

## COMMENT OPEN



# From models to tools: clinical translation of machine learning studies in psychosis

Andrea Mechelli<sup>1</sup> and Sandra Vieira<sup>1</sup>npj Schizophrenia (2020)6:4; <https://doi.org/10.1038/s41537-020-0094-8>

## INTRODUCTION

The past decade has seen a proliferation of neuroimaging-based machine learning studies in psychosis.<sup>1</sup> Furthermore, within the span of ten years, small local studies with few dozen participants have evolved into large multi-centre studies with several hundreds of participants.<sup>2–5</sup> In the midst of the search for accurate models, much attention has been given to methodological challenges including the impact of sample size,<sup>6,7</sup> the limitations of traditional case–control designs,<sup>8,9</sup> how to best deal with confounding variables<sup>10</sup> and the effects of heterogeneity<sup>11,12</sup> and inter-scanner variability,<sup>13</sup> just to mention a few. Although there are still important methodological challenges to overcome, substantial progress is being made, and a solution to these challenges is now considered to be a matter of when rather than if.<sup>14,15</sup> Wider discussions in the medical community about the ethical and legal implications of integrating machine learning models within diagnostic and prognostic assessment of patients are also underway.<sup>16–20</sup> Taken collectively, the progress being made towards the development and validation of neuroimaging-based machine learning models is encouraging, as if the different pieces of a very complex puzzle were slowly coming together. Less discussed however, are the challenges related to the development and validation of machine learning-based clinical tools. Here the critical distinction is between “models”, which tend to be developed and validated using a limited number of well characterised datasets with the aim of maximising accuracy, sensitivity and specificity, and “tools”, which must be feasible, acceptable and safe, and provide information that will guide clinical decision-making in real-world settings. This is a timely discussion, as a new generation of multi-centre studies aiming to develop machine learning tools to manage patients with psychosis is emerging (e.g., PSYSCAN,<sup>21</sup> PRONIA—[www.pronia.eu](http://www.pronia.eu)).

Let's imagine that we have developed a neuroimaging-based machine learning model with high levels of accuracy, sensitivity and specificity, after addressing the main methodological issues.<sup>2</sup> Next we'd like to translate this machine learning model into an actual clinical tool to support the assessment of individual patients. What are our main challenges along this translation? In this opinion piece, we discuss seven critical aspects that require careful consideration when moving from a “model” to a “tool”. These include real-world validation, clinical utility, feasibility, acceptability, safety and finally, dissemination.

### Real-world validation

After validating our model using several independent datasets, collected using different scanners across multiple research sites, we might feel reassured about its performance in a real-world setting. Yet our optimism might be premature. This is because

datasets collected for the purpose of research tend to include patients who meet stringent inclusion/exclusion criteria; unfortunately this highly selected group of patients differ from service users who do not take part in research (e.g., less severe, lower comorbidities, less medicated, and higher functioning).<sup>22</sup> Therefore, when it comes to clinical validation, we need to consider not only the size but also the type of sample. In practice, the validation of a clinical tool should be done in a naturalist design, where all service users that may benefit from the tool are approached whilst having minimal exclusion criteria. It is likely that this will result in lower accuracies, sensitivities and specificities than the original validation using research datasets. The silver-lining here is that, if permitted, the more “naturalistic” data could be used to improve our tool. Learning from experience is, after all, one of the essential properties of machine learning.

### Clinical utility

Successful validation of our model using real-world data does not necessarily guarantee clinical utility. For a tool to be clinically useful, two conditions must be met. First, the tool must provide the treating clinicians with information which is not already available to them via conventional clinical assessment. Second, the tool must provide the treating clinicians with information that will influence a patient's clinical management. Based on these conditions, there are at least four areas of potential clinical utility in the case of psychosis: (i) prediction of conversion to psychosis in individuals at clinical high risk (i.e., conversion vs. non-conversion); (ii) prediction of diagnostic outcome (i.e., affective vs. non-affective psychosis); (iii) prediction of response to conventional antipsychotic medication (responders vs. non-responders); and (iv) prediction of psychotic relapse amongst patients who have recovered from the acute phase of the illness (relapsers vs. non-relapsers). In each of these examples, it is not possible to predict clinical outcome based on the initial clinical presentation (condition 1), and the prediction has practical implications for clinical management (condition 2). A critical implication of this conceptualization of clinical utility is that very high levels of accuracy, sensitivity and specificity do not necessarily guarantee clinical utility. For example, a tool developed to distinguish between patients experiencing psychosis and healthy volunteers would not add to conventional clinical assessment, and as such would not be considered clinically useful even in the context of perfect accuracy, sensitivity and specificity. Conversely, accuracy, sensitivity, and specificity do not necessarily need to be very high for a tool to have clinical utility. For example, in circumstances where the clinical decision represents a “toss-up” and is subsequently adjusted via trial and error, even a modest improvement in prediction (e.g., from 50 to 70%) may be

<sup>1</sup>Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ✉email: [a.mechelli@kcl.ac.uk](mailto:a.mechelli@kcl.ac.uk)

considered clinically useful. For completeness it should be mentioned that tools which do not meet both conditions, but have the potential of changing the patient's perceptions and behaviours around their illness, could also have some clinical utility. For example, a tool predicting that someone has a very high chance of benefiting from a certain medication may make them more tolerant of side effects and less likely to discontinue treatment.

#### Feasibility

Next we will need to establish the feasibility of implementing our tool in a real-world setting. In practice, this would require carrying out a prospective naturalistic investigation using both quantitative and qualitative methodologies. Feasibility from the perspective of healthcare providers could be assessed based on the proportion of clinical services, which have the capacity to collect the required neuroimaging data, and have access to the technical infrastructure to implement the tool. Here potential barriers include absence of scanning facilities and insufficient computational resources to process the images in a timely manner. Feasibility from the perspective of service users could be assessed based on the proportion of patients who meet the inclusion and exclusion criteria for our tool, and the proportion of acceptances, refusals and nonresponses. Here patients with acute psychosis might find the scanner environment too stressful, thereby limiting the feasibility of our tool in this phase of the illness. At present, the feasibility of implementing neuroimaging-based machine learning tools in clinical services for psychosis is unclear due to the paucity of empirical data.

#### Acceptability

Not much is known about the acceptability of machine learning tools in health care services in what has been recognized as "blind spot" in AI research.<sup>23</sup> Yet such knowledge is particularly important for mental health applications, where the use of AI may be perceived as more controversial.<sup>24</sup> Negative attitudes towards AI-based clinical tools amongst clinicians and service users may pose a significant obstacle to translational implementation, and as such should be an integral part of any formal assessment of clinical viability. As part of our prospective naturalistic study, we would therefore need to measure acceptability from the perspectives of clinicians and service users, covering domains such as helpfulness, ease of use, attractiveness, safety, and privacy features. In addition, because a questionnaire might miss potential barriers, we might also want to carry out face-to-face qualitative interviews with both groups. These interviews will provide an opportunity to unearth and understand concerns about technical aspects (e.g., unreliable internet connection) as well as operational aspects (e.g., medical staff being uncomfortable or uncertain about how the output of our tool can be integrated into clinical assessment).

#### Safety

A first source of risk in the translational implementation of our tool relates to the collection of brain scans. While MRI is generally regarded as a safe procedure, some service users may have conditions that may pose a risk (e.g., pregnancy). However, the screening for these conditions and other issues that may affect the quality of the image is a routine exercise in clinical settings, and as such should not be a significant barrier. On the other hand, experiencing stress and anxiety before and/or during the MRI is not uncommon,<sup>25</sup> and services users with acute psychosis may find the procedure even more challenging.<sup>26</sup> A second source of risk in the translational implementation of our tool relates to the possible misuse of its output by medical staff. Here the potential risk is that a clinician will misinterpret and/or misuse the

information with detrimental consequences for the patient. As part of our prospective naturalistic study, we would therefore need to measure safety in terms of frequency of adverse events and inappropriate use of the tool by medical staff.

#### Dissemination

Having established the clinical utility, feasibility, acceptability, and safety of our tool using real-world data, we are now ready to make it available to the wider clinical and research communities. This raises the question of how to best disseminate our healthcare innovation. This question requires careful consideration of our aims and values as healthcare innovators as well as the opportunities and limitations of the current market. On the one hand, we would like our tool to improve the way psychosis is diagnosed, monitored and treated, reducing the burden of the illness on patients, their carers and the wider society; in order to achieve this aim, it is imperative that our tool is developed for wide adoption and is made available to the widest possible cohort of patients. On the other hand, we need to ensure the long-term sustainability of our tool in light of its future running costs (e.g., technical updates, cloud-based hosting, and bug fixing); in order to achieve this aim a robust plan for market adoption, continuous development, and financial income generation over a sustained period will be required. We will therefore need to develop a dissemination and commercialization plan which combines these aims, covering aspects such as medical device regulations, unique value proposition, market size, and revenue streams, data governance and ethics, cost structure, partnerships, and key risks.<sup>27</sup>

#### CONCLUSION

Neuroimaging-based machine learning studies of psychosis are generating a plethora of academic publications, many of which are reporting "promising" findings. The ultimate aim of our research, however, is to find ways of decreasing the burden of this illness on patients, their carers, and the wider society. In order to achieve this aim, we need to start translating these machine learning models into clinical tools. This translation is not straightforward, as it requires us to move away from the kind of metrics that are the cornerstone of academic publications (e.g., statistical significance of the accuracy) towards measures of feasibility, acceptability, safety and, of course, clinical utility. A further stumbling block is that the vast majority of published studies compared patients with psychosis against healthy volunteers using a cross-sectional design, and as such produced findings with little or no clinical utility. In contrast, to develop machine learning models capable of providing clinically useful information, we need access to longitudinal data (for example, whether a patient did or did not respond to a full cycle of conventional antipsychotic medication). In the near future, a number of ongoing large-scale studies using a longitudinal design are expected to come to completion (e.g., PSYSCAN,<sup>21</sup> PRONIA—[www.pronia.eu](http://www.pronia.eu)). It is hoped that the data resulting from these studies will provide our research community with opportunities to bridge the existing gap between models and tools.

Received: 3 December 2019; Accepted: 24 January 2020;

Published online: 14 February 2020

#### REFERENCES

1. Kambeitz, J. et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology* **40**, 1742–1751 (2015).
2. Vieira, S. et al. Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence. *Schizophr. Bull.* (2019). <https://doi.org/10.1093/schbul/sby189>.

3. Rozycki, M. et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. *Schizophr. Bull.* **44**, 1035–1044 (2018).
4. Dluhoš, P. et al. Multi-center machine learning in imaging psychiatry: a meta-model approach. *Neuroimage* **155**, 10–24 (2017).
5. Lei, D. et al. Detecting schizophrenia at the level of the individual: relative diagnostic value of whole-brain images, connectome-wide functional connectivity and graph-based metrics. *Psychol. Med.* (2019). <https://doi.org/10.1017/S0033291719001934>.
6. Schnack, H. G. & Kahn, R. S. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Front. Psychiatry* **7**, 50 (2016).
7. Nieuwenhuis, M. et al. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage* **61**, 606–612 (2012).
8. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biol. Psychiatry* **80**, 552–561 (2016).
9. Sato, J. R., Rondina, J. M. & Mourão-Miranda, J. Measuring abnormal brains: building normative rules in neuroimaging using one-class support vector machines. *Front. Neurosci.* **6**, 178 (2012).
10. Rao, A., Monteiro, J. M. & Mourao-Miranda, J. Predictive modelling using neuroimaging data in the presence of confounds. *Neuroimage* **150**, 23–49 (2017).
11. Schnack, H. G. Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophr. Res.* (2017). <https://doi.org/10.1016/J.SCHRES.2017.10.023>.
12. Thomas, R. M., Bruin, W., Zhutovsky, P. & van Wingen, G. in *Machine Learning: Methods and Applications to Brain Disorders* (eds. Mechelli, A. & Vieira, S.) 249–266 (Academic Press, 2020). <https://doi.org/10.1016/B978-0-12-815739-8.00014-6>.
13. Pomponio, R. et al. Harmonization of large multi-site imaging datasets: application to 10,232 MRIs for the analysis of imaging patterns of structural brain change throughout the lifespan. *bioRxiv* 784363 (2019). <https://doi.org/10.1101/784363>.
14. Vieira, S., Pinaya, W. H. L. & Mechelli, A. in *Machine Learning: Methods and Applications to Brain Disorders* (eds. Mechelli, A. & Vieira, S.) 1–20 (Academic Press, 2020). <https://doi.org/10.1016/B978-0-12-815739-8.00001-8>.
15. Scarpazza, C., Baecker, L., Vieira, S. & Mechelli, A. in *Machine Learning: Methods and Applications to Brain Disorders* (eds. Mechelli, A. & Vieira, S.) 45–65 (Academic Press, 2020). <https://doi.org/10.1016/B978-0-12-815739-8.00003-1>.
16. Cabitza, F., Rasoini, R. & Gensini, G. F. Unintended consequences of machine learning in medicine. *JAMA* **318**, 517 (2017).
17. Yuste, R. et al. Four ethical priorities for neurotechnologies and AI. *Nat. News* **551**, 159 (2017).
18. Vayena, E., Blasimme, A. & Cohen, I. G. Machine learning in medicine: addressing ethical challenges. *PLoS Med.* **15**, e1002689–e1002689 (2018).
19. Char, D. S., Shah, N. H. & Magnus, D. Implementing machine learning in health care—addressing ethical challenges. *N. Engl. J. Med.* **378**, 981–983 (2018).
20. Kellmeyer, P. in *Machine Learning: Methods and Applications to Brain Disorders* (eds. Mechelli, A. & Vieira, S.) 329–342 (Academic Press, 2020). <https://doi.org/10.1016/B978-0-12-815739-8.00018-3>.
21. Tognin, S. et al. Towards precision medicine in psychosis: benefits and challenges of multimodal multicenter studies—PSYSCAN: translating neuroimaging findings from research into clinical practice. *Schizophr. Bull.* (2019). <https://doi.org/10.1093/schbul/sbz067>.
22. Patel, R. et al. What proportion of patients with psychosis is willing to take part in research? A mental health electronic case register analysis. *BMJ Open* **7**, e013113 (2017).
23. Crawford, K. & Calo, R. There is a blind spot in AI research. *Nature* **538**, 311–313 (2016).
24. Fiske, A., Henningsen, P. & Buys, A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J. Med. Internet Res.* **21**, e13216–e13216 (2019).
25. Meléndez, J. C. & McCrank, E. Anxiety-related reactions associated with magnetic resonance imaging examinations. *J. Am. Med. Assoc.* **270**, 745 (1993).
26. Falkenberg, I. et al. Clinical utility of magnetic resonance imaging in first-episode psychosis. *Br. J. Psychiatry* **211**, 231–237 (2017).
27. Greenhalgh, T. & Abimbola, S. The NASSS framework—a synthesis of multiple theories of technology implementation. *Stud. Health Technol. Inform.* **263**, 193–204 (2019).

## ACKNOWLEDGEMENTS

Prof. Andrea Mechelli is supported by a Wellcome Trust's Innovator Award (208519/Z/17/Z).

## AUTHOR CONTRIBUTIONS

A.M. and S.V. have co-developed and co-drafted this opinion piece.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to A.M.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020