Psychological Medicine

cambridge.org/psm

Commentary: the ethical challenges of machine learning in psychiatry: a focus on data, diagnosis, and treatment

Daniel S. Barron[1,2,3,4]

[1]Department of Psychiatry, Yale University, New Haven, CT, USA; [2]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA; [3]Department of Psychiatry, Brigham & Women's Hospital, Harvard University, Boston, MA, USA and [4]Department of Anesthesiology & Pain Medicine, Brigham & Women's Hospital, Harvard University, Boston, MA, USA

Invited Commentary

Cite this article: Barron DS (2021). Commentary: the ethical challenges of machine learning in psychiatry: a focus on data, diagnosis, and treatment. *Psychological Medicine* **51**, 2522–2524. https://doi.org/10.1017/S0033291721001008

Received: 4 January 2021
Revised: 22 February 2021
Accepted: 1 March 2021
First published online: 12 May 2021

Key words:
Schizophrenia; diagnosis; machine learning

Author for correspondence:
Daniel S. Barron,
Email: daniel.s.barron@yale.edu

Which data are useful?

The clinical interview is the psychiatrist's data gathering procedure. However, the clinical interview is not a defined entity in the way that 'vitals' are defined as measurements of blood pressure, heart rate, respiration rate, temperature, and oxygen saturation. There are as many ways to approach a clinical interview as there are psychiatrists; and trainees can learn as many ways of performing and formulating the clinical interview as there are instructors (Nestler, 1990). Even in the same clinical setting, two clinicians might interview the same patient and conduct very different examinations and reach different treatment recommendations. From the perspective of data science, this mismatch is not one of personal style or idiosyncrasy but rather one of uncertain salience: neither the clinical interview nor the data thereby generated is operationalized and, therefore, neither can be rigorously evaluated, tested, or optimized.

Consider a standard psychiatric evaluation, wherein a thorough clinical interview will span a patient's biologic, psychologic, and social history. A clinical interview might yield thousands of datapoints that can range from a patient's visible and audible behavior (posture, speech, and expression); their reported narrative and symptomatology; results from clinical tests like blood work, urine toxicology, and electroencephlogram (EKG); collateral information from family members, legal authorities, or other health care providers; and the patient's socioeconomic status. Whether a clinical datapoint is useful is a testable hypothesis, one which depends on the specific *use* in question; for example, a patient's response to an selective serotoninreuptake inhibitor (SSRI) over 4–6 weeks (Chekroud et al., 2017) might be useful to an outpatient clinician but not useful to an emergency room psychiatrist assessing a patient's acute suicide risk (Just et al., 2017).

Defining and operationalizing which clinical data are useful for which decisions are no small matter, one that decades of research have been unable to answer. And yet, the very thing that machine learning (ML) algorithms offer is the ability to identify data that optimize some yet undefined purpose. The question becomes which purpose to optimize. Two answers might lie in diagnosis and treatment.

Why diagnose?

Schizophrenia is not schizophrenia in the way that hypertension is hypertension. Hypertension is diagnosed in one way: measured blood pressure is greater than a defined value. Though schizophrenia is also a defined diagnosis, if we consider the criteria for schizophrenia (see Table 1), there are 7 696 580 419 045 sets of symptoms that meet both criteria A and B as defined in the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) (SCID-5) (First, Williams, & Karg, 2016). Crucially, these sets do not differentiate symptom severity: for example, two patients might have so-called 'tangential speech,' but *how* tangential is irrelevant to diagnosis.

Because no quantitative measures exist for the signs or symptoms of schizophrenia, 'mild' is the only modifier that Stark could apply to patient D's 'psychotic symptoms' (of which there could be many variants, see Table 1). Contrast 'mild' with a blood pressure of 200/120, which can be readily understood in relation to 120/80. It is not at all clear whether or to what extent D's symptoms overlap with R's or T's (and, indeed, it is statistically unlikely that they do). And yet, Starke accurately describes how each patient's unknown symptom profile would be represented in an ML study: schizophrenia.

The larger purpose of diagnosis in psychiatry remains unclear. Current psychiatric diagnoses are not motivated by etiology or treatment or symptom severity. Psychiatric diagnosis is a vestige of the pre-computer era: in a world of hand-written clinical notes, diagnosis's virtue was to tidily communicate and standardize the general flavor of a clinical interview (Lieberman & Ogas, 2015). In this sense, psychiatric diagnosis met (and meets) its mark: although

CAMBRIDGE UNIVERSITY PRESS

Table 1. Schizophrenia is not schizophrenia

Case 1: R is presenting with newly developed negative and positive symptoms at a university psychiatry department. Based on a clinical interview, R is diagnosed with schizophrenia by a psychiatrist			
Case 2: D is presenting at a psychiatric day-clinic with mild psychotic symptoms and is diagnosed with schizophrenia after a clinical interview			
Case 3: T is diagnosed with the first episode of schizophrenia based on a clinical interview			
SCID-5 core symptoms required for schizophrenia diagnosis		Minimum required	Possible sets
A.	1. Delusions: reference, persecutory, grandiose, somatic, guilt, jealous, religious, erotomanic, being controlled, thought insertion, thought withdrawal, thought broadcasting, other (B1-B13, 13 total)	2	1 099 511 488 435[a]
	2. Hallucinations: auditory, visual, tactile, somatic, gustatory, olfactory (B14-B19, six total)		
	3. Disorganized speech: derailment, tangential, neologism, word salad (B20, four total)		
	4. Grossly disorganized behavior: dress, sexually inappropriate, agitation (B21, three total); or catatonic behavior: stupor, grimacing, mannerisms, posturing, agitation, stereotypy, mutism, echolalia, negativism, echopraxia, catalepsy, waxy flexibility (B22, 12 total)		
	5. Negative Symptoms: diminished emotional expression or avolition (B23–24, 2 total)		
B.	Decreased level of function: work, interpersonal relationships, self-care (3 total)	1	7[a]
Total subsets			7 696 580 419 045[a]

Each case vignette from Starke et al., (2020) actually describes one of 7 696 580 419 045 possible types of schizophrenia, based on how many sets of symptoms meet SCID-5 criteria for schizophrenia (First et al., 2016).

[a]Based on the SCID-5, Criteria A is met if at least two of the A-criteria symptoms are present, and at least one symptom is from either A1, A2, or A3. Mathematically, the total combination of symptoms that meet criteria A can be represented as a power set. To compute the total symptom sets possible across A1–A5, we simply calculate $[(2^{40})]$. From this total, we subtract the number of unwanted or redundant symptom sets. A set can be unwanted in two ways: (1) if it only includes symptoms from A4 and A5, $([(2^{17})]$; (2) if it involves symptoms from a single A group, some of which are already accounted for in (1) with the remaining from A1, A2, A3: $[(2^{13}-1)+(2^{6}-1)+(2^{4}-1))]$. So overall, the total number of symptom sets for criterion A is: $[(2^{40})] - [(2^{17})+(2^{13}-1)+(2^{6}-1)+(2^{4}-1))] = =1\ 099\ 511\ 488\ 435$. Criteria B adds $(2^{3}-1)=7$ sets. So there are a total of $1\ 099\ 511\ 488\ 435 \times 7 = 7\ 696\ 580\ 419\ 045$ sets of symptoms that meet SCID-5 criteria for schizophrenia. Of course, this number assumes that each individual symptom has a clear, monolithic meaning (which they do not). The author acknowledges and is grateful for the mathematical assistance of Drs. Leo A. Harrington (Department of Mathematics, UC Berkeley), W. Hugh Woodin (Department of Mathematics & Philosophy, Harvard University), and Gabriel Goldberg (Department of Mathematics, UC Berkeley) who, separately, helped me converge on the above solution.

Abbreviation: SSCID-5, Structured Clinical Interview for DSM-5.

schizophrenia can be diagnosed in 7 696 580 419 045 ways, most clinicians (and even non-clinicians) still have a notion for what is communicated by 'schizophrenia' and how this differs from, say, 'PTSD' (Young, Lareau, & Pierre, 2014). Diagnosis is a latent variable, a summary statistic of salient information from the clinical interview, varied as it might be. While data loss is necessary to define any summary statistic, data scientists are understandably suspicious of a latent variable that represents at least 7 696 580 419 045 sets of symptoms, each set possibly representing a unique etiology or pathophysiology.

Now that technology is relieving some of the burden of data creation, storage, and transmission, we might ask ourselves: if the best a diagnosis can offer is a latent variable summarizing a clinical interview, then why not produce high-definition audio and visual recording of the entire interview without any loss of data? Furthermore, given the complexity that arises in defining ground truth for a psychiatric diagnosis, ML analyses have begun to look for mechanistic understanding that might more ably pair clinical data with underlying biology or etiology (Bzdok & Ioannidis, 2019). It could be that ML classifiers might represent an evolution beyond psychiatric diagnostic groupings and that the question of which clinical data are most relevant might be better answered by data scientists than by clinicians.

How best to treat?

As Starke describes, many ML studies attempt to circumvent diagnoses entirely and inform treatment. This parallels what a clinician's brain does: gather and sift through clinical data primarily to inform treatment and, only later, to diagnose (Waszczuk et al., 2017). This makes sense given the lack of specificity between diagnosis and treatment; antipsychotics, antidepressants, and mood stabilizers are routinely used in treating psychosis or depression or mood instability. Furthermore, patients with the same diagnosis often receive different treatments: Starke's patient R might be prescribed an antipsychotic and antidepressant while patient T is prescribed only an antipsychotic.

ML studies very well might help clinicians optimize treatment, yet as Stark notes, examples should be taken with a grain of salt: there is no consensus on how to measure treatment outcome in psychiatry (Zimmerman, Morgan, & Stanton, 2018). For example, would antipsychotic treatment be 'successful' if patient R's hallucinations decrease by 50%? By 90%? What if the hallucinations stop entirely but, even though R no longer requires frequent hospitalization, R cannot return to university because the treatment itself is too sedating? Or what if R's hallucinations do not dissipate but they *are* able to return to university? There is no clear answer to this question and, I suspect, any ML analysis attempting to optimize treatment selection would require not simply exhaustive phenotyping but also a 'personalized tuning' of the algorithm based on that patient's unique goals and expectations for treatment (Barron, 2021). There, very well, maybe as many definitions of treatment success as there are patients in treatment and, even so tailored, that definition might change over time.

Overall, it remains possible that ML algorithms and the data scientists that produce them might bring clarity to the questions raised by decades of research. Starke's discussion of the ethical challenges for ML algorithms in psychiatry was a welcome addition to the growing dialogue. At base, the moral virtue of an algorithm is not simply whether it works but what it does, and for whom.

Acknowledgements. I thank the editors for inviting me to comment on Starke et al.'s ((Starke, Clercq, Borgwardt, & Elger, 2020); hereafter, Starke)

https://doi.org/10.1017/S0033291721001008 Published online by Cambridge University Press

timely paper about the ethical challenges for ML in psychiatry. I hope to further magnify three challenges, which are fundamental questions about data, diagnosis, and treatment in psychiatric disorders.

References

Barron, D. (2021). *Reading Our minds: The rise of Big data psychiatry*. New York City: Columbia University Press, Columbia Global Reports.

Bzdok, D., & Ioannidis, J. P. A. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in Neurosciences*, *42*, 251–262.

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., & McCarthy, G. (2017). Reevaluating the efficacy and predictability of antidepressant treatments: A symptom clustering approach. *JAMA Psychiatry*, *74*, 370.

First, M. B., Williams, J., & Karg, R. (2016) Structured Clinical Interview for DSM-5 Disorders (SCID-5), Clinician Version (SCID-5-CV).

Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., & Brent, D. (2017). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour*, *1*, 911–919.

Lieberman, J. A., & Ogas, O. (2015). *Shrinks: The untold story of psychiatry*. New York, NY: Back Bay Books.

Nestler, E. J. (1990). The case of double supervision: A resident's perspective on common problems in psychotherapy supervision. *Academic Psychiatry*, *14*, 129–136.

Starke, G., Clercq, E. D., Borgwardt, S., & Elger, B. S. (2020). Computing schizophrenia: Ethical challenges for machine learning in psychiatry. *Psychological Medicine*, 1–7. https://doi.org/10.1017/S0033291720001683.

Waszczuk, M. A., Zimmerman, M., Ruggero, C., Li, K., MacNamara, A., Weinberg, A., … Kotov, R. (2017). What do clinicians treat: Diagnoses or symptoms? The incremental validity of a symptom-based, dimensional characterization of emotional disorders in predicting medication prescription patterns. *Comprehensive Psychiatry*, *79*, 80–88.

Young, G., Lareau, C., & Pierre, B. (2014). One quintillion ways to have PTSD comorbidity: Recommendations for the disordered DSM-5. *Psychological Injury and Law*, *7*, 61–74.

Zimmerman, M., Morgan, T. A., & Stanton, K. (2018). The severity of psychiatric disorders. *World Psychiatry*, *17*, 258–275.

https://doi.org/10.1017/S0033291721001008 Published online by Cambridge University Press