AMA Journal of Ethics®
September 2018, Volume 20, Number 9: E804-811

CASE AND COMMENTARY

Is It Ethical to Use Prognostic Estimates from Machine Learning to Treat Psychosis?

Nicole Martinez-Martin, JD, PhD, Laura B. Dunn, MD, and Laura Weiss Roberts, MD, MA

Abstract

Machine learning is a method for predicting clinically relevant variables, such as opportunities for early intervention, potential treatment response, prognosis, and health outcomes. This commentary examines the following ethical questions about machine learning in a case of a patient with new onset psychosis: (1) When is clinical innovation ethically acceptable? (2) How should clinicians communicate with patients about the ethical issues raised by a machine learning predictive model?

Case

Dr K is a psychiatrist who regularly attends in an inpatient psychiatric ward at an academic medical center. In this role, Dr K regularly sees patients admitted from the emergency department who present with new onset psychosis. A major challenge with these patients is that clinicians are unable to predict an individual patient's clinical outcomes: some return to baseline, others experience only mild symptoms, while others deteriorate and might even become severely disabled.

Dr K is interested in piloting a study based on a recently published predictive model for patients who present with their first episode of psychosis.[1] The model was developed by applying machine learning methods to a large, multisite European database of patients with psychosis and offers 2 potentially helpful pieces of information to clinicians. First, the model yields a prognostic estimate. Using the patient's baseline information such as sex, occupational status, and history of major depressive episodes, the model predicts whether the patient will have a good or a poor outcome 1 year later. A good or poor outcome is defined by the Global Assessment of Function (GAF), a validated method for quantifying a patient's overall functional status. A good outcome—defined as a GAF score of greater than or equal to 65—typically indicates that a patient is able to function with minimal impairments. A poor outcome—a GAF score of less than 65—can indicate a broad range of impairment severity.[2] At GAF scores at the higher end of poor outcome (50-65), patients can experience moderate impairment in their social or occupational functioning. At GAF scores at the lower end of poor outcome (0-10), patients might be

unable to handle their own personal hygiene or be persistently suicidal. Of note, the model predicts a good or poor outcome with approximately 75% accuracy. Second, the model guides treatment choice for some patients. Although psychiatrists have access to a variety of antipsychotic agents to treat psychosis, patients predicted to have a poor outcome benefit more from amisulpride or olanzapine than other agents.

Dr K and colleagues think this model can enhance their treatment of psychotic patients and would like to incorporate it into their practice. However, they wonder whether the prognostic estimate, in particular, should be disclosed. While this information might help patients and their families plan and make decisions, they also wonder whether, when, and how this information could cause more harm than good.

Commentary

Dr K is considering piloting a predictive model for patients with first-episode psychosis that relies on machine learning applied to large data sets drawn from European sites and patients. Machine learning is a technique used to build algorithms for computational analysis that improves as a function of experience.[3] Algorithms can be used to analyze massive data sets to determine patterns and predict future outcomes. Machine learning is expected to bring major advances to psychiatry by improving prediction, diagnosis, and treatment of mental illness.[4,5] The above scenario illustrates some of the ethical considerations that will arise as machine learning techniques move from the lab to the clinic. Although the model in this case has been statistically validated, it is not yet validated as a clinical intervention that will lead to improved outcomes. This essay first examines whether Dr K is ethically justified in implementing this clinical innovation. We then discuss whether the target population for the predictive algorithm—ie, patients with psychotic disorders—raises special ethical issues regarding informed consent that should be considered.

When Should Clinicians Implement a Clinical Innovation?

Dr K's piloting of the predictive model would be considered a clinical innovation—that is, a novel use of an intervention or model that has not been shown to be definitively clinically superior to standard practice. Clinical innovation falls into a category somewhere between clinical practice and research, as these activities were distinguished in terms of their ethical mandates in the Belmont Report.[6] What would constitute sufficient ethical justification to implement the clinical innovation described in this case?

First, there must be a demonstrated need for the innovative practice.[7] Psychotic disorders exert a considerable personal, social, and financial burden on those affected. The recovery rate (10%-15%) after a first episode of psychosis, with routine clinical care, has remained the same for decades.[8] Timely intervention after a first episode of psychosis and treatment with antipsychotic medications can improve outcomes,[9] but treatment tolerance, adherence, and response can be highly variable.[10,11] Given the

potential severity of new onset psychosis, as well as the lack of adequate treatments, there is a demonstrable need for the proposed innovation.

Next, we must consider whether the risk posed by the innovation is ethically acceptable relative to risks of the underlying condition.[7] First, Dr K will need sufficient evidence that the proposed innovation can deliver the promised benefit. While accuracy of the proposed psychosis predictive model is supported by the study conducted in Europe, it is not known whether variables present in the local context—such as differences in psychiatric practice and social support—would affect the model's validity and ability to improve outcomes for Dr K's patient population. The model will need to be calibrated to account for relevant local variables.[12] Because of the "black box" nature of machine learning algorithms, software developers do not always know or might not understand how the system has used input data to arrive at decisions.[13] Thus, designers of the system will not likely know exactly which variables need to be addressed to validate the model for a new context; additional patients' data from the local clinical setting will be needed to perform a calibration.

Calibration will need to take into account not only local variables but also error and bias. Machine learning is often presented as more objective than human judgment, but it is susceptible to operator error. When faulty data are used as input, flawed analyses can result.[14] Machine learning algorithms can also reinforce existing biases in data.[15] For example, depending upon the way an algorithm accounts for socioeconomic status or race, decisions made on the basis of that algorithm could unintentionally reinforce existing structural deficits for vulnerable patients. On the other hand, with proper calibration, the algorithm could be used to reduce bias in health care. Finally, use of the predictive model will itself influence the care patients receive, impacting how psychiatrists make treatment decisions and allocate resources. Initially, the effect of the predictive model on cases might not be adequately accounted for in its analyses. In order to ensure an ethically acceptable balance of risks and benefits in implementing a predictive model, clinicians will need to be actively involved in validating the predictive algorithm in the local context by ensuring that the calculations are attuned to the particular patient population and by outlining the associated protocols for moving from prediction to treatment.

At the same time, physicians using the algorithm may not know the variables and rationale behind predictions it generates, making it difficult for them to assess and justify resulting treatment decisions. Justifying use of a predictive model will require addressing issues of transparency and bias that arise in the use of machine learning systems by implementing strategies such as training physicians on how machine learning systems work, including physicians in their creation, and even supporting efforts to implement machine learning systems that can give insight into the reasons for their predictions. Clinicians who use the machine learning systems will need to learn more about how they

are constructed, the underlying data sets that inform their recommendations, and their limitations.[16]

Informed Consent for the Target Population

In order to ensure trust and transparency in using predictive models, there must be careful attention to ethical issues related to informed consent. Currently, informed consent is not explicitly required to use patients' data in applying and improving predictive algorithms.[12] Furthermore, patients are generally not aware when physicians use computer-based decision aids in the course of their care and are rarely informed of sources that inform their physician's judgment.[12] These facts raise the question: Do machine learning predictive algorithms such as this psychosis prediction model involve novel ethical issues that necessitate a different ethical approach?

How machine algorithms differ from existing risk assessment tools, such as those used to assess risk of heart attack or stroke, has to do with their potential impact on therapeutic relationships. As physicians increasingly turn to machine learning algorithms to inform diagnostic and treatment decisions, these algorithms might become more than just support tools.[16] Furthermore, as machine learning systems are integrated into health care settings, decisions regarding treatment or resource allocation that stem from a predictive tool could come from rules or protocols set by hospital administration rather than a treating physician. Thus, machine learning tools can reconfigure physicians' roles in their relationships with patients.[16] As machine learning systems become more integrated into care, careful examination of the fiduciary dimension of relationships between patient and machine learning decision systems in health care institutions will be needed.[16]

Because technology can intrude upon patient-clinician relationships by influencing how a physician makes decisions and directs resources to care for a patient and will impact confidentiality as machine learning systems are integrated with electronic health records,[16] patients should be notified about uses of predictive algorithms at their health care institutions. Patients will need sufficient information to consider how machine learning systems can influence their care, the confidentiality of their information, and the privacy of their data. We suggest that patients should be alerted that their data could be used to formulate or improve predictive algorithms and that predictive tools might play a role in their care. In the case of early psychosis, decisions would need to be made about when to notify patients, given that patients and families are invariably coping with severe disease-related and psychosocial stress at the time of patients' hospitalization for psychosis that could make it even more difficult to digest and retain complex information, such as the use of predictive algorithms. Community stakeholders could provide input on how to formulate the content of such notices and the procedures for engaging with patients and families meaningfully.[12]

Should a prognosis delivered by a predictive algorithm be disclosed to a patient as a part of informed consent for treatment? Informed consent does not require explanation of *all* details that inform a treatment recommendation, but it does require that explanation of *pertinent* information about the nature, risks, and benefits of treatment options be conveyed to a patient.[17,18] Therefore, clinicians would need sufficient education regarding a machine learning system in order to communicate information about an algorithm's treatment recommendation. Before disclosing a prognosis generated by a predictive model, it would be helpful to have at least some data generated by a machine learning algorithm on the effects of sharing a prognosis on patient distress and outcomes.

Given that Dr K's patients have new onset psychosis, there might be concerns that providing information regarding the algorithm's prediction could lead to psychological distress in some patients or their families. In general, assumptions that persons with severe mental illness have impaired ability to make autonomous and well-informed research and treatment decisions have frequently not borne up under rigorous scrutiny.[19,20] Such concerns need to be empirically examined rather than accepted at face value.[21] Patients might want more or less detail regarding treatment depending upon factors such as their education levels, how they assess their own decisional capacity, or their satisfaction with treatment.[22] The capacity for voluntarism—ie, the ability to make choices that are free from coercion and are consonant with an individual's values and history—is another critical component of informed consent,[18] one that necessitates engaging with patients to discern their preferences in the context of specific decisions. Attending to the individual needs and capacity of a patient for informed consent remains key, including supporting a patient's capacity to engage meaningfully in health care decisions and identifying tools that help assess decisional capacity,[23] especially relative to understanding predictive algorithms.

Conclusion

In order to implement the predictive tool in an ethical manner, Dr K will need to carefully consider how to give appropriate information—in an understandable manner—to patients and families regarding use of the predictive model. In order to maximize benefits from the predictive model and minimize risks, Dr K and the institution as a whole will need to formulate ethically appropriate procedures and protocols surrounding the instrument. For example, implementation of the predictive tool should consider the ability of a physician to override the predictive model in support of ethically or clinically important variables or values, such as beneficence. Such measures could help realize the clinical application potential of machine learning tools, such as this psychosis prediction model, to improve the lives of patients.

References

1. Koutsouleris N, Kahn RS, Chekroud AM, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry*. 2016;3(10):935-946.
2. Hall RC. Global assessment of functioning. A modified scale. *Psychosomatics.* 1995;36(3):267-275.
3. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. 2018;14:91-118.
4. Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med*. 2016;46(12):2455-2465.
5. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *JAMA*. 2016;315(6):551-552.
6. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html. Published April 18, 1979. Accessed July 23, 2018.
7. Ghaemi SN, Goodwin FK. The ethics of clinical innovation in psychopharmacology: challenging traditional bioethics. *Philos Ethics Humanit Med*. 2007;2(1):26. doi:10.1186/1747-5341-2-26.
8. Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry*. 2017;16(3):251-265.
9. Singh SP. Early intervention in psychosis. *Br J Psychiatry*. 2010;196(5):343-345.
10. Stevenson JM, Reilly JL, Harris MS, et al. Antipsychotic pharmacogenomics in first episode psychosis: a role for glutamate genes. *Transl Psychiatry*. 2016;6:e739. doi:10.1038/tp.2016.10.
11. Millan MJ, Andrieux A, Bartzokis G, et al. Altering the course of schizophrenia: progress and perspectives. *Nat Rev Drug Discov*. 2016;15(7):485-515.
12. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff (Millwood)*. 2014;33(7):1139-1147.
13. Maini V. Machine learning for humans. *Medium*. August 19, 2017. https://medium.com/machine-learning-for-humans/why-machine-learning-matters-6164faf1df12. Accessed March 12, 2018.
14. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2017;3(3):223-230.
15. Tunkelang D. Ten things everyone should know about machine learning. *Forbes*. September 6, 2017. https://www.forbes.com/sites/quora/2017/09/06/ten-things-everyone-should-know-about-machine-learning/. Accessed January 13, 2018.

16. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378(11):981-983.
17. Hall DE, Prochazka AV, Fink AS. Informed consent for clinical treatment. *CMAJ*. 2012;184(5):533-540.
18. Roberts LW. Informed consent and the capacity for voluntarism. *Am J Psychiatry*. 2002;159(5):705-712.
19. Dunn LB, Candilis PJ, Roberts LW. Emerging empirical evidence on the ethics of schizophrenia research. *Schizophr Bull*. 2006;32(1):47-68.
20. Humphreys K, Blodgett JC, Roberts LW. The exclusion of people with psychiatric disorders from medical research. *J Psychiatr Res*. 2015;70:28-32.
21. Roberts LW, Roberts B. Psychiatric research ethics: an overview of evolving guidelines and current ethical dilemmas in the study of mental illness. *Biol Psychiatry*. 1999;46(8):1025-1038.
22. Hamann J, Mendel R, Reiter S, et al. Why do some patients with schizophrenia want to be engaged in medical decision making and others do not? *J Clin Psychiatry*. 2011;72(12):1636-1643.
23. Palmer BW, Harmell AL. Assessment of healthcare decision-making capacity. *Arch Clin Neuropsychol*. 2016;31(6):530-540.

Nicole Martinez-Martin, JD, PhD is a postdoctoral fellow at the Stanford Center for Biomedical Ethics in Stanford, California. She attained a JD from Harvard Law School and a PhD from the University of Chicago in comparative human development. Her research focuses on neuroethics as well as the ethics of digital health technology and machine learning with a focus on mental health issues and special populations.

Laura B. Dunn, MD is a professor of psychiatry and behavioral sciences, the director of the Geriatric Psychiatry Fellowship Training Program, and the section chief of Geriatric Psychiatry in the Department of Psychiatry and Behavioral Sciences at Stanford University School of Medicine in Stanford, California. Her research focuses on ethical issues in clinical research, including informed consent, decision-making capacity, and influences on research participation, and she has published extensively on empirical ethics issues in vulnerable populations.

Laura Weiss Roberts, MD, MA serves as the chair and the Katharine Dexter McCormick and Stanley McCormick Memorial Professor in the Department of Psychiatry and Behavioral Sciences at Stanford University School of Medicine in Stanford, California. She has received scientific, peer-reviewed funding from the National Institutes of Health, the US Department of Energy, and private foundations to perform empirical studies of modern ethical issues in research, clinical care, and health policy with a particular focus on vulnerable and special populations.

Editor's Note
The case to which this commentary is a response was developed by the editorial staff.

Citation
AMA J Ethics. 2018;20(9):E804-811.

DOI
10.1001/amajethics.2018.804.

Conflict of Interest Disclosure
Dr Dunn is a consultant to Otsuka America Pharmaceuticals, Inc., and a member of Lundbeck's advisory boards. The other authors had no conflicts of interest to disclose.

The people and events in this case are fictional. Resemblance to real events or to names of people, living or dead, is entirely coincidental. The viewpoints expressed in this article are those of the author(s) and do not necessarily reflect the views and policies of the AMA.

ISSN 2376-6980