Psychological Medicine

cambridge.org/psm

Mind the prevalence rate: overestimating the clinical utility of psychiatric diagnostic classifiers

Ahmad Abu-Akel[1], Chad Bousman[2,3], Efstratios Skafidas[2,4] and Christos Pantelis[2,4]

[1]Institute of Psychology, University of Lausanne, Lausanne, Switzerland; [2]Department of Psychiatry, Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Carlton South, Victoria, Australia; [3]Departments of Medical Genetics, Psychiatry, and Physiology & Pharmacology, University of Calgary, Calgary, AB, Canada and [4]Centre for Neural Engineering, University of Melbourne, Parkville, Victoria, Australia

Editorial

Cite this article: Abu-Akel A, Bousman C, Skafidas E, Pantelis C (2018). Mind the prevalence rate: overestimating the clinical utility of psychiatric diagnostic classifiers. *Psychological Medicine* **48**, 1225–1227. https://doi.org/10.1017/S0033291718000673

Received: 24 November 2017
Revised: 21 February 2018
Accepted: 22 February 2018
First published online: 20 March 2018

Key words:
Autism; clinical classifiers; negative predictive value; positive predictive value; prevalence rate; psychiatric conditions; psychosis; risk calculators

Author for correspondence:
Christos Pantelis, E-mail: cpant@unimelb.edu.au

Abstract

Currently, there is an intense pursuit of pathognomonic markers and diagnostic ('risk-based') classifiers of psychiatric conditions. Commonly, the epidemiological prevalence of the condition is not factored into the development of these classifiers. By not adjusting for prevalence, classifiers overestimate the potential of their clinical utility. As valid predictive values have critical implications in public health and allocation of resources, development of clinical classifiers should account for the prevalence of psychiatric conditions in both general and high-risk populations. We suggest that classifiers are most likely to be useful when targeting enriched populations.

The last few years have witnessed a surge in the development of predictive classifiers to ascertain the probability of an individual to develop a particular mental health condition at some point in the future. These efforts are potentially important as they could make significant contributions to developing preventive approaches (Cannon *et al.* 2016). While such classifiers have yet to be adopted generally in clinical practice, recent influential models, reporting high predictive values (from ~70–100%), raise such prospects, and have spurred startup companies to generate enthusiasm and attract investors (Hayden, 2017), as well as ambitious strategies for prevention (Couzin-Frankel, 2017).

Within psychiatric conditions, predictive classifiers have probably been mostly applied in autism and psychosis spectrum disorder populations (ASD and PSD, respectively). Researchers suggest that various indices, from simple demographics to biological, could be used in clinical practice to improve diagnosis of ASD (Pramparo *et al.* 2015; Yahata *et al.* 2016; Howsmon *et al.* 2017; Emerson *et al.* 2017; Hazlett *et al.* 2017) and PSD (Cannon *et al.* 2008; Bedi *et al.* 2015; Cannon *et al.* 2016; Fusar-Poli *et al.* 2017; Hafeman *et al.* 2017), within both general and high-risk populations. However, since these classifiers can have critical implications in public health, inspection of the validity of their predictive values is important because 'false diagnostic predictions have the potential to adversely affect individuals and families' (Hazlett *et al.* 2017).

Generally, the accuracy of predictive classifiers depends on: (1) the data collected before the onset of the condition (which could include but not be limited to demographic, clinical, genetic, and brain-based markers/indices), (2) the clinical classification instrument used to determine the presence or absence of the condition, and (3) the prevalence of diagnosed individuals in the test population. While classifiers have been criticized on both methodological and statistical grounds (Studerus *et al.* 2017), accounting for the epidemiological prevalence of the condition in question has largely been overlooked. By not adjusting for epidemiological prevalence, which is unfortunately a commonplace practice (Cannon *et al.* 2008; Sundermann *et al.* 2014; Bedi *et al.* 2015; Pramparo *et al.* 2015; Yahata *et al.* 2016; Emerson *et al.* 2017; Hazlett *et al.* 2017; Just *et al.* 2017), classifiers often seriously overestimate their clinical potential even for enriched, high-risk populations.

The clinical utility of the classifier is estimated in terms of two values: the positive predictive value (PPV; i.e. how likely it is that the individual *with* the condition is correctly identified) and the negative predictive value (NPV; i.e. how likely it is that the individual *without* the condition is correctly identified). Importantly, these values are sensitive to the prevalence of the condition in the population of interest. In cases where there is a mismatch between the prevalence of the condition in the test sample and its prevalence in the population (general or high-risk), the clinical value of the classifier needs to be estimated by calculating the Bayes' adjusted positive and

CAMBRIDGE UNIVERSITY PRESS

negative predicted values for the prevalence of the condition in the population as follows:

$$\mathrm{PPV}=\frac{\text{sensitivity}\times\text{prevalence}}{\text{sensitivity}\times\text{prevalence}+\left(1-\text{specificity}\right)\times\left(1-\text{prevalence}\right)}$$

and

$$\mathrm{NPV}=\frac{\text{specificity}\times\left(1-\text{prevalence}\right)}{\left(1-\text{sensitivity}\right)\times\text{prevalence}+\text{specificity}\times\left(1-\text{prevalence}\right)}$$

We emphasize that what we are presenting here is not a new analysis or method, but an overlooked necessary step for calculating a classifier's predictive value and thus estimating its clinical utility. In fact, the influence of the prevalence of disease on the predictive values of diagnostic/screening classifiers has long been recognized such that increasing prevalence increases PPV and decreases NPV (Mausner & Kramer, 1985; Altman & Bland, 1994), as also shown in our prior study of ASD (Skafidas *et al.* 2014).

We illustrate this point by an examination of two recent influential studies reporting on the promise of such diagnostic classifiers. The first study (Hazlett *et al.* 2017) reports on a diagnostic classifier that uses brain surface area of 6–12-month-old siblings of children with ASD, to predict whether these infants would develop the condition at age 24 months. It is reported that a deep-learning algorithm that primarily used this brain measure, correctly classified which of the infants developed the condition at a 94% level of accuracy, with 88% sensitivity, and 95% specificity. This corresponded to 81% PPV, and 97% NPV.

Relevant to our argument, the reported sensitivity and specificity values in this study were based on the analysis of 179 infants of high familial risk, of whom 34 infants developed ASD at 24 months of age. The crucial point to which we would like to draw the readers' attention is that the resultant predictive values are based on the *prevalence* of ASD in this test sample, which is at 19% (or 34/179). However, the epidemiological prevalence of ASD in children having siblings with ASD in this sample is likely to be overestimated (Szatmari *et al.* 2016), and estimates from a large population study suggest a much lower prevalence of 6.9% for full siblings, 2.4% for maternal half-siblings, and 1.50% for paternal half-siblings (Gronborg *et al.* 2013). Under such uncertainty, of over- or under-estimation of prevalence rates, the clinical utility of the classifier simply cannot be evaluated; substantiated epidemiological prevalence rates are a necessary first step to assessing the true predictive validity of any diagnostic classifier. Therefore, when adjusting for a prevalence of 6.9%, for example, their test with 88% sensitivity and 95% specificity yields 57% PPV and 99% NPV. These values translate to a high false discovery rate of 43% (i.e. the probability of misclassifying those with a condition as without), and low false omission rate of 1% (i.e. the probability of misclassifying those without a condition as having the condition), which substantially undermines the clinical utility of the classifier to detect risk for ASD among infants at high familial risk for ASD.

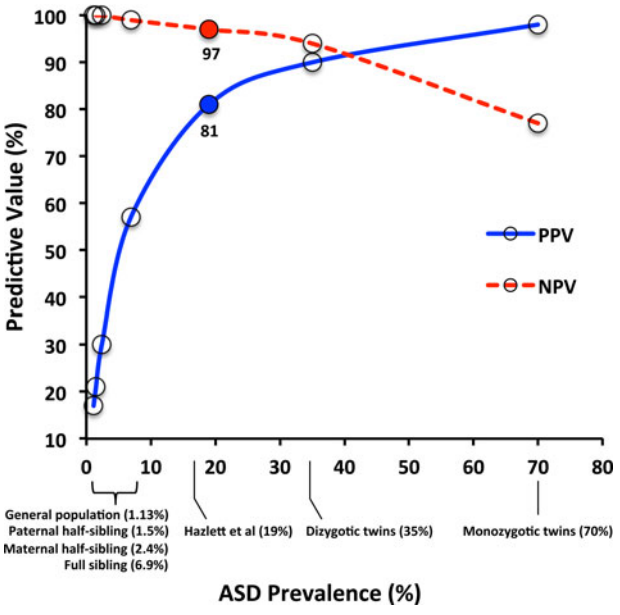

Fig. 1. Dependence of predictive values on condition prevalence. We show the dependency of predictive values on the prevalence rate of autism spectrum disorders (ASD) in seven populations, including the Hazlett *et al.* sample (filled circles), based on a classifier with 88% sensitivity and 95% specificity.

In a second study, Pramparo *et al.* (2015) reported that genomic biomarkers correctly classified 83% of boys with ASD in general pediatric settings in the discovery population (80% specificity and 85% sensitivity) and 75% of the replication sample (72% specificity and 77% sensitivity). These estimates were based on a 52% prevalence in the discovery sample, and 47% prevalence in the replication sample, both of which do not reflect the population prevalence of ASD in boys in the USA, currently estimated at 1 in 42 boys, or 2.38% (C.D.C.P, 2012). Using the specificity and sensitivity estimates from the replication sample (Specificity = 72.41%; Sensitivity = 77.27%) while adjusting for epidemiological prevalence, the PPV is only 6.39% and the NPV is 99.24%, reflecting extremely high false discovery rate of about 93%, and extremely low false omission rate of about 0.8%. These adjusted values undermine the promise of the classifier in 'detecting risk for ASD among infants in the general pediatric population' (Pramparo *et al.* 2015).

To illustrate more fully the dependency of predictive values on the prevalence rate within a study cohort compared with the estimates in the population, we repeated the same analysis from the Hazlett *et al.* (2017) study by adjusting for the prevalence of ASD in the general population (1.13%) (C.D.C.P, 2012), paternal (1.5%) and maternal (2.4%) half-siblings (Gronborg *et al.* 2013), as well as for the prevalence of ASD in dizygotic (35%) and monozygotic (70%) twins (Hallmayer *et al.* 2011) (see Fig. 1). This figure underscores the importance of understanding the sources of variation in the prevalence estimates of the condition of interest and their impact on the predictive values of screening assessments or biological markers.

Our discussion underscores that the utility of a diagnostic classifier for a particular condition depends both on the discriminatory value of the classifier and on the prevalence of the condition in the population of interest. Accurate predictive values are particularly important when screening assessments or biological markers are considered as early indicators that lead to early identification of diseases. Failure precisely to calculate predictive values may lead to misleading conclusions about the utility of the assessment tools and/or biological markers for the professional community as well as the general population. Therefore, clinical classifiers should be developed in tandem with rigorous epidemiological studies to ascertain the prevalence of psychiatric conditions in both general and high-risk populations. Above all,

https://doi.org/10.1017/S0033291718000673 Published online by Cambridge University Press

the importance of accurate prevalence estimates cannot be overlooked as it allows for the allocation of resources including use of screening assessments and biological markers for efficient interventions/preventions to lessen disease burden.

References

Altman DG and Bland JM (1994) Diagnostic tests 2: predictive values. *British Medical Journal* **309**, 102.

Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB *et al.* (2015) Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia* **1**, 15030.

Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E *et al.* (2008) Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Archives of General Psychiatry* **65**, 28–37.

Cannon TD, Yu C, Addington J, Bearden CE, Cadenhead KS, Cornblatt BA *et al.* (2016) An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry* **173**, 980–988.

C.D.C.P (2012) Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years –Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *Centers for Disease Control and Prevention: Surveillance Summaries* 1–19.

Couzin-Frankel J (2017) A change of mind. *Science* **358**, 856–859.

Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L *et al.* (2017) Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science Translational Medicine* **9**, eaag2882.

Fusar-Poli P, Rutigliano G, Stahl D, Davies C, Bonoldi I, Reilly T *et al.* (2017) Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry* **74**, 493–500.

Gronborg TK, Schendel DE and Parner ET (2013) Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. *JAMA Pediatrics* **167**, 947–953.

Hafeman DM, Merranko J, Goldstein TR, Axelson D, Goldstein BI, Monk K *et al.* (2017) Assessment of a person-level risk calculator to predict New-onset bipolar spectrum disorder in youth at familial risk. *JAMA Psychiatry* **74**, 841–847.

Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T *et al.* (2011) Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry* **68**, 1095–1102.

Hayden EC (2017) The rise, fall and rise again of 23andMe. *Nature* **550**, 174–177.

Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ *et al.* IBIS NETWORK (2017) Early brain development in infants at high risk for autism spectrum disorder. *Nature* **542**, 348–351.

Howsmon DP, Kruger U, Melnyk S, James SJ and Hahn J (2017) Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation. *PLoS Computational Biology* **13**, e1005385.

Just MA, Pan L, Cherkassky VL, Mcmakin DL, Cha C, Nock MK *et al.* (2017) Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour* **1**, 911–919.

Mausner JS and Kramer S (1985) *Epidemiology: An Introductory Text*. Saunders: Philadelphia, PA.

Pramparo T, Pierce K, Lombardo MV, Carter Barnes C, Marinero S, Ahrens-Barbeau C *et al.* (2015) Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* **72**, 386–394.

Skafidas E, Testa R, Zantomio D, Chana G, Everall IP, Pantelis C (2014) Response to Belgard *et al. Molecular Psychiatry* **19**, 407–409.

Studerus E, Ramyead A and Riecher-Rossler A (2017) Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. *Psychological Medicine* **47**, 1163–1178.

Sundermann B, Herr D, Schwindt W and Pfleiderer B (2014) Multivariate classification of blood oxygen level-dependent FMRI data with diagnostic intention: a clinical perspective. *AJNR American Journal of Neuroradiology* **35**, 848–855.

Szatmari P, Chawarska K, Dawson G, Georgiades S, Landa R, Lord C *et al.* (2016) Prospective longitudinal studies of infant siblings of children with autism: lessons learned and future directions. *Journal of the American Academy of Child and Adolescent Psychiatry* **55**, 179–187.

Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y *et al.* (2016) A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature Communication* **7**, 11254.

https://doi.org/10.1017/S0033291718000673 Published online by Cambridge University Press