AA P&P Association for the Advancement of Philosophy and Psychiatry

Bulletin

Volume 27, Number 1 **2020**

From the Editor

Welcome to another lively discussion, this one led by Dr. Phoebe Friesen, who is tackling the issue of digital psychiatry. This is a very broad area that leads in many directions. Phoebe focuses in on the ability of digital technology to generate huge data sets that allow prediction and diagnosis, all with no awareness on the part of the subjects of the research. She questions both the utility and the ethics of such research. As you will see from the commentaries, her analysis has stimulated a lot of discussion.

As is now our routine, this issue of the Bulletin will be accompanied by a target article and call for commentaries for the next Bulletin issue. Our author is Dr. Louis Charland, and his piece is titled Consent and Capacity in the Age of the Opioid Epidemic: The Drug Dealer's Point of View.

Meanwhile, my own commentary on Phoebe Friesen's article.

Data and People

Phoebe Friesen has eloquently introduced us to a language game than includes data sets, machine learning, algorithms, digitalization, AI, and natural language processing. And she has used this cluster to examine a new type of research that rates high on diagnosis and outcome prediction, but with the scientific and ethical challenges posed by such research. In this brief commentary I wish to focus on one aspect, cluster, data, and to do this I will include another feature of the new research protocol, metrics. Working with data always involves measurement, and measurement in research always involves data.

Data are pieces or units of information, and such pieces may exist in many different formats. Whatever form they start as, the fact that they can be transformed into digital units allows for the study and manipulation of huge data sets. And the manipulation involves measurements. That is what happens with the kind of research Friesen is describing. When a suicide prediction model can predict with 70% to 85% accuracy who is at risk for suicide, the model is measuring the accuracy of the prediction. And finally, as Friesen points out, all this manipulation and measurement can be done by machines and algorithms.

Friesen contends that the research findings in prediction do not lead in an obvious way to patient care. It's not hard, however, to invoke examples of data meas urements where they do accomplish something meaningful and important . A

(continued on page 12)

President's Column

In the latter part of 2019 I started seeing colleagues using digital signatures on administrative paperwork and investigated setting up a digital signature for myself, but it seemed like it would take more work than it was worth. By late March 2020 we were in lockdown in response to COVID-19 and I found myself at home. As the associate dean in my college I had to sign anywhere from three to seven grade change forms a day nearly every day of the week. I printed them out, signed them, and emailed a scanned copy to the registrar's office. By early May I decided I had to implement the digital signature option and it took me maybe 10 minutes to set one up. I do not see myself signing administrative forms by hand ever again if I have a choice. Nor will I pull a credit card out of my wallet if a store takes Apple Pay or write a check to deposit money into my mother's bank account rather making a direct transfer using Zelle.

Technology forces itself into our lives and changes our worlds. We sometimes adopt it out of openness and curiosity, or because those we work with have adopted it, or we may be shamed into it, or in the case of COVID-19, forced into it by the situation. In this issue of the *AAPP Bulletin*, Phoebe Friesen and several of our colleagues raise important considerations about yet another potential technological development – informatics-assisted diagnosis and psychopathology under the name of digital psychiatry.

Rather than summarizing the many important concerns and reservations about digital psychiatry articulated in this issue, I let me briefly ponder a world that introduces new technology at a dizzying pace.

If we could bring someone from the late 19th century into our current world, what might they think of it and of us? Quite likely, the farther back through the centuries we reach, the harder it would be for our time traveling ancestor to adapt to and become part of society. Assuming our society continues to progress along the same trajectory it has been on since the Renaissance, the same would be true of us if we could be transported to late 22nd or 23rd century.

More proximately, we should think about adapting to the society that will exist 25 years into the future. Many of us will end up there one day no matter what.

I remember how 25 year ago one of the upper level administrators at my university took pride in not having an email account. During that same time, I had colleagues who insisted on computing an analysis of variance by hand rather than using SPSS or SAS. I thought they were ridiculous and never wanted to be 'that person" and still don't, but neither do I tweet or use TikToc. Last year I found myself trying to explain to my incredulous younger brother why I was still buying CDs rather than paying ten dollars a month for a streaming music service. It is stunning, but the amount of time it goes from being an early adopter to an

(Continued on page 12)

Digital Psychiatry: Promises and Perils

Phoebe Friesen, Ph.D.
phoebe.friesen@mcgill.ca

Hype is not new to the field of psychiatry. Following waves of enthusiasm related to neuroscience and genetic research, psychiatry is now looking towards digital technologies to help relieve suffering. Machine learning techniques, combined with enormous data sets, have opened up a new frontier, in which algorithms can predict and identify the diagnoses of patients with tremendous accuracy and efficiency. It may seem obvious that this development is worth celebrating. However, while there is no shortage of predictive success in this domain, it is not always clear whether and how these technologies are contributing to patient wellbeing. Furthermore, these techniques now allow for psychiatric research to take place outside of existing structures of research ethics governance. Predictions and diagnoses can now be made about unsuspecting users who are not aware that their data is being used for research purposes, and yet few protections are in place.

AI-driven techniques, including machine learning, natural language processing, and predictive analytics, are changing the nature of health research. These technologies, combined with enormous and widely available data sets, now allow for medical research to take place in new settings (in online forums, on mobile phones) using novel data sets (Twitter posts, Google searches) and to make predictions well in advance of medical events. This new frontier has made its way into all domains of medicine, but psychiatry, in particular, is diving in with both feet. A recent systematic review of research that utilized machine learning techniques to analyze online personal health data by Yin et al. found that mental health was the most common investigational target in their sample, appearing in 39 of 103 papers [1].

In many of these papers, the predictive results are impressive, if not incredible. For example, De Choudhury et al. analysed Twitter posts of 376 new mothers and developed a classifier that can predict, with 71% accuracy, which mothers' online behaviors will change significantly after birth, signalling a risk of post-partum depression [2]. In another example, using social media data from 547 users who had either posted publicly about a past suicide attempt or donated their data to OurDataHelps.org, Coppersmith et al. built a suicide prediction model that demonstrated a 70% to 85% true positive rate (depending on how false positives were weighted) [3]. Similarly, Corcoran et al. reported the use of linguistic analysis to predict the onset and severity of psychosis in a sample of high-risk youth. Using an automated machine learning speech classifier, the authors were able to identify which features of speech best predict the onset of psychosis and use the classifier to predict the onset of psychosis with 83% accuracy and distinguish between patients and healthy controls will 72% accuracy [4].

While such predictions are undoubtedly impressive, more than mere predictive power is often promised in these manuscripts. In their abstract, Corcoran et al. claim that "automated linguistic analysis can be a powerful tool for diagnosis and treatment across neuropsychiatry" and that these findings can help to "identify linguistic targets for remediation and preventive intervention" [4]. It is unclear precisely what is being suggested by the authors here in terms of preventative intervention though. The features picked out by the classifier as predictive of the onset of psychosis included the use of possessive pronouns as well as decreases and variance in semantic coherence. Are these the "linguistic targets" the authors refer to? It seems very unlikely that by teaching patients to correct these language patterns, by using more possessive pronouns (e.g., her, his, mine, our) for example, that they would be less likely to develop psychosis. Unfortunately, nothing more is said within the manuscript about how the research might contribute to the development of such preventative interventions.

This example points to a much larger issue that arises frequently within digital psychiatry: the gap between making a prediction and identifying causal pathways. Prediction is not causation, and in many cases, it does not get us any closer to understanding causation, which is what is likely to support the development of novel interventions [5]. Arguably, the gap between prediction and causation is even more significant in digital psychiatry than the one we saw in previous waves of psychiatric research enthusiasm involving brains and biomarkers; since the body is no longer involved, putative targets for intervention seem even more distant. Unfortunately, in the digital realm, it's hard to see how predictions based on speech patterns, social media use, and texting behaviors are going to translate into novel interventions.

Similarly, gaps related to resources loom large in predictive analytics related to mental health. After a recent paper reported the development of a model that could predict loneliness based on social media data, the study's lead author, Sharath Chandra Guntuku, was quoted as saying that this identification system, combined with early interventions, could have "long-lasting effects on public health" [6, 7]. The distance between identifying lonely Facebook users and offering effective interventions to those users is a significant one though. The 'epidemic of loneliness' has been under discussion for decades now and it's far from clear that we have the tools or resources to address it [8]. Pinpointing which Facebook users are most lonely may help us to understand who, and when, people are most lonely, and perhaps where to direct our scarce resources, but it is unlikely to solve the problem.

Much of this research takes place online and uses publicly available data sets, creating challenges for the systems of research ethics governance that were established in the mid-twentieth century, when health research looked very different. Importantly, these new technologies allow for the creation of 'emergent medical data' from non-medical data sets without the awareness of those whose data is being used [9]. While the users on Twitter that De Choudhury et al. collected data from had publicly announced their new motherhood, they were never told that their data was being collected for research purposes and that they were being assessed in relation to their risk of post-partum depression. Because the data the researchers were interested in, that which represented the social engagement, emotions, social networks, and linguistic styles of the new moth-

ers, were publicly available online, there was no need to ask for consent.

In their paper, De Choudhury et al. acknowledge that "people may be uncomfortable with others performing and sharing these predictions", while also pointing out that the authors "chose Twitter because it is public and provides a longitudinal record of the events, thoughts, and emotions experienced in daily life" [2]. The creation of medical data about unsuspecting and nonconsenting users is especially worrisome in relation to psychiatric research, where the medical data being created is stigmatized and can introduce significant risks. Despite the increases in risk that arise when psychiatric data is created about unsuspecting users, De Choudhury and colleagues were not required to submit their research protocol for ethics review. As the research was conducted by employees of Microsoft, federal regulations requiring ethics review, as well as legal protections for health data, do not apply [2, 10]. Even if the same project had been pursued within an academic setting, however, it would have likely been exempt from ethics review, because such 'publicly available' data does not qualify as 'human subjects' research [11].

While the predictive promise of computational psychiatry is real, it is worth keeping in mind that there is no golden road from prediction to treatment, and much of this research is taking place outside of current protections. So, along with the enthusiasm, we might want to keep on hand a little pile of salt.

References

1. Yin, Z., L.M. Sulieman, and B.A. Malin, A systematic literature review of machine learning in online personal health data. *Journal of the American Medical Informatics Association*, 2019. **26**(6): p. 561-576.

2. De Choudhury, M., S. Counts, and E. Horvitz. Predicting postpartum changes in emotion and behavior via social media. in *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013.

3. Coppersmith, G., et al., Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical informatics insights,* 2018. **10**: p. 1178222618792860-1178222618792860.

4. Corcoran, C.M., et al., Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry,* 2018. **17**(1): p. 67-75.

5. Kleinberg, J., et al., Prediction policy problems. *American Economic Review*, 2015. **105**(5): p. 491-95.

6. Guntuku, S.C., et al., Studying expressions of loneliness in individuals using twitter: an observational study. *BMJ Open*, 2019. **9** (11): p. e030355.

7. Tweets from Twitter users could predict loneliness, in *Penn Today*. 2019.

8. Killeen, C., Loneliness: an epidemic in modern society. *Journal of Advanced Nursing,* 1998. **28**(4): p. 762-770.

9. Marks, M., *Tech companies dangerous practice: Using AI to infer hidden health data*, in *STAT News*. 2019.

10. *Health Insurance Portability and Accountability Act (HIPAA)*, E.B.S.A. U.S. Dept. of Labor, Editor. 2004: Washington, D.C.

11. Metcalf, J. and K. Crawford, *Where are Human Subjects in Big Data Research? The Emerging Ethics Divide.* The Emerging Ethics Divide (May 14, 2016). Big Data and Society, Spring, 2016.

***

Another Peril: Machine Learning as a Comprehensive Theory

Paul Lieberman, M.D.
paul_lieberman@brown.edu

Dr. Friesen is right to warn us of the perils of over-hyping machine learning in psychiatry. We are apt to forget that, as she reminds us, correlation is not causation, prediction is not treatment, and data mining is not ethically innocuous.

To this list of important perils, I would like to suggest adding another: machine learning is not a comprehensive theory of mind. By this I mean that, while we are tempted to move from considering machine learning as a description of how we learn *some* things, to its being a description of how we learn *anything*, and, therefore, of what *knowing* is, or even (a mathematical description) of *what* we know, these moves we should not make.

It seems natural to apply associationist theories, however vaguely formulated, including machine learning, very widely. For example, in an introduction to machine learning Alpaydin reminds us in completely general terms that, "When we learn the best strategy in a certain situation that knowledge is stored in our brain, and when the situation arises again – when we recognize ('cognize' means to know) the situation – we recall the suitable strategy and act accordingly." (1) The scope of scientific applications for machine learning and other associationist approaches has been breathtaking – language acquisition and reading, vision and facial recognition, learning and problem-solving, memory and motor action (2) – suggesting that it may be universally appropriate. In psychiatry, as well, it has recently been proposed that *all* effective psychotherapy for all forms of psychopathology (that is, therapy which is 'transdiagnostic') can be explained in terms of cognitive-behavioral processes – which are essentially associationist. (3) And Wittgenstein appears to be addressing this temptation to generalize an associationist model of mind when he *opens* Philosophical Investigations with St. Augustine's version of associationist language learning and his own characterization of it: "These words, it seems to me, give us a particular picture of the essence of human language." (4)

But as appealing as the idea of a comprehensive machine learning or associationist model of mind may be, there are causes for concern, if not alarm, because such a model would pass by or leave out much of what we ordinarily think about when we think about our mental lives. For example, if thoughts were neural connections, then thinking would be a brain process and the content of thoughts in mind, their presentation if you like in conscious-

ness, would have no causally explanatory role (5), for example, providing reasons for belief or for action. Not only would machine learning be universal, we would be nothing more than learning machines. The model could also lead to solipsism or skepticism. If thoughts were neural connections and thinking were a brain process, why should I believe that my thoughts are *true* – since they are only associations in my mind, or weighted connections in my brain? And why should you believe something I tell you is true because I believe it (6), since you also know that my belief is nothing but neural connections?

But we don't have to accept these unwelcome conclusions because machine learning is not the comprehensive, universally applicable theory of mind that some might imagine.

Machine learning or other associationist approaches can be meaningful and empirically testable only within larger networks of beliefs and practices. For example, in order to determine what a machine has 'learned,' or to test and confirm that an associationist model is true, one must correlate inputs and outputs and also have knowledge of the structure and functioning (program) of the intervening machine or brain – all of which must be identified and measured which, in turn, occurs within a much broader set of processes and practices: reliable recognition, labeling and correlation with other observations and with numbers, by multiple observers, using specified protocols, calculations made and repeatable using procedures appropriate to the task. These further rely upon the capacities of people, suitably trained, to reach agreement which may depend, in turn, on their experience, selection, education, motivation, temperament, intellectual endowment, physical health, economic security, and so on.

Some of these beliefs and practices are highly specialized and require extensive formal education (for example being able to program or use statistical methods), while others are also parts of and necessary for ordinary, everyday life (such as being able to see and remember reliably, in agreement with the relevantly similar judgments of others). All are examples of what Wittgenstein famously referred to as our 'forms of life' which all share the properties of being holistic (terms and propositions having meaning only within webs or networks of beliefs and practices), normative (practices, including using language, requiring standards which we accept as their correct applications and implications), and externalist (contents of thoughts being identified and constituted by objects, events, properties and behaviors which are publicly observable).

From these considerations, it seems we should conclude that machine learning or other associationist models of mind are a specialized subset of a much more general and diverse set of formulations, and that it is this larger, more diverse universe of our multiple, shared forms of life which is in fact the more comprehensive model of mind which we searched for in machine learning or other associationist models, which turned out to be only specialized subsets.[1]

If this is correct, then a psychiatrist should accept that while machine learning or other associationist approaches may have very useful applications in her field – as machine learning may have for highlighting vulnerable groups and potential treatments and as CBT in its various forms has for ameliorating distress – these approaches are not comprehensive and other useful means of understanding, explaining and treating psychiatric conditions *may* exist (although whether they do exist is an empirical matter).

As obvious as this conclusion may be, it is often lost sight of in psychiatry which, as Dr. Friesen notes, is prone to hype and, we might say, a weakness for sectarianism. The predictive successes of machine learning, the treatment successes of cognitive-behavioral psychotherapies and a number of discoveries in neural science (for example, of processes which inhibit or potentiate neural activity at the cellular or network level) are collectively reinforcing of the notion that a fundamentally associationist model of mind may be a comprehensive one. Dr. Friesen's warning is that, rather than being comprehensive, associationist approaches, including machine learning, must be situated within a much broader universe of thought, talk and practice.

These divergent views about machine learning and other associationist approaches – one seeing them as fundamental and comprehensive, the other as specialized and partial – are apt to have divergent effects on the people, including psychiatrists, who hold them – as any models or pictures do. Like psychoanalysts and psychopharmacologists, cognitive-behavioral psychiatrists are also susceptible to self-imposed limits on the questions they ask and the answers they suggest. For example, modifying weightings, whether in fact or in the imagination of a cognitive-behavioral therapist treating someone who feels anxious or in a learning machine which models an anxious patient simply passes by a number of questions which can be asked. Is there something *about* this person (what his experiences are like or how he thinks) in this situation (of uncertainty or exposure) which brings on or contributes to anxiety? Why does his being here now cause a *symptom* (an experience inside that something is wrong) as opposed, for example, to a behavior, and might that 'inner-ness' be telling us something important about what anxiety, or any symptom *is* (a retreat or adaptation, for example)? Does it matter, or is it just coincidence, that anxiety phenomenologically resembles *fear* (and *anger*)? And why is it something that has *those* resemblances and not others (depression-connected to-grief, for example) that this person experiences here and now? Why do we feel that anxiety normatively, usually, has a *psychological* explanation (seems to allow for 'a certain kind of 'why question"' (7)) when feeling pain or cold do not? Why is anxiety *embarrassing* or stigmatizing? We don't for the most part know if these questions have any cash out in clinical psychiatry but psychiatrists may reasonably wonder if they might. Ignoring them altogether in enthrallment to over-hyped machine learning and other associationist theories is a real peril for us.

Note

1. At this point, one might think that, if a machine learning or associationist model is true, it would apply to all our thoughts, talk and practices, so

the argument we're making seems flawed. But it should be recalled, first, that we can't learn or *know* that the application is universal without already knowing and participating in our ordinary forms of life and, second, that we can't even formulate or *mean* the thought (that all learning is machine-like or associationist) without a prior acceptance of our ordinary forms of life, since the terms in which the thought is formulated have meanings only within such webs of belief and practice.

References

1. Alpaydin E. *Machine Learning*. Cambridge, MA, The MIT Press, 2016, p. 18.

2. Posner, ML (editor). *Foundations of Cognitive Science.* Cambridge, MA, The MIT Press, 1989.

3. Caspi, A and Moffitt, TE. All for one and one for all: *mental disorders in one dimension. American Journal of Psychiatry* 175(9):831-844, 2018.

4. Wittgenstein, L. *Philosophical Investigations,* Third Edition, New York, The Macmillan Company, 1958, para.1.

5. Kim.J. *Supervenience and Mind*. Cambridge, Cambridge University Press, 1993.

6. Moran, T. Getting told and being believed. *Philosopher's Imprint* 5:1-29, 2005.

7. Anscombe, GEM. Intention. *Proceedings of the Aristotelian Society* 57:321-332, 1957.

***

Digital Selves and Bodied Realities in Psychiatry: More Perils?

Suze Berkhout, MD, PhD, FRCPC
suze.berkhout@uhn.ca

It was with great enthusiasm that I read Dr. Friesen's sharp and insightful piece on the promises and perils of digital psychiatry. The notion that a digitized psychiatry will modernize the field to respond to challenges in diagnosis, prognosis, and intervention holds sway across numerous domains of psychiatric medicine. As Friesen explains, much of this is directed toward the promises of both prediction *and* prevention—whether in relation to perinatal depression, suicidal behaviour, or psychosis. It is the latter field—and the area of first episode psychosis (FEP) in particular—that I want to take a moment to reflect on, as a way to extend the cogent argument that Dr. Friesen puts forward.

Friesen aptly points to a number of underlying issues within the claims of digital psychiatry: (1) the gap between prediction and causation (presumably an important one to address if intervention is the end goal); (2) the distance between conceptualizing a broad public issue and how one best directs limited resources to address it; (3) the ethical challenges in using "emergent medical data," i.e. clinical information generated from large, non-medical data sets for which consent is not required. Most (if not all) of these concerns likewise apply to digital technologies in FEP. That said, the scope of ethical concerns regarding digital technologies in psychiatry can be extended even further in the field of FEP.

FEP refers generally to an early point in time in the diagnosis and management of psychotic illnesses stemming from a range of potential causes (Breitborde et al. 2009) and as a clinical organizing concept, FEP is structured by what is termed the "early intervention" paradigm. This paradigm holds that the ability to prevent or reduce morbidity (in this case, from psychosis) is best accomplished through the provision of treatment early in the course of illness (Csillag et al. 2016, 540). And it is within FEP that excitement for digital technologies has grown in a dramatic fashion, related to the hopes of approximating cure vis a vis early intervention. Digital platforms have been lauded here for their role in self-management as well as their ability to generate biometric data and digitized "phenotypes" that may accomplish a much sought-after goal of understanding heterogeneity in clinical presentations. Much of this is based on a collection of research practices termed "ecological momentary analysis" (EMA) or "experience sampling method" (ESM), which can be likened to one another in light of their use of digital platforms (smartphone apps predominantly) to track behaviour and phenomenological experiences (Ben-Zeev et al. 2014; Bell et al. 2018; Firth and Torous 2015).

If we turn to considering the language and metaphors used to depict digital care tools in FEP, something interesting is revealed. There is a striking overlap between how digital tools are conceptualized, and the imagery that shapes FEP and the early intervention paradigm itself. Within early intervention in psychosis, hope is indexed through the temporality of interventions. The framing of biological treatments in FEP, for instance, is one of a medical "cure" (or near-cure)—intervention in FEP is frequently referred to in editorials, scientific articles, and infographics as a new era, a bridge to the future; early intervention as it relates to FEP additionally links hope to novelty, knowledge, and innovation (McGorry et al. 2015; Saraceno 2007). Similarly, digital care tools are a "new hope" (Torous et al. 2019); innovation and novelty are predominant themes. The temporality of the early intervention paradigm as well as of digital care tools attaches hope and optimism to the rational deployment of scientific facts. In both instances, we hear of a futurity that casts mental disability as an obstacle to the arc of progress (Berkhout 2018). This enlightenment-laden progress narrative is part of what Allison Kafer calls the curative imaginary: an understanding of disability that sees medical intervention as unquestionable; expected and assumed, any other way of living is unimaginable (Kafer 2013, 27-28). And, to Friesen's point about the attention of digital psychiatry as being focused on prediction as well as prevention, prognostication is likewise part of a futurity that demarcates disa-

bled bodyminds as unquestionably in need of biomedical intervention (Clare 2016). Digital technologies in FEP are meant to predict and control uncertainty so as to return individuals to (or approximate, as closely as possible) a pre-morbid sate, presumably with the help of antipsychotic medications that are also more easily managed through digital care tools—micro-political technologies of health at the individual and molecular level (Berkhout 2018). To say that such technologies give "voice" to lived experience as some proponents of digital care tools in FEP do (see Torous et al. 2019) misses a larger issue of health politics. When biodata is seen as somehow speaking for itself while serving as a proxy for the complexity of experiential knowledge, it needs to be acknowledged that this lived experience serves as knowledge *only* when translated through expert-designed tools (Swartz 2018).

What might we say if we hold crip and disability-informed insights regarding futurities alongside the notion of *co-production*? As a concept stemming from the field of science and technology studies, co-production refers to the ways in which evolving scientific concepts, technological artifacts, and associated beliefs may emerge hand-in-hand with representations, discourses, and identities (Jongsam et al. 2018). Within FEP, EMA/ESM apps built into smartphones and tablets can be said to be coeval with beliefs about what constitutes psychotic phenomena and with psychiatric service users themselves. To paraphrase Latour, digital care tools have never been modern: such tools do not simply carve up the world at its joints but materialize particular kinds of subjects. I see this very much akin to the concerns Dr. Friesen raises in their excellent piece and look forward to more discussion on this topic from within philosophy of psychiatry and psychiatric ethics.

References

Bell, Imogen H., et al. 2018. "Smartphone-based ecological momentary assessment and intervention in a blended coping-focused therapy for distressing voices: Development and case illustration." *Internet Interventions.* 14: 18-25.

Ben-Zeev, Dror, et al. 2014. "Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia." *Schizophrenia Bulletin*. 40(6): 1244-1253.

Berkhout, S. 2018. "Paradigm Shift? Purity, Progress, and the Origins of First Episode Psychosis." *Medical Humanities*. 44: 172-180.

Breitborde, N., Srihari, V.H., and Woods, S. 2009. "Review of the Operational Definition for First-Episode Psychosis/EIP." *Early Intervention in Psychiatry*. 3(4): 259-265.

Clare, E. 2017. *Brilliant Impfetion: Grappling with Cure.* Durham: Duke University Press.

Csillag, C., Nordentoft, M., Mizuno, M. et al. Early Intervention Services in Psychosis: From Evidence to Wide Implementation. *Early Intervention in Psychiatry* 2016; 10 (6): 540-546.

Firth, J. and Torous, J. 2015. "Smartphone apps for schizophrenia: A systematic review." *JMIR mhealth and uHealth.* 3(4): e102.

Kafer, A. 2013. *Feminist, Queer, Crip.* Bloomington: Indiana University Press.

Jongsma, K., Bredenoord, A.L, and Lucivero, F. 2018. "Digital Medicine: An Opportunity to Revisit the Role of Bioethicists." *American Journal of Bioethics*.18(9): 69-70.

McGorry, P. Early Intervention in Psychosis: Obvious, Effective, Overdue. *Journal of Nervous and Mental Disease* 2015; 203(5): 310-318.

Saraceno, B. New Knowledge and New Hope to People with Emerging Mental Disorders. *Early Intervention in Psychiatry* 2007; 1(1): 3-4.

Swartz, A. 2018. "Ethical Challenges of Digital Medicine for Serious Mental Illness." *American Journal of Bioethics*.18(9): 65-67.

Torous, J., Woodyatt, J., Keshavan, M., and Tully, L. 2019. "A New Hope for Early Psychosis Care: The Evolving Landscape of Digital Care Tools." *The British Journal of Psychiatry*. 214: 269-272.

* * *

Prediction Without Explanation in Digital Psychiatry

Robyn Bluhm, Ph.D.
rbluhm@msu.edu

Phoebe Friesen's paper raises a number of important questions about digital psychiatry research. I will focus here on the issue of the gap between production and identifying causal pathways or developing treatments. Friesen acknowledges that a number of studies have reported impressive success in predicting, for example, post-partum depression or the onset of psychosis. (It is worth noting, though, that a 50% success rate is expected by chance, and that the true test of a predictive algorithm comes when it is used on datasets other than the one used to generate it.) Yet even when specific features of the data can be shown to drive successful production, the predictive features cannot be situated with a clear causal pathway. As Friesen points out, knowing that possessive pronoun use is linked to the onset of psychosis does not suggest a means of intervening to prevent the episode. Thus, these statistical relationships fail to explain how a predictive features is related to a clinical outcome, or to suggest a potential avenue for developing a treatment. This does not, though, stop researchers from promising more than predictive power. I suggest that their optimism is unwarranted, not merely because the field of AI-driven psychiatry is fairly new, but because the relationships it discovers are almost never going to be the right kind for explanation.

The philosopher Heather Douglas has written about the relationship between prediction and explanation in science, arguing that the purpose of explanations is (or should b) to inform better predictions. Her claim takes on an additional urgency in medical research, where prediction is necessary not only for accurate prognosis, but also to ground good treatment decisions. Some of the studies Friesen discusses do show promise for prognosis; examples are the prediction of post-partum depression or of an episode of psychosis. But the kind of predictive/explanation relationships Douglas discusses can't be achieved by digital psychiatry. Detailed causal explanations of phenomena that can lead to new scientific insights are notoriously im-

possible in the kinds of AI-driven techniques used in digital psychiatry, where the algorithm driving prediction cannot be mapped onto known causal variables. Nor can what we know about causal processes be used to inform predictions in any meaningful sense: this requires the development of hypotheses that can be experimentally tested, or at least a model that works "forward" from potentially causal variables to outcome data. By contrast, in digital psychiatry, prediction works "backwards" from existing data to identify patterns within the data, with no guarantee that those patterns have any deeper causal, or more clinically useful, meaning. This means that, ultimately, the prospect for these techniques to contribute to the development of new therapies is low.

References

Douglas, HE (2009) Reintroducing prediction to explanation *Philosophy of Science* 76(4):444-463.

***

Digital healthcare can expand the mental health industrial complex

John Z Sadler M.D.
John.Sadler@utsouthwestern.edu

In his prescient 2002 science-fiction film, *Minority Report,* Steven Spielberg's digital future portrays a world where crime can be prevented before it happens, through pre-emptive intervention based upon technology. What he doesn't get quite right, is he depends upon a triad of gifted humans, 'precogs', who can foresee with complete accuracy criminal actions. The data for their input is supplied by networked big data which feeds the precogs constantly. What the ensuing 18 years of real-world digital development has shown is that the precogs are not needed, because AI potentially can do criminal, and other forms of surveillance, without superpowered beings. Only supercomputers are needed.

What Spielberg gets exactly right in the film is the following world-building component of a digital future. In a particularly revealing scene, the star Tom Cruise's character walks through a department store where commercial messages are fed directly to him, appearing as holographic images fading in and out to him throughout the store. The results of his unique big-data digital phenotyping gets its ultimate capitalist expression: selling stuff to him, of course!

In her *Bulletin* essay Phoebe Friesen raises more than a few very interesting questions, for which I can only response to a few, and briefly. Dr. Friesen wonders about the utility of AI profiling, or digital phenotyping (Insel, 2017) in terms of treating people with mental disorders. My point through mentioning *Minority Report* is to mention one implication: Assuming that the Western world maintains some substantive commitment to free- or minimally-constrained market capitalism, we can count on digital phenotyping to sell (presumably more effectively) products to people with mental disorders These products will most likely be medications or somatic therapies of the future. It's possible that services, such as psychotherapies, will also be 'sold' or promoted, though if the past and present are any indication, psychotherapy and other mental health services are obligations framed by clinician's conscientious interests and not very appealing to lobbyists in Washington, nor to politicians, whose campaign coffers are much more likely to be swollen by pharmaceutical and industry money (Sadler 2013). Will psychosocial treatments be further marginalized by digital psychiatry? The answers are not clear. We already see computerized and online psychotherapies, now aimed at enabling these services for patients in underserved areas.

A second, more explictly philosophical question implied by Dr. Friesen's essay, concerns the concept of 'health information', and what counts as protected, and protect-able health information. While my electrocardiogram, my MMPI, and my serum sodium are 'personal health information' (PHI), the information collected and even synthesized by digital phenotyping appears to not be protected health information of the 18 kinds mentioned in the HIPAA regulations, cited by Friesen. While my preference for Levi's jeans, arugula, and Archie Shepp's jazz may predict some health vulnerability or outcome for me in the digital health universe, this kind of information is far from health information as formulated in the HIPAA regulations We might consider the consequences of this. If currently innocuous, seemingly transparent, non-clinical information about us turns out to be parts of clinically-significant diagnostic or prognostic sets of information, this poses a substantive epistemic and practical challenge in protecting health privacy. Should we expand the kinds of information about us to include, well, *everything* counting as PHI? Besides being a practical challenge for enforcement (of which already there is little of, if HIPAA convictions are telling), I can't imagine the mental health industry will tolerate constraints upon the kinds of precise marketing information available that also happens to be the same information that is 'clinical' information. Further, the potential for my digital phenotype to identify me is real, if a currently-unrealized ideal sought by commercial interests illustrated in *Minority Report*. A looming difficulty already identified in the PHI front is that one's DNA profile is already capable of identifying me; meaning my personal identifiers are carried around with me, and through my skin cells, shedded everywhere (shades of another sci-fi film, *Gattaca*). So what we seem to be facing is the loss of the *feasibility* of health privacy. As two millenia of stigma has demonstrated, this doesn't seem to be a very attractive development for people with mental illnesses and disabilities.

References

Insel, T. R. (2017). Digital phenotyping: technology for a new science of behavior. *Jama*, *318*(13), 1215-1216.

Sadler, J. Z. (2013). Considering the economy of DSM alternatives. In *Making the DSM-5* (pp. 21-38) edited by J. Paris and J Phillips. Springer, New York, NY.

***

(Don't) Believe the Hype

Paolo Corsico, Søren Holm
paoo.corsico@postgrad.
manchester.ac.uk

What is the problem with digital psychiatry? In her article, Phoebe Friesen effectively highlights three key challenges posed by the development of digital psychiatric applications. One of these challenges is practical whilst the other two are ethical. First, there is the practical issue of research *governance*: if digital psychiatric research takes place outside of clinical contexts, how can we regulate—and who should oversee—such research so that participants are safeguarded according to established research ethics principles? Second, there is the ethical issue of *beneficence*: will digital applications be beneficial to psychiatric patients? Third, and closely linked to the second, is the issue of the *gap* between prediction and intervention in digital psychiatry: what do we do with increased predictive capabilities if we cannot provide (access to) timely intervention? This is not only a clinical problem but also an ethical one. It raises the moral question of how we *ought to act* upon our increased predictive capabilities in mental health care. These three challenges, Friesen argues, make digital psychiatry currently problematic.

We do not directly address these three challenges here. Rather, we believe that Friesen poses a more fundamental question that embeds and puts the others into context. We address this question here: Why the hype? How cautious should we be in believing that digital psychiatry will bring about the revolution it promises to bring about?

We might call this question the 'hype problem'. As Friesen argues, the hype problem posed by digital psychiatry has much in common with the hype problem posed in the last decades by genomics and neuroscience, which often promised to bring about a revolution in the way we understand, diagnose, and treat mental illness [1]. The extent to which we believe that genomics and neuroscience are (or may be) in a position to hold on to that promise defines the boundaries amongst different aetiological theories of mental illness. Yet, from neuroscience and genomics to machine learning, the hype appears to be similar. On a more general level, the current hype on machine learning extends far beyond psychiatry to a variety of medical domains. One thing must be noted. The hype is evident not only in medicine but also in ethics. It is evident in the growing field of Artificial Intelligence (AI) and digital ethics, to which we are contributing right now. Does ethics also suffer from the hype problem? Are we co-responsible for the hype on digital ethics, or are we justified in investigating the ethical implications of digital innovation? Udo Schuklenk has interestingly suggested that we should reflect on the 'ethics of AI ethics', and to avoid that only the voices of those who are willing to investigate the moral implications of AI whilst not being overly critical of the digital agenda are heard [2].

So, why the hype on digital psychiatry? We propose a hypothesis. The hype on digital psychiatry may be the result of not only technological innovation but also of how new technologies interact with longstanding issues in a given clinical speciality. Not only technological development but also the cultural background where this unfolds may be held accountable for the hype. Two major issues seem to characterise psychiatry:

1. *Diagnostic uncertainty*, meaning that (i) a lot of patients receive many different diagnoses over their lifetime, and that (ii) diagnostic categories have been heavily debated and contested and are subject to constant revision [3, 4].

2. The structural *lack of access* to effective mental health care [5]. This refers to both (i) access to mental health services, and (ii) the difficulty in developing novel and effective interventions for severe mental illness.

Our hypothesis is that these two issues may create the necessary 'room' for the hype on digital psychiatry. Digital applications seem to be currently used to address these two longstanding issues. In their scoping review, Shatte et al. show that most machine-learning applications are being developed in mental health for the detection and diagnosis of mental health conditions [6]. Digital phenotyping—the collection and analysis of cognitive and behavioural data via digital technologies—holds promise to transform mental health care because it uses widespread low-cost technologies such as smartphones. As Martinez-Martin et al. claim, "Because digital phenotyping uses a ubiquitous technology and is inexpensive to deploy, it will likely transform the diagnosis and treatment of mental illness globally by enabling passive, continuous, quantitative, and ecological measurement-based care" [7]. Digital psychiatry promises to fill the two gaps identified above. Hence, the hype.

If our hypothesis is true, then only by addressing issues (1) and (2) in the first place may we properly understand the hype on digital psychiatry and evaluate the potential role of digital innovation in improving people's mental health. The important point seems to be, do we have *good reasons* to believe the hype? Do we have good reasons to think that digital innovation will really help us to address those two issues? If we do *not* have good reasons to believe the hype then we should probably be *wary* of the claimed disruptive potential of digital technology in psychiatry. Conversely, if we have good reasons to believe the hype then it may be worth to sketch out the practical and ethical issues that arise in digital psychiatry, and to strive to regulate emerging practices. Figuring out whether we have, and what these good reasons may be can stem from a thorough discussion amongst the clinical community, ethico-legal scholars, and patients. After all, the hype problem of digital psychiatry is situated at the interface between (promising) research results and appropriate clinical translation. In describing their results, researchers are often prompted to consider—and sometimes overstate—the *potential* clinical impact of their findings.

Assessing the hype of digital psychiatry and the ethical issues thereof hence implies that we do two things. Firstly, we should carefully evaluate the *real* potential of digital applications to ameliorate diagnostic and treatment practices in psychiatry. We should investigate how digital applications might effectively reduce diagnostic uncertainty and improve access to

(quality) mental health care. By doing so, we might be able to start distinguishing what is only hype from what may indeed benefit patients. Secondly—not chronologically but theoretically—we should ensure the ethical conduct of digital psychiatric research within and outside of clinical settings, and ensure the appropriate (ethical) clinical translation of research findings in mental health care. In doing so, we may end up contributing to the hype problem in digital ethics by incidentally highlighting its promise, but that is what bioethicists do; they investigate the ethical implications of technological innovation in medicine. Indeed, they might help to provide that little pile of salt that is needed along with technological enthusiasm.

References

1.Insel, T.R., Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Archives of General Psychiatry*, 2009. **66**(2): p. 128-33.

2.Schuklenk, U., On the ethics of AI ethics. *Bioethics,* 2020. **34**(2): p. 146-147.

3. Read, J. and J. Dillon, *Models of madness. Psychological, social and biological approaches to psychosis*. 2nd ed. 2013, London & New York: Routledge.

4. Guloksuz, S. and J. van Os, The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. *Psychological Medicine*, 2018. **48**(2): p. 229-244.

5. NHS England and Department of Health, *Achieving Better Access to Mental Health Services by 2020*. 2014.

6. Shatte, A.B.R., D.M. Hutchinson, and S.J. Teague, Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine* .2019 ,**49**(9): p. 1426-1448.

7.Martinez-Martin, N., T.R. Insel, P. Dagum, H.T. Greely, and M.K. Cho, Data mining for health: staking out the ethical territory of digital phenotyping. npj *Digital Medicine*, 2018. **1**.68,

***

Analog Follies in the Age of Digital Psychiatry

Awais Aftab, M.D.
awaisaftab@gmail.com

Phoebe Friesen does a splendid job in outlining how the reality of computational psychiatry falls short of the hyped promise and how much of this research is happening without oversight from the traditional apparatus for protection of human subjects. I find myself in agreement with her, and I will use this opportunity to elaborate on some of my own concerns about the rise of computational methods in psychiatry.

I will do so in the setting of a partly hypothetical scenario in which methods of digital psychiatry have made it possible to analyze publicly available online data (as well as any personal data that users may be willing to share from their social media profiles) and make some psychiatric diagnoses with a high degree of accuracy.

1) Concerns about Diagnostic Reification

Dr. Friesen makes an excellent point that methods of digital psychiatry are not likely to be causally informative. It'll be valuable for us to look at this point in the context of Derek Bolton's discussions of biomarkers:

> …if and when a biomarker were found, the science of mechanisms, causes, treat- ment, prevention may well stay as it is. It depends on the extent and nature of the causal of the biomarker. At one of the spectrum, a biomarker may be just an (other) sign of the illness, *internal* (inside the skin) as opposed to external, but as yet hardly worth distinguishing from the *external* signs and symptoms of the illness, from the point of view of the etiological model, which, we may suppose, stays as highly complex and multifactorial as before. (1)

Digital markers of diagnostic entities will in all likelihood be surface phenomena and epiphenomena. The science of mechanisms, causes, treatment, prevention may well stay as it is even if successful digital markers are discovered. There is a mistaken belief in psychiatry that a biological or digital marker of some sort will somehow establish the disorder are more *real*, as less abstract. A digital marker may therefore lead to a false sense of validity of the construct and may further contribute to the already rampant reification of DSM diagnoses. It will be essential for clinicians to understand that digital markers will be *designed* through machine learning to detect the presence of a construct.

I am reminded of something Douglas Adams wrote:

> This is rather as if you imagine a puddle waking up one morning and thinking, 'This is an interesting world I find myself in — an interesting hole I find myself in — fits me rather neatly, doesn't it? In fact it fits me staggeringly well, must have been made to have me in it!' This is such a powerful idea that as the sun rises in the sky and the air heats up and as, gradually, the puddle gets smaller and smaller, frantically hanging on to the notion that everything's going to be alright, because this world was meant to have him in it, was built to have him in it; so the moment he disappears catches him rather by surprise." (2)

Like the puddle, we should not be caught by surprise. Our digital markers may fit psychiatric diagnoses *staggeringly well*, but that is not evidence that we are carving nature at its joints.

2) Concerns about Clinical Validity and Ethical Use

Until now, for the most part, diagnosis of a psychiatric disorder has relied on patients or families seeking help, thereby creating a certain threshold of "clinical significance". This is important because it means that the diagnosis is mostly, in some sense, *invited*. Either the patients or the patient's social system is experiencing distress/impairment/harm and is in need for help. That is why psychiatry and psychology are healing professions.

The ability to make a diagnosis based on publicly available online data such as social media use divorces

diagnosis from that *invitation*, from that threshold of clinical significance. We can legitimately wonder what meaning a diagnosis holds if it is not tied to clinical significance. This is an objection that has been raised with regards to epidemiological surveys (which evaluate presence of descriptive symptoms and typically produce inflated estimates of prevalence of psychiatric disorders). The same objection will also apply to digital psychiatry.

When digital psychiatry divorces diagnosis from the 'invitation' by consumers/clients who are in need of help, it creates the tools with which anyone can be diagnosed without their consent or their knowledge. The diagnosis is depersonalized and decontextualized.

Consider the implications of this using the highly controversial example of the Goldwater rule: with regards to public individuals, American Psychiatric Association maintains that "it is unethical for a psychiatrist to offer a professional opinion unless he or she has conducted an examination and has been granted proper authorization for such a statement." (3) This rule is often defended partly on an epistemological and partly on an ethical basis. The epistemological argument states that any diagnosis made in the absence of a personal examination is highly fallible. The epistemological argument is less convincing, since the professional lives of psychiatrists and psychologists are full of instances where we offer professional opinions in the absence of personal examination (when supervising trainees, doing chart reviews, doing psychological autopsies, etc.) The ethical argument states that even if it is possible to make a diagnosis without personal examination, a psychiatrist shouldn't do it because such an opinion is likely to be exploited for political purposes.

If all you need to make a diagnosis is a computer algorithm, however, digital psychiatry may possibly remove the psychiatrist from the equation entirely. Any newspaper may use such an algorithm to examine the publicly available tweets of a president and will have all the predictive power of digital psychiatry to offer a diagnosis to a hungry public!

Will this be the beginning of political psychiatric warfare? Or perhaps psychiatric diagnoses will become ubiquitous, a benign and common aspect of our existence? Facebook apps will inform us of our psychiatric morbidity profiles with as much nonchalance as they tell us which wizarding house at Hogwarts we belong to?

Digital psychiatry is unlikely to revolutionize understanding of etiology and treatment, but it may very well drastically alter society's relationship with psychiatric diagnoses. Like the ancient philosopher Thales, we may grow so absorbed in contemplating the digital heavens that we may stumble into the well at out feet.

References

1. Bolton D. Classification and causal mechanisms: a deflationary approach to the classification problem. In *Philosophical Issues in Psychiatry II: Nosology*, K. S. Kendler & J. Parnas (Eds). 2012. Oxford: Oxford University Press, pp. 10.

2. Adams D. *The Salmon of Doubt: Hitchhiking the Galaxy One Last Time.* Pan Macmillan. 2002.

3. American Psychiatric Association. Goldwater Rule's Origins Based on Long-Ago Controversy. https://www.psychiatry.org/newsroom/goldwater-rule (accessed 5/12/2020)

***

Response to Commentaries

Phoebe Friesen

Thank you for a number of fascinating and creative responses to my discussion of digital psychiatry, each of which has supplied me with an ample serving of food for thought.

While I sought to map out a few emerging issues with digital psychiatry within my commentary, I will say that I have a lot more to worry about now. The responses point to a number of possible harms, implications, and misuses, of digital psychiatric technologies. Both Jim Phillips and Paul Lieberman note the way in which large data sets can erase individual narratives. John Sadler alerts us to the potential for digital psychiatry, particularly in the context of capitalism, to lead to the further marginalization of psychosocial treatments. Suze Berkhout raises concerns about how digital tools can serve to reinforce the assumption that mental disability is an obstacle for which a cure is the solution to work towards. Furthermore, Paul Lieberman examines how the growth of digital psychiatry may lead us to adopt a narrow theory of the mind, causing us to view ourselves as "nothing more than learning machines" and to neglect complex and interesting research questions that do not present themselves in such a paradigm. Especially worrisome is the picture of "political psychiatric warfare" painted by Awais Aftab. In this (oh-so-) possible future, digital diagnostic tools are widely available for use on unconsenting users; journalists, who have no obligation to follow the Goldwater rule, can use these tools to feed "a hungry public" who desire daily psychiatric assessments following presidential tweets!

While it may not balance out these considerable worries, I enjoyed the tiny bit of optimism I found within the commentaries, voiced by Jim Phillips. Nodding to our current pandemic context, he aptly points out the many ways in which epidemiological data can be of use, and how we might also find such uses within digital psychiatry. He also reminds us that digital predictive technologies won't always have the last say. Just as psychiatrists (and patients, I would add) can interpret the conclusions of meta-analyses as relevant to an individual's care or not, the outputs of digital diagnostic algorithms will ideally be filtered through additional routes whereby false positives and false negatives can be caught, and individual experiences and values can be taken into account.

The particular worries I raised in my commentary, related to jumping from predictions to the promise of interventions, regulatory gaps that research takes place within, and the growing hype surrounding digital psychiatry, were also illuminated and expanded upon within several commentaries. Robyn Bluhm is entirely on point in her explanation of why the gap between predictions and intervention is so significant in digital psychiatry. Precisely because of the nature of these

technologies and the way they seek to exploit any predictive features that might be present, the statistical relationships discovered "are almost never going to be the right kind for explanation". As Bluhm puts it, drawing on the work of Heather Douglas, a good scientific explanation generates predictions by developing models and hypothesis that can then be tested empirically. However, in machine learning, "prediction works 'backwards' from existing data to identify patterns within the data, with no guarantee that those patterns have any deeper causal, or more clinically useful, meaning". I couldn't agree more. Although I fear that the enthusiasm surrounding machine learning may be inspiring others to move backwards rather than forwards [1].

John Sadler further advances my worries regarding how frequently digital psychiatry evades research oversight and health privacy protections. Noting that it is not merely health data that can be exploited today, but any large data sets that can be translated into health data, he notes that "what we seem to be facing is the loss of the *feasibility* of health privacy". Given this predicament, he asks whether we might expand health privacy laws to "include, well, *everything*"? While I share Sadler's concerns, I'm not ready to give up hope on demarcating what ought to fall within the boundaries of regulatory oversight just yet. While it seems likely that 'medical data' may no longer serve the function of providing such a boundary, it may be that we need to expand protection to include not only types of data, but also models and technologies that can create such sensitive data; through this route, perhaps we can still salvage some form of health data privacy.

Other responses engaged with the topic of hype which surrounds digital psychiatry. Suze Berkhout notes the resemblance and overlap between discussions of digital psychiatric tools and early intervention paradigms, particularly those which focus on first episode psychosis (FEP). In a fascinating characterization of the language and metaphors used within these two realms, she observes the way in which they both portray themselves as associated with hope, novelty, and innovation, and as frontiers which offer "a bridge to the future" that can help us enter "a new era". She also observes how these initiatives often align themselves with participatory approaches, claiming to incorporate the voices of those with lived experience into their tools and treatments. In doing so, she offers another kind of warning about the narratives produced by those working in this digital space. Beyond concerns about overpromising interventions as a result of predictive power, such initiatives may also make unfounded promises related to representation. As Berkhout points out, however, these "tools do not simply carve up the world at its joints but materialize particular kinds of subjects". I see this as an especially important warning, given the increasing pressure to democratize psychiatric research, and the concerns related to tokenism and cooptation that often arise within such initiatives [2, 3]

Paolo Corsico and Søren Holm also narrow in on the issue of hype, asking why such hype exists and how we ought to response to it. They insightfully identify two features of psychiatry that render it vulnerable to such digital hype: diagnostic uncertainty and the lack of access to effective treatments. They suggest further that we should "be *wary* of the claimed disruptive potential of digital technology in psychiatry" if we do not have good reasons to believe the hype. If we do have good reasons, however, they suggest that "it may be worth to sketch out the practical and ethical issues that arise in digital psychiatry, and to strive to regulate emerging practices". While I agree with these directives, I think we should embrace them both, regardless of whether we have good reasons to believe the hype or not. I would suggest we should always be wary of hype, particularly in a field like psychiatry, which has such a long history of hype followed by disappointment. Ethical and practical analyses, as well as regulatory reform, are perhaps even more necessary if unfounded hype is being generated, particularly as it corresponds with mountains of research dollars being spend on such digital developments.

In terms of Corsico and Holm's discussion of digital psychiatry filling in the gaps of diagnostic uncertainty and access to care in psychiatry, I can't say I'm feeling very optimistic on this front. As Awais Aftab points out, digital tools in psychiatry are "*designed* through machine learning to detect the presence of a construct", and so will merely affirm our existing systems of psychiatric classification, rather than generating alternatives; this means it's unlikely that the contested categories will become any less contested within the digital revolution. In terms of accessing care, predictive technologies certainly promise to identify more individuals at risk or in crisis. But, as noted above by Bluhm and myself, such predictions, which narrow in on finger strokes, the use of personal pronouns, and prosodic features of speech, are unlikely to generate effective treatments anytime soon. As is often the case, it seems, the proof will be in the pudding. The more important question in this case, however, may be who gets to decide whether the pudding constitutes proof. As John Sadler points out, in this realm, as in all parts of psychiatry, market forces will be shaping determinations of efficacy. As a result, along with Patyusha Kallari, I suggest we focus on how not only on how good or fair uses of AI tend to be, but how they shift, or fail to shift, distributions of power [4].

That said, if Corsico and Holm are right, by merely writing this commentary, I've already contributed to the hype. So I might not sleep very well tonight!

References

1. Yarkoni, T. and J. Westfall, Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 2017. **12**(6): p. 1100-1122.

2. Kalathil, J., Beyond Tokenism: Participation of mental health service users from racialised groups in mainstream user involvement initiatives. *Agenda*, 2010. **34**: p. 16-18.

3. Eriksson, E., Four features of cooptation: User involvement as sanctioned resistance. Nordisk välfärdsforskning| *Nordic Welfare Research*, 2018. **3**(1): p. 7-17.

4. Kalluri, P., Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 2020.

***

(President, continued from page 1)

emerging dinosaur might be around 25 years, or less.

We should not just passively accept that recent technology is the new best practice (it is not always better), nor try to deny it once it arrives. It would be better to actively make it ours and, in the process, both remake ourselves and mold the technology to our purposes and goals.

Each of us will always be who we are - children of a particular time and place - but part of who we are should include asking who we want to be. We should be grounded enough to not follow every fad or buy the hype about the next big product, but we should also appreciate and benefit from what human ingenuity and creativity produces. For example, I find this array of all-access max plus pay for television bewildering. As someone who was in elementary school in 1970s, I almost miss those days of only three television networks plus PBS, not but quite. What people are calling 'quality television' is better. Likewise, in the past month, several artists I follow have released new albums and I have listened to all of them via my streaming music service. It is better in many ways.

Now I have to decide what to do with all these CDs. To be honest, many of them are unopened because they came with free digital downloads into iTunes. It seems ridiculous to have been buying them given what the new technology offers. But I will keep them. Who knows? I should have held on to all my old vinyl records instead of giving them away in 1995 – today's teenagers might pay good money for the 'outmoded' analog technology. They say it's better.

***

(Editor, continued from page 1)

ready example is the current coronavirus crisis. We depend heavily on the gathering of data and its measurement to assess where we are in relation to the virus. Think of incidence of illness, incidence of death, positive and negative test results,

effects of social distance, benefits of masks, declining or rising rates of hospitalization, etc. – all involving the measurement of data.

On the other hand, in addition to the limitations suggested by Friesen, I wish to point out one big limitation of data analyses – that they leave out individual narratives. To appreciate this, consider how the PBS `i datum, we weep.` Data are what they sound like – impersonal, numerical, anonymous, shallow.

How might these thoughts apply to psychiatry? Let's begin with the notion of a symptom. Is a symptom a piece of data – a datum? If so, symptoms are among the t the gravity of the diagnosis in each case. As we can see, in this diagnostic process data and metrics are keeping close company.

The above represents the diagnostic manual approach to diagnosis. Scales are also used in the diagnostic process. The Beck, Hamilton, and PhQ-9 scales all represent ways of counting symptoms, with thresholds to meet for the diagnosis. For psychiatric treatment, on the other hand, we rely on the claims of Evidence Based Psychiatry (EBP) and the randomized controlled trials (RCT) that provide the 'evidence'. EBP and RCT have been soundly criticized, but that is not the point I want to develop here. I want rather to underline the fact that, along with symptoms, the findings of EBP and RCT form the data of psychiatry, and that as data they pose the challenges to data described above.

To illustrate these challenges I will focus on psychiatric treatment and the use of RCTs and meta-analyses in the choice of the best medication. In an RCT each subject in each group is a data point, in all important respects the same as other subjects in its group. We are not interest in their differences. Inasmuch as the meta-analysis is an analysis of RCTs, it includes huge numbers of data-point/subjects. Further, the decision as to the best answer is made semi-automatically and algorithmically. The challenge here is the same one raised by Friesen with her examples: does the meta-analysis lead to better psychiatric care? In treatment the subject is not a datum; it's the person in the office with me. And the treater is not an algorithm. Although in the hierarchy of EBP, the opinion of the treater is at the bottom of the hierarchy, way below the meta-analysis, the fact remains that it is this treater who decides whether the meta-analytic recommendation fits the needs of this patient. No piece of data, and no meta-analysis will answer that question.

JP

***

Bulletin Editor
James Phillips, M.D.
88 Noble Avenue
Milford, CT 06460
Phone (203) 877-0566
Fax (203) 877-1404
E-mail james.phillips@yale.edu

Philosophy, Psychiatry, & Psychology

K.W.M. Fulford, D.Phil., MRCPsych.
Founding Editor

John Z. Sadler, M.D.
Editor-in-Chief

AAPP Web Site
www3.utsouthwestern.edu/aapp

ASSOCIATION FOR THE ADVANCEMENT OF PHILOSOPHY & PSYCHIATRY (AAPP)
MEMBERSHIP APPLICATION

Membership in AAPP is open to all individuals interested in the subject of philosophy and psychiatry by election through the Membership Committee. The Association welcomes Student Members (enrollees in degree-granting programs in colleges and universities and physicians enrolled in approved psychiatric training programs and post-graduates in post-doctoral programs). In order to join AAPP please detach this form and mail to: Ms. Alta Anthony, Journal Subscriptions/ Memberships, The Johns Hopkins University Press, P.O. Box 19966, Baltimore, Maryland 21211.

Annual Dues: $115 Members; $42 Student Members (this includes a year's subscription to ***Philosophy, Psychiatry, &Psychology (PPP)***. Make checks payable to The Johns Hopkins University Press.

Name

Qualifications (clinical and/or philosophical)/Speciality/Interests

Address

Telephone
FAX

Amount Enclosed: Check:_____ VISA:_____
Exp.Date:_____

MasterCard:_____ Exp.Date:_____