry*. 2026;87(1):25com16224.

Author Affiliations: Birmingham VA Health Care System, Birmingham, Alabama, and University of Alabama at Birmingham Heersink School of Medicine, Birmingham, Alabama.

Corresponding Author: Lori L. Davis, MD, Birmingham VA Health Care System, 7901 Crestwood Blvd, Birmingham, AL 35210 (lori.davis@va.gov).

Relevant Financial Relationships: In the past 24 months, Dr. Davis has received consulting fees from Otsuka and research funding from Alkermes, Relmada, and Fisher Wallace.

Funding/Support: None.

Disclaimer: This commentary is based on the author's views and does not convey the opinions of the U.S. government or Otsuka Pharmaceuticals.

Acknowledgments: Dr. Davis acknowledges the generous input of A. John Rush, MD.

References

1. Zimmerman M, Snyder M. Lack of generalizability of PTSD treatment trials: the recent brexpiprazole-sertraline trials as an example. *J Clin Psychiatry*. 2026;87(1):25m15921.
2. Davis LL, Behl S, Lee D, et al. Brexpiprazole and sertraline combination treatment in posttraumatic stress disorder: a phase 3 randomized clinical trial. *JAMA Psychiatry*. 2025;82(3):218–227.
3. Hobart M, Chang D, Hefting N, et al. Brexpiprazole in combination with sertraline and as monotherapy in posttraumatic stress disorder: a full-factorial randomized clinical trial. *J Clin Psychiatry*. 2025;86(1):24m15577.
4. Koek RJ, Schwartz HN, Scully S, et al. Treatment-refractory posttraumatic stress disorder (TRPTSD): a review and framework for the future. *Prog Neuropsychopharmacol Biol Psychiatry*. 2016;70:170–218.
5. Fonzo GA, Federchenko V, Lara A. Predicting and managing treatment non-response in posttraumatic stress disorder. *Curr Treat Options Psychiatry*. 2020;7(2):70–87.
6. Parmenter ME, Lederman S, Weathers FW, et al. A phase 3, randomized, placebo-controlled, trial to evaluate the efficacy and safety of bedtime sublingual cyclobenzaprine (TNX-102 SL) in military-related posttraumatic stress disorder. *Psychiatry Res*. 2024;334:115764.
7. Davis LL, Urganus A, Gagnon-Sanschagrin P, et al. Patient journey of civilian adults diagnosed with posttraumatic stress disorder-A chart review study. *Curr Med Res Opin*. 2024;40(3):505–516.
8. Davis LL, Urganus A, Gagnon-Sanschagrin P, et al. Patient journey before and after a formal post-traumatic stress disorder diagnosis in adults in the United States - a retrospective claims study. *Curr Med Res Opin*. 2023;39(11):1523–1532.
9. Taquet M, Fazel S, Rush AJ. Transdiagnostic early warning score for psychiatric hospitalisation: development and evaluation of a prediction model. *BMJ Ment Health*. 2025;28(1):e301622.
10. Taquet M, Griffiths K, Palmer EOC, et al. Early trajectory of clinical global impression as a transdiagnostic predictor of psychiatric hospitalisation: a retrospective cohort study. *Lancet Psychiatry*. 2023;10(5):334–341.
11. Zimmerman M, Clark HL, Multach MD, et al. Have treatment studies of depression become Even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. *Mayo Clin Proc*. 2015;90(9):1180–1186.
12. Davis LL, Behl S, Lee D, et al. Fixed-dose brexpiprazole and sertraline combination therapy for the treatment of posttraumatic stress disorder: a phase 3, randomized trial. *J Clin Psychopharmacol*. 2025;45(6):580–589.

Scan Now



Cite and Share
this article at
Psychiatrist.com