BJPsych Open (2026)
12, e183, 1–3. doi: 10.1192/bjo.2026.12048

Commentary

Beyond discrimination: calibration and decision utility are required before GREAT can guide individual ECT decisions



Juan Carlos Munguia, Yolly Molina, Patricia Soriano and Isaac Zablah

Summary

Belz and colleagues present GREAT, a seven-item clinical instrument for predicting electroconvulsive therapy response in unipolar depression, with promising discriminatory validity (AUC 0.841). We identify three methodological gaps – absent calibration, limited sample representativeness and unquantified incremental value – that must be addressed before GREAT can guide individual clinical decisions.

Keywords

Electroconvulsive therapy; depression; prediction model; calibration; decision curve analysis.

Copyright and usage

© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Juan Carlos Munguía, MD (pictured), is a physician and surgeon, specialist in psychiatry, forensic medical examiner and specialist in scientific crime investigation, and works at the Faculty of Medical Sciences, National Autonomous University of Honduras, Tegucigalpa, Honduras.

Yolly Molina, MSc, is a researcher specialising in genetics, forensic science, neuroscience, medical informatics and genetically based neurological/metabolic diseases, and works at the Faculty of Medical Sciences, National Autonomous University of Honduras, Tegucigalpa, Honduras.

Patricia Soriano, MSc, is a researcher specialising in molecular biology applied forensic science and population genetics, and works at the Directorate of Forensic Medicine, Public Ministry, Tegucigalpa, Honduras.

Isaac Zablah, PhD, is a researcher on artificial intelligence, specialising in computing, high performance, medical informatics and technological innovation, and works at the Faculty of Medical Sciences, National Autonomous University of Honduras, Tegucigalpa, Honduras.

Electroconvulsive therapy (ECT) remains the most effective intervention for severe or treatment-resistant unipolar depression, yet the persistent challenge of anticipating individual response continues to impede timely and equitable treatment allocation. We read with considerable interest the prospective study by Belz and colleagues, who introduce the Göttingen Response to the ECT Assessment Tool (GREAT), a seven-item scoring instrument designed to predict response prior to treatment initiation.[1] The initiative is scientifically timely: access to ECT remains constrained by resource limitations and institutional barriers in many healthcare systems,[2] and a validated, infrastructure-independent predictor could meaningfully improve triage efficiency and shared decision-making. GREAT's pragmatic design, drawing exclusively on clinician-assessed routine data, requiring no neuroimaging or molecular assays and yielding an area under the receiver-operating characteristic curve (AUC) of 0.841, represents a genuine conceptual advance in a field where prior multivariate models have reported comparatively modest performance.[3,4] We offer three methodological considerations that, if addressed, would substantially strengthen the case for clinical implementation.

Calibration: the missing half of predictive validity

The most consequential gap in the current analysis concerns calibration, the degree of agreement between predicted and observed response probabilities across the full risk spectrum. The authors present a compelling case for discrimination (the capacity to rank patients by relative response likelihood), but the manuscript is entirely silent on whether the numerical probabilities implied by the GREAT score are accurate at the individual level. These are conceptually distinct properties: a model can achieve a high AUC and still systematically overestimate or underestimate response rates within clinical subgroups.[5,6] Contemporary prediction-model reporting standards, including TRIPOD + AI guidance, recommend calibration-in-the-large, calibration slope and graphical calibration plots as minimum requirements for any model intended to inform patient-level decisions.[5] The relevance of this omission is not merely methodological. GREAT's seven item gradations were established by clinical judgement and are acknowledged by the authors as 'partly arbitrary', introducing the possibility of systematic miscalibration at the extremes of the score distribution. Furthermore, 7 patients who exceeded the proposed ≥7-point threshold nonetheless failed to respond, while 2 patients below it achieved clinically meaningful Montgomery-Asberg Depression Rating Scale improvement.[1] These discordant cases are precisely where calibration determines whether the model's probabilistic framing supports or misleads bedside reasoning.

Sample composition and external validity

The preliminary nature of this derivation study warrants explicit attention to sample composition. The cohort is markedly sex-imbalanced, with women representing 80% of included patients ($n = 36/45$) – substantially higher than the female predominance

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 01:32:18, subject to the Cambridge Core terms of use.

reported in large multicentre ECT registries.[7,8] The relative predictive weight of age, psychomotor symptoms and pharmacotherapy resistance may differ in male patients, and this imbalance limits generalisability across broader clinical populations. In addition, item-level analyses revealed that four of the seven GREAT items (pharmacotherapy resistance, depression severity, psychotic symptoms and psychomotor symptoms) did not individually achieve statistical significance as predictors of response. In small samples, such non-significant weights can reflect insufficient statistical power rather than true clinical irrelevance, and their inclusion without recalibration may introduce noise into individual predictions.[9] Two further predictors, childhood trauma and maltreatment, have recently been associated with attenuated ECT outcomes in independent cohorts,[10,11] as have broader personality disorder traits beyond the borderline subtype.[12] Incorporating these dimensions into a revised GREAT version and adopting validated staging instruments such as the Maudsley Staging Method for treatment-resistance quantification,[1] would broaden the instrument's prognostic scope and clinical relevance.

Incremental value and the benchmark problem

A central question for any new clinical prediction tool is whether it offers net benefit beyond what clinicians already achieve through structured judgement or parsimonious models. Larger ECT prediction studies employing multivariable regression and Bayesian network approaches on independent, multicentre data-sets have reported substantially more modest performance adjusted $R^2 \approx$ 19% and AUC values of 0.686–0.693, even when drawing on broader predictor sets.[3,4] Machine learning approaches applied to ECT prediction have likewise underperformed on independent samples relative to derivation performance,[13,14] underscoring that high within-sample accuracy is a necessary but insufficient criterion for clinical recommendation. GREAT's superior metrics in a single-centre sample of 45 patients may partly reflect optimism inherent to derivation cohorts without external validation, a well-characterised source of inflated performance in clinical prediction research.[9] To establish GREAT's added value rigorously, we encourage the authors to compare its performance against a parsimonious baseline model restricted to the two statistically significant item-level predictors identified in their own analysis (age and episode duration), and against structured clinical judgement alone. Decision-curve analysis would then quantify net benefit across clinically plausible threshold probabilities relative to default strategies of treating all eligible or no patients.[15]

In conclusion, these considerations are not intended to diminish the importance of what Belz and colleagues have accomplished. The prospective design, pragmatic item selection and time-efficient administration of GREAT constitute a meaningful step toward individualised ECT planning that respects the resource constraints of real-world clinical environments. We encourage the authors to supplement the current analysis with calibration statistics, to prioritise sex- and age-stratified subgroup reporting, and to embed decision-curve analysis within the planned multicentre validation protocol.[16] The difficult-to-treat depression framework, which emphasises functional recovery and quality of life beyond binary response thresholds,[17] may also offer a clinically richer outcome target for the next GREAT iteration. Until calibration and incremental validity are demonstrated prospectively on independent samples, GREAT is best characterised as a promising screening adjunct rather than a standalone decision-support instrument, one whose full potential remains to be realised.

Juan Carlos Munguia, MD, Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras; **Yolly Molina**, MSc, Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras; **Patricia Soriano**, MSc, Directorate of Forensic Medicine, Public Ministry, Tegucigalpa, Honduras; **Isaac Zablah**, PhD, Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras

Correspondence: Isaac Zablah. Email: jose.zablah@unah.edu.hn

First received 16 Mar 2026 UTC, final revision 29 Apr 2026 UTC, accepted 4 Jun 2026 UTC

Data availability

No new data were generated or analysed for this correspondence.

Author contributions

Conceptualisation: J.C.M. and I.Z.; investigation: J.C.M. and Y.M.; validation: J.C.M. and P.S.; writing – original draft preparation: J.C.M. and I.Z.; writing – review and editing: Y.M.; supervision: P.S. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Declaration of interest

None.

Declaration of generative AI use

During the preparation of this work the author(s) used the QUILLBOT service to improve language and semantics concerns. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1 Belz M, Methfessel I, Besse M, Heinisch M, Strube W, Tritsch J, et al. Little information, great impact? A clinical tool for the prediction of electroconvulsive therapy effectiveness in depression. *BJPsych Open* 2026; **12**: e66.

2 Wilkinson ST, Kitay BM, Harper A, Rhee TG, Sint K, Ghosh A, et al. Barriers to the implementation of electroconvulsive therapy (ECT): results from a nationwide survey of ECT practitioners. *Psychiatr Serv* 2021; **72**: 752–7.

3 Loef D, Hoogendoorn AW, Somers M, Mocking RJT, Scheepens DS, Scheepstra KWF, et al. A prediction model for electroconvulsive therapy effectiveness in patients with major depressive disorder from the Dutch ECT Consortium (DEC). *Mol Psychiatry* 2025; **30**: 1915–24.

4 van der Does Y, Turner RJ, Bartels MJH, Hagoort K, Metselaar A, Scheepers F, et al. Outcome prediction of electroconvulsive therapy for depression. *Psychiatry Res* 2023; **326**: 115328.

5 Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 2024; **384**: e074819.

6 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; **17**: 230.

7 van Diermen L, van den Ameele S, Kamperman AM, Sabbe BCG, Vermeulen T, Schrijvers D, et al. Prediction of electroconvulsive therapy response and remission in major depression: meta-analysis. *Br J Psychiatry* 2018; **212**: 71–80.

8 Haq AU, Sitzmann AF, Goldman ML, Maixner DF, Mickey BJ. Response of depression to electroconvulsive therapy: a meta-analysis of clinical predictors. *J Clin Psychiatry* 2015; **76**: 1374–84.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 01:32:18, subject to the Cambridge Core terms of use.

9 Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; **368**: m441.

10 Jelovac A, Mohan C, Whooley E, Igoe A, McCaffrey C, McLoughlin DM. Childhood maltreatment and outcomes following electroconvulsive therapy in adults with depression. *Acta Psychiatr Scand* 2025; **151**: 46–55.

11 Thompson S, Finnegan M, Galligan T, Jelovac A, McLoughlin DM. Childhood trauma and response to electroconvulsive therapy for depression: a pilot study. *Brain Stimul* 2024; **17**: 864–6.

12 Ferrea S, Petrides G, Ehrt-Schäfer Y, Angst J, Seifritz E, Olbrich S, et al. Outcomes of electroconvulsive therapy in patients with depressive symptoms with versus without comorbid personality disorders/traits: a systematic review and meta-analysis. *Acta Psychiatr Scand* 2024; **149**: 18–32.

13 Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* 2016; **73**: 557–64.

14 Nakajima K, Takamiya A, Uchida T, Kudo S, Nishida H, Minami F, et al. Individual prediction of remission based on clinical features following electroconvulsive therapy: a machine learning approach. *J Clin Psychiatry* 2022; **83**: 21m14293.

15 Zhao L, Leng Y, Hu Y, Xiao J, Li Q, Liu C, et al. Understanding decision curve analysis in clinical prediction model research. *Postgrad Med J* 2024; **100**: 512–5.

16 Blanken TF, Kok R, Obbels J, Lambrichts S, Sienaert P, Verwijk E. Prediction of electroconvulsive therapy outcome: a network analysis approach. *Acta Psychiatr Scand* 2025; **151**: 521–8.

17 McIntyre RS, Alsuwaidan M, Baune BT, Berk M, Demyttenaere K, Goldberg JF, et al. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. *World Psychiatry* 2023; **22**: 394–412.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 01:32:18, subject to the Cambridge Core terms of use.