Acta Neuropsychiatrica

cambridge.org/neu

Precision psychiatry needs causal inference

Martin Bernstorff[1,2,3] and Oskar Hougaard Jefsen[4,5]

[1]Department of Affective Disorders, Aarhus University Hospital – Psychiatry, Aarhus, Denmark; [2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [3]Center for Humanities Computing, Aarhus University, Aarhus, Denmark; [4]Psychosis Research Unit, Aarhus University Hospital – Psychiatry, Aarhus, Denmark and [5]Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

Perspective

Cite this article: Bernstorff M and Jefsen OH. (2025) Precision psychiatry needs causal inference. *Acta Neuropsychiatrica*. **37**(e32), 1–5. doi: 10.1017/neu.2024.29

Received: 14 February 2024
Revised: 23 May 2024
Accepted: 9 June 2024

Keywords:
Machine learning; causality; psychiatry; precision medicine

Corresponding author:
Oskar Hougaard Jefsen;
Email: oskar.jefsen@clin.au.dk

Abstract

Objective: Psychiatric research applies statistical methods that can be divided in two frameworks: causal inference and prediction. Recent proposals suggest a down-prioritisation of causal inference and argue that prediction paves the road to 'precision psychiatry' (i.e., individualised treatment). In this perspective, we critically appraise these proposals. *Methods:* We outline strengths and weaknesses of causal inference and prediction frameworks and describe the link between clinical decision-making and counterfactual predictions (i.e., causality). We describe three key causal structures that, if not handled correctly, may cause erroneous interpretations, and three pitfalls in prediction research. *Results:* Prediction and causal inference are both needed in psychiatric research and their relative importance is context-dependent. When individualised treatment decisions are needed, causal inference is necessary. *Conclusion:* This perspective defends the importance of causal inference for precision psychiatry.

Summations

- Psychiatric research applies statistical methods from two different frameworks: causal inference and prediction.
- If prediction methods (i.e., machine learning algorithms) are causally agnostic, their ability to inform clinical decision-making is limited.
- Common pitfalls can be avoided by considering key causal structures such as confounders, mediators, and colliders.

Perspectives

- The relative need for prediction vs. causal inference is context-dependent.
- New methods combining prediction and causal inference may hold great promise for precision psychiatry.

The recent move towards prediction in psychiatry

Psychiatric care involves deciding upon the optimal course of action for individual patients (i.e., clinical decision-making). Psychiatric research aids clinical care by developing diagnostic methods, new treatments, evaluating safety and efficacy, and much more. This research depends on the application of appropriate statistical frameworks to answer specific research questions. These frameworks may generally be divided into frameworks for causal inference and for prediction, as previously described (Breiman, 2001; Bzdok *et al.*, 2018). Causal inference aims to determine the effect of one variable on another, which is crucial for selecting between alternative courses of action. In contrast, prediction aims at forecasting, independently of whether the patterns observed cause the predicted data, or simply correlate with them. Colloquially, it is one thing to build a barometer to predict a storm (prediction), another to know how to improve the weather (causal inference). Importantly, answering causal questions does not require a granular understanding of mechanisms. For example, randomised trials may provide evidence for the effects of a given medication on an outcome, irrespective of whether the mechanism of that medication is known. Likewise, knowing whether smoking causes cancer does not require extensive knowledge about all mediating mechanisms. Traditionally, psychiatric research has focused mostly on understanding, rather than predicting, and on population-level, rather than individual-level, questions, but this focus may be shifting.

Advances in machine learning (ML) and the increasing availability of large datasets have led to wide propositions about the potential applications of ML-based prediction at the level of individual patients in healthcare (Matheny *et al.*, 2020). ML methods are able to incorporate

© The Author(s), 2024. Published by Cambridge University Press on behalf of Scandinavian College of Neuropsychopharmacology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

large amounts of data, detect complex dependencies, and often focus explicitly on optimising generalisability. These strengths are suggested to furnish reliable predictions for individual patients, and thus more individually tailored treatments – that is, 'precision medicine' (Bzdok *et al.*, 2021). In psychiatry, ML prediction methods have been argued to be superior to 'traditional' methods (aimed at causal inference) when it comes to individualising psychiatric care, and several authors thus propose a wider adaptation of ML methods in psychiatric research (Paulus, 2015; Bzdok *et al.*, 2021). For example, Bzdok et al. argue that '*Prediction, not association, paves the road to precision medicine*' and Paulus proposes '*. . . that we shift from a search for elusive mechanisms to implementing studies that focus on predictions to help patients now*'. These viewpoints not only suggest a wider application of the prediction framework, but also a down-prioritisation of causal inference. The growing interest in prediction is also reflected in the rising number of studies applying ML methods in psychiatry (Chekroud *et al.*, 2021; Salazar de Pablo *et al.*, 2021; Koutsouleris *et al.*, 2022). The promise of prediction is great: If we combine 'big data' with ML methods, we may be able to make individual-level predictions that can guide clinical, psychiatric care. However, prediction may be insufficient. As argued recently (Prosperi *et al.*, 2020; Wilkinson *et al.*, 2020), precision medicine – that is, choosing the optimal course of action for individual patients – cannot be built on prediction alone, but requires causal inference. Here, we elaborate why, and how it applies to psychiatry.

Prediction is not enough

In this Viewpoint, we claim that the direction set forth by Bzdok et al., Paulus, and others, may not deliver precision psychiatry as intended. We argue that prediction models may very well yield accurate prognostic predictions, but that this is insufficient for improving clinical decision-making. This claim is based on the central difference between determining prognosis – a *factual* prediction – and deciding between alternative treatment options – a causal (*counterfactual*) question. Here, we elaborate the claim that precision psychiatry needs causal inference first by describing key causal structures, second, by showing how prediction models may be misinterpreted if causality is neglected, and third, by describing how causal inference and prediction may complement each other in psychiatric research, going forward.

Key causal structures

The field of causal inference highlights three fundamental causal structures which, if not handled appropriately during analysis, will result in incorrect conclusions about the intervention of interest (Hernán & Robins, 2016; Pearl *et al.*, 2016). These are 1) confounders, 2) mediators, and 3) colliders. We will discuss each in turn, placing a particular focus on their role in psychiatric research.

Confounders

Confounders are variables that have a causal effect on both the exposure (e.g., an intervention) and the outcome (Figure 1A). If these variables are not conditioned on, estimates of the effect of the intervention on the outcome will be biased. Note that 'conditioned on' can mean: a) Participant selection is dependent on the variable or b) the variable is included in the statistical model. A typical example in healthcare is confounding by indication; where the intervention is administered based on some criterion (typically a disease). For example, we may study patients with depression, some of which have been treated with electroconvulsive therapy (ECT). As a predictor of post-treatment depressive symptoms, ECT would likely perform tremendously well (Kellner *et al.*, 2020). However, we know that ECT is indicated for patients with severe depression. Initial symptom severity both causes the intervention and causes post-treatment symptom severity (Figure 1B). Pre-treatment symptom severity thus confounds the association between ECT and post-treatment symptoms. The association would, if interpreted incorrectly, lead us to the erroneous conclusion that ECT harms patients, even though randomised trials show consistent benefit. To remove bias from confounding, we can add the confounding variable to our model. If we have measured the variable with sufficient granularity and precision, this will solve the problem. So, is causal inference just a problem of measuring a sufficient number of variables? Can we solve our problems with big(ger) data? Unfortunately, not necessarily.

Mediators

When a variable is caused by the exposure of interest (i.e., an intervention) and has a causal effect on the outcome, it is a mediator (Figure 1C). If this variable is conditioned on, estimates of the intervention's effect on the outcome will be biased. Mediators, as well as colliders, are examples where adding more variables to your model will not only decrease statistical precision, it will also lead you to draw wrong conclusions. For example, a theory of cognitive behavioural therapy (CBT) assumes that it causes changes to a patient's cognitive schemas, which causes changes to rumination, which alleviates depressive symptoms (Watkins, 2009). In this case, CBT will be a strong predictor of depressive symptoms if rumination is not included in the analyses, but may or may not be a strong predictor if rumination *is* included (Figure 1D). The association depends on the particularities of the prediction model chosen, not the actual effect of treatment. For causal estimates, mediators should not be included in the models.

Colliders

When a variable is caused by the exposure (i.e., intervention) and the outcome (or another variable with a causal effect on the outcome), it is a collider (Figure 1E). If this variable is conditioned on, estimates of an exposure's effect on the outcome will be biased. An example of this is the paradoxical observation that post-traumatic stress disorder (PTSD) is negatively associated with suicide in some studies (Zivin *et al.*, 2007), despite strong evidence that PTSD increases the risk of suicide in other studies (Gradus *et al.*, 2010; Fox *et al.*, 2021) As described by (H. Jiang *et al.*, 2022), this paradox may arise due to conditioning on mediators (e.g., depression) that share a common cause (e.g., other mental illness, lack of social support, etc.) with the outcome (Figure 1F). In this case, it acts as a collider. For causal estimates, colliders should not be conditioned on.

With these concepts in hand, we can turn to how modern ML-based prediction models can be misinterpreted if causality is neglected.

Prediction pitfalls

Pitfall #1 - mistaking feature importance for causal importance

When ML methods are used to predict an outcome, researchers may wish to know which variables were most important for the prediction, that is, evaluate 'feature importance'. Here, an

https://doi.org/10.1017/neu.2024.29 Published online by Cambridge University Press

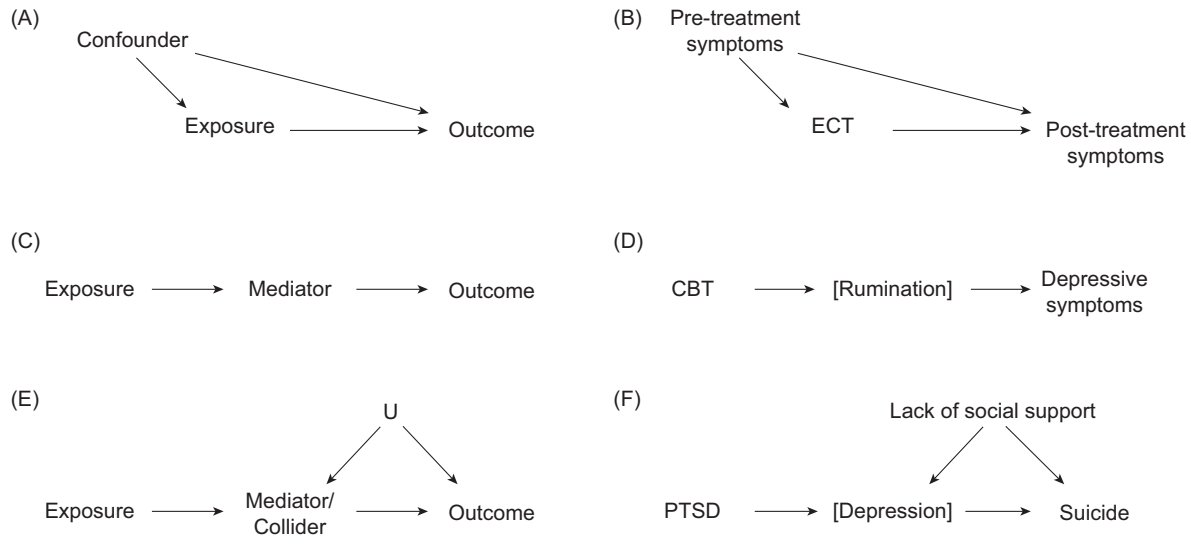


Figure 1. Directed acyclic graphs (DAGs) describing key causal structures. CBT = cognitive behavioural therapy; ECT = electroconvulsive therapy; PTSD = post-traumatic stress disorder; U = unknown variable.

important pitfall is to mistake feature importance for causal importance, forget the key causal structures described above, and assume that modification of the variable will necessarily alter the risk of the outcome. An example of such misinterpretation is seen in Liu et al., who identified higher body mass index (BMI) as an important variable for predicting cognitive impairment. They concluded: *'Therefore, interventions for cognitive function among the elderly should target weight management'*. This statement assumes a causal effect of BMI on cognitive function, without considering alternative explanations. High feature importance could just as well arise if BMI and lowered cognitive function share a common cause (i.e., confounding), for example historical lack of exercise (Pitrou *et al.*, 2022). While weight management is likely beneficial to most patients, there may be cases where mistaking feature importance for causal importance will be less beneficial, or even harmful to patients (Guglin *et al.*, 2023).

Pitfall #2 - mistaking symptom changes for treatment effects

ML-based methods are suggested to support selection of treatment. This calls for a model that can determine which treatment will improve the patient's state the most. Research in prediction of 'treatment response' attempts to answer this question by training models to predict which patients will improve after being administered treatment. If the patient improves, the logic goes, they were given the right treatment. However, improving after being administered treatment can be due to a plethora of factors besides the treatment itself, for example, confounding, regression to the mean, placebo effects, natural course of illness, etc. An example of this pitfall is seen in (Redlich *et al.*, 2016), which recruited patients with major depressive disorder (MDD) and obtained baseline structural magnetic resonance imaging (sMRI) data before treatment with ECT + antidepressants. Crucially, the ML algorithms were trained solely on the ECT group. Redlich et al., found ML on baseline sMRI could predict treatment response and concluded: *'Although determining which ECT recipients will respond remains difficult in clinical practice, a routine assessment with structural MRI before treatment could serve as a decision guide for clinical psychiatrists'*. While these predictions may hint at outcomes after treatment with ECT, they do not estimate how patients would have fared if they were not given ECT. For example, the MRI may simply identify patients that would have recovered on their own, irrespective of whether they were treated or not. If the study had been designed to for causal inference, it could have served as a decision guide for treatment.

Pitfall #3 - avoiding causal inference altogether

In order to avoid the pitfalls described above, researchers applying ML methods for prediction purposes may rightfully refrain from making causal interpretations of their predictive models, and many studies indeed do so. For example, Jiang et al., applied ML methods to register-based data to predict suicide in the 30 days post discharge from a psychiatric hospital and stated: *'It is noteworthy that these predictors should be interpreted as risk markers and not causal risk factors, given that our analyses were not intended to quantify the causal effect of any of these predictors, but rather to examine their contribution to accurate prediction of postdischarge suicide' (T. Jiang et al., 2021)*. With this statement, they acknowledge that they leave a crucial question unanswered: How can we better prevent suicide? Brief suicide prevention interventions reduce the number of suicide attempts by roughly 30% (Doupnik *et al.*, 2020), but what of the remaining 70%? Prediction models can identify which patients are missed and should receive interventions, and when we know which interventions to administer, this is valuable. However, as prediction accuracy improves and fewer patients are missed, intervention efficacy becomes the limiting factor for clinical care, and research may be centred around methods to identify causal mechanisms and develop more effective interventions.

Ways forward

When is prediction enough?

In the example above (i.e., suicide prevention), the challenges lie both in knowing who to act upon and knowing how to act. However, there may be contexts where we know how to act (i.e., treat/prevent), but systematically miss patients who we should act upon. This may be the case for type 2 diabetes (T2D).

https://doi.org/10.1017/neu.2024.29 Published online by Cambridge University Press

The causal mechanisms underlying T2D development and the effectiveness of different interventions is well established through RCTs (i.e., causal inference) (Knowler, 2002), but in some populations, for example psychiatric patients, they are systematically undertreated (Scott & Happell, 2011). Hence, ML-based prediction need not provide causal knowledge, but only identify at-risk individuals. In this case, the lack of causal inference in the prediction is compensated by the strong causal understanding of the mechanisms involved in T2D, and this may generalise many clinical issues.

Inferring causality

The development of psychiatric disorders is highly complex, and the underlying causal effects are typically not known, motivating causal inference. Although RCTs remain the gold standard for causal inference, they are often unfeasible or unethical (e.g., for determining the effects of childhood trauma or substance abuse on mental health). Methods to infer causal effects from observational data are thus needed. Algorithm-based identification of causal networks is a promising, ongoing research field, (Eberhardt, 2017), but literature in psychiatry is scarce. Instead, the dominant approach relies on experts to specify a set of assumptions which is agreed upon and then to acquire data to estimate causal effects (Hernán & Robins, 2016). Interactive tools have been developed to exactly this end (Textor *et al.*, 2016). Broader approaches to causal inference in health sciences have also been described, such as 'inference to the best explanation', 'triangulation', and the classical Hill criteria (Krieger & Davey Smith, 2016; Ohlsson & Kendler, 2020). Regardless of the exact approach, we are convinced that causal inference frameworks will play a defining role in developing the future of psychiatric care.

Acknowledgements. None.

Author contributions. MJ and OHJ wrote the paper.

Funding statement. OHJ is funded by the Health Research Foundation of the Central Denmark Region (Grant no. R64-A3090-B1898).

Competing interests. The authors declare no competing interests.

References

Breiman L (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Statistical Science* **16**, 199–231. doi: 10.1214/ss/1009213726.

Bzdok D, Altman N and Krzywinski M (2018) Statistics versus machine learning. *Nature Methods* **15**, 233–234. doi: 10.1038/nmeth.4642.

Bzdok D, Varoquaux G and Steyerberg EW (2021) Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry* **78**, 127–128. doi: 10.1001/jamapsychiatry.2020.2549.

Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen Z, Belgrave D, DeRubeis R, Iniesta R, Dwyer D and Choi K (2021) The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)* **20**, 154–170. doi: 10.1002/wps.20882.

Doupnik SK, Rudd B, Schmutte T, Worsley D, Bowden CF, McCarthy E, Eggan E, Bridge JA and Marcus SC (2020) Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings: a systematic review and meta-analysis. *JAMA Psychiatry* **77**, 1021–1030. doi: 10.1001/jamapsychiatry.2020.1586.

Eberhardt F (2017) Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics* **3**, 81–91. doi: 10.1007/s41060-016-0038-6.

Fox V, Dalman C, Dal H, Hollander A-C, Kirkbride JB and Pitman A (2021) Suicide risk in people with post-traumatic stress disorder: a cohort study of 3.1 million people in Sweden. *Journal of Affective Disorders* **279**, 609–616. doi 10.1016/j.jad.2020.10.009.

Gradus JL, Qin P, Lincoln AK, Miller M, Lawler E, Sørensen HT and Lash TL (2010) Posttraumatic stress disorder and completed suicide. *American Journal of Epidemiology* **171**, 721–727. doi: 10.1093/aje/kwp456.

Guglin M, Li B, Kanwar M, Abraham J, Kataria R, Bhimaraj A, Vallabhajosyula S and Kapur N (2023) Obesity and outcomes in cardiogenic shock due to acute myocardial infarction. *European Heart Journal* **44**, 655.1144. doi: 10.1093/eurheartj/ehad655.1144.

Hernán MA and Robins JM (2016) Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* **183**, 758–764. doi: 10.1093/aje/kwv254.

Jiang H, Huang N, Tian W, Shi S, Yang G and Pu H (2022) Factors associated with post-traumatic stress disorder among nurses during COVID-19. *Frontiers in Psychology* **13**, 282. doi: 10.3389/fpsyg.2022.745158.

Jiang T, Rosellini AJ, Horváth-Puhó E, Shiner B, Street AE, Lash TL, Sørensen HT and Gradus JL (2021) Predicting suicide in the 30 Days postdischarge from psychiatric hospitalization in Denmark using machine learning. *The British Journal of Psychiatry : The Journal of Mental Science* **219**, 440–447. doi: 10.1192/bjp.2021.19.

Kellner CH, Obbels J and Sienaert P (2020) When to consider electroconvulsive therapy (ECT). *Acta Psychiatrica Scandinavica* **141**, 304–315. doi: 10.1111/acps.13134.

Knowler (2002) Reduction in the incidence of Type 2 Diabetes with lifestyle intervention or metformin. *The New England Journal of Medicine* **11**, 319–22.

Koutsouleris N, Hauser TU, Skvortsova V and Choudhury MD (2022) From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health* **4**, e829–e840. doi: 10.1016/S2589-7500(22)00153-4.

Krieger N and Davey Smith G (2016) The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology* **45**, 1787–1808. doi: 10.1093/ije/dyw114.

Matheny ME, Whicher D and Thadaney Israni S (2020) Artificial intelligence in health care: a report from the National Academy of Medicine. *JAMA* **323**, 509–510. doi: 10.1001/jama.2019.21579.

Ohlsson H and Kendler KS (2020) Applying Causal inference methods in psychiatric epidemiology: a review. *JAMA Psychiatry* **77**, 637–644. doi: 10.1001/jamapsychiatry.2019.3758.

Paulus MP (2015) Pragmatism instead of mechanism: a call for impactful biological psychiatry. *JAMA Psychiatry* **72**, 631–632. doi: 10.1001/jamapsychiatry.2015.0497.

Pearl J, Glymour M and Jewell NP (2016) *Causal inference in statistics: a primer*. Chichester, West Sussex: John Wiley & Sons.

Pitrou I, Vasiliadis H-M and Hudon C (2022) Body mass index and cognitive decline among community-living older adults: the modifying effect of physical activity. *European Review of Aging and Physical Activity* **19**, 3. doi: 10.1186/s11556-022-00284-2.

Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, Rich S, Wang M, Buchan IE and Bian J (2020) Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence* **2**, 369–375. doi: 10.1038/s42256-020-0197-y.

Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, Münker S, Mühlmann L, Wahl P, Heindel W, Arolt V, Alferink J, Zwanzger P, Zavorotnyy M, Kugel H and Dannlowski U (2016) Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* **73**, 557–564. doi: https://doi.org/10.1001/jamapsychiatry.2016.0316.

Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, Irving J, Catalan A, Oliver D, Baldwin H, Danese A, Fazel S, Steyerberg EW, Stahl D and Fusar-Poli P (2021) Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. *Schizophrenia Bulletin* **47**, 284–297. doi: 10.1093/schbul/sbaa120.

Scott D and Happell B (2011) The high prevalence of poor physical health and unhealthy lifestyle behaviours in individuals with severe mental illness. *Issues in Mental Health Nursing* **32**, 589–597. doi: 10.3109/01612840.2011.569846.

Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M and Ellison GT (2016) Robust causal inference using directed acyclic graphs: the R package 'dagitty. *International Journal of Epidemiology* **45**, 1887–1894. doi: 10.1093/ije/dyw341.

Watkins ER (2009) Depressive rumination: investigating mechanisms to improve cognitive behavioural treatments. *Cognitive Behaviour Therapy* **1**, 8–14. doi: 10.1080/16506070902980695.

Wilkinson J, Arnold KF, Murray EJ, Smeden Mvan, Carr K, Sippy R, Kamps Mde, Beam A, Konigorski S, Lippert C, Gilthorpe MS and Tennant PWG (2020) Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health* **2**, e677–e680. doi: 10.1016/S2589-7500(20)30200-4.

Zivin K, Kim HM, McCarthy JF, Austin KL, Hoggatt KJ, Walters H and Valenstein M (2007) Suicide mortality among individuals receiving treatment for depression in the veterans affairs health system: associations with patient and treatment setting characteristics. *American Journal of Public Health* **97**, 2193–2198. doi: 10.2105/AJPH.2007.115477.

https://doi.org/10.1017/neu.2024.29 Published online by Cambridge University Press