Administration and Policy in Mental Health and Mental Health Services Research
https://doi.org/10.1007/s10488-020-01056-9

COMMENTARY

Can Artificial Intelligence Improve Psychotherapy Research and Practice?

Rachel L. Horn[1] · **John R. Weisz**[1]

Abstract

Leonard Bickman's article on the future of artificial intelligence (AI) in psychotherapy research paints an encouraging picture of the progress to be made in this field. We support his perspective, but we also offer some cautionary notes about the boost AI can provide. We suggest that AI is not likely to transform psychotherapy research or practice to the degree seen in pharmacology and medicine because the factors that contribute to treatment response in these realms differ so markedly from one another, and in ways that do not favor advances in psychotherapy. Despite this limitation, it seems likely that AI will have a beneficial impact, improving empirical analysis through data-driven model development, tools for addressing the limitations of traditional regression methods, and novel means of personalizing treatment. In addition, AI has the potential to augment the reach of the researcher and therapist by expanding our ability to gather data and deliver interventions beyond the confines of the lab or clinical office.

Keywords Artificial intelligence · Psychotherapy · Methodological limitations · Machine learning

Leonard Bickman's characteristically rich, thoughtful, and thought-provoking article "Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence" (2020) puts the spotlight on artificial intelligence (AI) as the future of psychotherapy research, or at least a big part of that future. AI, he argues, can help us finally overcome barriers that have impeded treatment progress over the last several decades. Dr. Bickman has a strong track record as a predictor of future directions in psychology, and he is likely to be right in this case as well. It does seem reasonable to expect that AI, and particularly machine learning, will ultimately enrich our understanding of psychopathology and psychotherapy, including when, with whom, and how to intervene.

In this article, we offer support for Bickman's core idea, noting some examples of what AI may bring to psychotherapy, but we also offer cautionary thoughts, including a perspective on the pace at which AI may lead to measurable benefits for psychotherapy. To preview the last point, we note that whereas the contributions of AI in medicine and the biological sciences are already robust, it is early days yet for applications to mental health. Much of the support for AI in psychotherapy consists of proof-of-concept or simulation studies or research that is still in progress and not yet published or peer-reviewed. While this research may prove valuable, it is perhaps best viewed at this point as preliminary, foundational work that may possibly set the stage for exciting directions in the days ahead. While noting the potential of AI, and our hope for breakthroughs, we think it is also useful to identify some limitations and challenges of work in this area to date, and that is where we begin…

Some Challenges AI May Face in Psychotherapy Research

The existing research that most clearly highlights the utility of AI for intervention is primarily focused on biological constructs. When it comes to pharmacological interventions,

This article is part of the Festschrift for Leonard Bickman Special Issue on The Future of Children's Mental Health Services.

✉ Rachel L. Horn
rachelhorn@g.harvard.edu

John R. Weisz
john_weisz@harvard.edu

[1] Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138, USA

Published online: 26 July 2020

research in the medical field is already showing the impressive utility of the more advanced methods, and this utility is likely to extend to psychopharmacology. Why is this the case? AI or machine learning for studying response to pharmaceutical treatment focuses mainly on the biological person. The genetic and other biological measures emphasized most in precision medicine may account for most of what's needed to select appropriate treatments and predict outcomes for a number of medical conditions, including some forms of cancer. It is possible that prescribing practices for anxiety—for example, whether a selective serotonin reuptake inhibitor or a selective norepinephrine reuptake inhibitor should be administered—may eventually be best predicted by genetic information and other biological data on the client that can be collected at baseline, making AI a boon to psychopharmacology.

In contrast, factors that are traditionally considered relevant to psychosocial intervention (e.g., personality or cognitions of the client, or clinician characteristics) may not account for large proportions of the variance in psychotherapy response. Moreover, the factors that make a difference in psychotherapy may be more numerous, more subtle or abstract, and more complex in the ways they combine with one another than is the case in medical treatment. Unlike the pharmacology example, psychotherapy for anxiety may involve a combination of psychoeducation, exposure, relaxation techniques, and other evidence-based treatment protocols, all of which must be delivered in an engaging way by a clinician who can keep the client motivated to persevere despite fear, create a mix of difficulty levels that includes some highly anxiety-provoking stimuli without overwhelming and discouraging the client, monitor the client's response to each attempt and appropriately adjust difficulty level of the next exposures, while building and sustaining the kind of relationship that keeps the client coming back until adaptive functioning is achieved. Speaking more generally, most psychological treatments focus on the cognitive, behavioral, and emotional aspects of a person far more than the biological features. These client features, which may vary dramatically between *and* within a person—in addition to the interactions between these features and the environment—would need to be accounted for by AI for it to be as precise and as useful as AI is for medical and pharmaceutical intervention. The therapist's personal and professional characteristics in addition to the treatment elements, both independently and in interaction with the client's traits and cognitions, would also need to be accounted for if an AI program is to assess the full array of factors potentially relevant to psychotherapy outcomes.

Is it really likely that these often subtle and usually dynamic personal characteristics and procedural calibrations could be fully captured by the kinds of measures AI would employ? Could the measurement process be sufficiently dense and real-time to capture moment-to-moment shifts and thus provide the moment-to-moment guidance a therapist would need during a session? In raising these questions, we are not suggesting that AI cannot be applied to psychological intervention. To the contrary, there is already evidence that machine learning can improve our ability to choose between different psychotherapy options (e.g., van Bronswijk et al. 2018). What we are suggesting is that without a massive increase in our measurement capabilities it is unlikely that AI will be able to have *as much* impact on psychosocial intervention research or practice as it has and will continue to have in the medical and pharmaceutical domains.

It is possible that some of the factors that may be particularly difficult to quantify in AI-friendly ways simply matter less in medical care than in psychological treatment. As one example, our personal feelings about our medical care provider probably have less impact on our medical outcomes than our personal feelings about our psychotherapist have on outcomes of anxiety or depression treatment. Our capacity to efficiently and accurately measure personal feelings toward others, which of course may shift frequently, is nowhere near our capacity to measure the genetic or otherwise biological characteristics that may figure most prominently in response to medical treatments.

Some Opportunities AI May Bring to Psychotherapy Research and Practice

Our reservations and caveats aside, we do believe—consistent with the Bickman perspective—that AI is likely to produce significant advances in psychotherapy research in the coming decades. Among these likely benefits, one is improved predictive modeling. The ability of AI to detect patterns without human direction offers a major advantage over traditional statistical methods. Traditional randomized controlled trial analyses generate models that quantify the effects for all specified variables and only specified relationships. When conducting a regression analysis of psychotherapy trial data, for example, the analyst generally specifies a treatment group by time interaction to verify that symptom change is related to the intervention under study. Other interactions may also be included if hypotheses or prior knowledge indicate a possible relationship. However, any interactions that are not directly specified in the model are not quantified, and important variable interactions therefore may not be incorporated into the final model. Excluding these meaningful terms worsens our predictive accuracy and limits our capacity for advancement to tests of our own a priori hypotheses and planned comparisons.

An additional limitation of traditional regression approaches is that we focus attention on statistically significant effects, largely dismissing the role of nonsignificant predictors, except in cases where an anticipated result did not materialize (e.g., when a treatment expected to be beneficial is not associated with symptom improvement). Such selective attention ignores the fact that nonsignificant variables do enter into our models and do influence outcome estimation, and their inclusion can impact predictive accuracy. By contrast, AI's ability to detect patterns can identify interactions among variables that would not be noted by the data analyst, and the final model can also exclude nonsignificant predictors and thereby prevent them from unduly influencing predictions. This is a valuable benefit of AI. To be clear, it is subject to some of the limitations of traditional methods. Insufficient power to detect effects, for example, can be as big a problem for AI as for more traditional methods; but experts are finding ways to address such limitations. As one illustration, researchers have developed methods like random forest to better cope with analyses of small samples with many variables (Gunduz and Fokoué 2015).

Another advantage is the power of machine learning to improve our identification of clinical populations and our ability to match interventions to the subgroups they are most likely to help. Regarding clinical population identification, Chung et al. (2018) trained a machine learning model to estimate a client's age using neuroimaging of a healthy sample. The model generated highly accurate age predictions when applied to healthy young people from a new sample. However, they found that this model consistently overestimated the age of youths at clinically high risk of developing psychosis who developed symptoms between the ages of 12 and 17, and greater deviation between model-predicted age and chronological age was associated with greater psychosis risk and poorer functioning in younger adolescents. Since there are very few circumstances under which a physician would not know the true age of a client, the disparity between model-predicted age and true age is a novel option for identifying clients at high risk for early psychosis onset and poor outcomes.

The literature testing artificial intelligence for enhancing treatment assignment encompasses a wide variety of statistical methods and clinical populations that collectively gives us hope for the future of personalized treatment assignment. As one example, van Bronswijk et al. (2019) used a two-step machine learning process to select variables for calculating the Personalized Advantage Index (PAI; DeRubeis et al. 2014) for clients who received cognitive therapy (CT) or interpersonal psychotherapy (IPT); their analysis revealed that reports of traumatic childhood experiences and recent life events were associated with treatment benefit, such that more of the aforementioned experiences was associated with greater benefit for IPT over CT. Lorenzo-Luaces et al. (2017) used machine learning to develop a prognostic index for depressed adults, including factors such as unemployment, hostility, and sleep problems that predicted outcomes. Though no differences were originally found between the 3 tiers of care being studied—treatment as usual, brief therapy, and cognitive-behavioral therapy (CBT)—analysis with the prognostic index revealed that clients with a poor prognosis showed greater benefit from CBT than either of the lower-tier interventions. These are only two examples of the multiple potential applications of machine learning to personalized treatment assignment; fuller accounts of these methods and their utility may be found elsewhere (e.g., DeRubeis 2019).

Another aspect of the promise of AI for psychotherapy is its capacity to expand the therapist's reach beyond the therapy session itself. This has implications for existing therapies and creates potential for entirely new research and intervention applications. Technological methods of data collection illustrate this point. For decades, research on suicide and self-harm focused on group-level differences. Clinical psychologists were understandably interested in which demographic or clinical populations were at the greatest risk of engaging in harmful, potentially fatal behavior. Advances in technology now allow researchers to passively collect data in real time and even administer brief questionnaires frequently. Consequently, researchers today can investigate dynamic intra-individual factors associated with increased risk of harming behavior for the individual person. Kleiman et al.'s (2017) study using ecological momentary assessment data showed that factors such as hopelessness and loneliness, which were previously known as interindividual risk factors, also correlated with suicidal ideation *within* clients over time. This kind of knowledge illuminates a path to the kinds of idiographic, real-time assessment that can inform truly personalized intervention of the kind so many of us dream of (e.g., Ng and Weisz 2016; Weisz et al. 2019).

Similarly, the potential for AI to inform and guide intervention in places where there are insufficient resources—even insufficient numbers of clinicians—may be a significant boon to psychotherapy practice. We do not see clinicians being replaced by robots anytime soon, but there are many ways in which AI can complement the work of clinicians by informing inexpensive, accessible, and flexible intervention for those in need of mental health services. There is substantial evidence showing that very brief (even single-session) and computer-guided intervention can produce marked benefit (Andrews et al. 2018; Schleider and Weisz 2017); what AI may add to the picture is an ability to fit each such intervention to the individuals most likely to benefit from it. Going beyond this simple notion, Fitzpatrick et al. (2017) used an automated conversational agent named Woebot to guide a small sample of college students experiencing depression or anxiety through a course of CBT, and they found significant depression reduction after a two-week intervention period. This proof of concept study certainly warrants replication, but the approach and the findings may foreshadow what a significant portion of psychotherapy research and practice could look like decades from now—particularly where access and cost barriers need to be addressed.

Concluding Comment

The excellent Bickman review describes much of the psychotherapy landscape that may be altered by AI. The prospects are exciting, but these are early days, and applications to psychotherapy will face challenges that go beyond those encountered in the worlds of medical care and pharmacotherapy. We look forward to seeing whether, and if so, how, psychotherapy research and practice are reshaped by the rich array of tools provided by artificial intelligence.

Compliance with Ethical Standards

Conflict of interest We have no known conflict of interest to disclose.

References

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., et al. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders, 55*, 70–78. https://doi.org/10.1016/j.janxdis.2018.01.001.

Chung, Y., Addington, J., Bearden, C. E., Cadenhead, K., Cornblatt, B., Mathalon, D. H., … Walker, E. (2018). Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. *JAMA Psychiatry, 75*(9), 960–968. https://doi.org/10.1001/jamapsychiatry.2018.1543.

DeRubeis, R. J. (2019). The history, current status, and possible future of precision mental health. *Behaviour Research and Therapy, 123*, 103506. https://doi.org/10.1016/j.brat.2019.103506.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendation. A demonstration. *PLoS ONE, 9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4*(2), e19. https://doi.org/10.2196/mental.7785.

Gunduz, N., & Fokoué, E. (2015). *Robust classification of high dimension low sample size data*. http://arxiv.org/abs/1501.00592.

Kleiman, E. M., Turner, B. J., Fedor, S., Beale, E. E., Huffman, J. C., & Nock, M. K. (2017). Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of Abnormal Psychology, 126*(6), 726–738. https://doi.org/10.1037/abn0000273.

Lorenzo-Luaces, L., DeRubeis, R., Van Straten, A., & Tiemens, B. (2017). A prognostic index (PI) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders, 213*, 78–85. https://doi.org/10.1016/j.jad.2017.02.010.

Ng, M. Y., & Weisz, J. R. (2016). Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry, 57*(3), 216–236. https://doi.org/10.1111/jcpp.12470.

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects: Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*, 107–115. https://doi.org/10.1016/j.jaac.2016.11.007.

van Bronswijk, S. C., DeRubeis, R. J., Lemmens, L. H., Peeters, F. P., Keefe, J. R., Cohen, Z. D., et al. (2019). Precision medicine for long-term depression outcomes using the personalized advantage index approach: Cognitive therapy or interpersonal psychotherapy? *Psychological Medicine*. https://doi.org/10.1017/S0033291719003192.

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., et al. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, ADHD, and conduct problems. *Perspectives on Psychological Science, 14*(2), 216–237. https://doi.org/10.1177/1745691618805436.