

# **Clinical translation of computational brain models: Understanding the salience of trust in clinician-researcher relationships.**

Author list: Datta Burton, S.\* , Mahfoud, T.\* , Aicardi, C., Rose, N.

## **Dr Saheli Datta Burton**

Research Fellow,  
Department of Science, Technology, Engineering and Public Policy (Steapp)  
University College London, London, UK  
saheli.burton@ucl.ac.uk

## **Dr Tara Mahfoud**

Lecturer,  
Department of Sociology,  
University of Essex, Colchester, UK  
Tara.mahfoud@essex.ac.uk

## **Dr Christine Aicardi**

Senior Research Fellow,  
Department of Global Health and Social Medicine,  
King's College London, London, UK  
christine.aicardi@kcl.ac.uk

## **Professor Nikolas Rose**

Professor of Sociology,  
Department of Global Health and Social Medicine,  
King's College London, London, UK  
nikolas.rose@kcl.ac.uk

\* Saheli Datta Burton and Tara Mahfoud contributed equally to the work and are joint first authors, with Saheli Datta Burton as corresponding author.

## **Abstract**

Computational brain models use machine learning algorithms and statistical models to harness big data for delivering disease-specific diagnosis or prognosis for individuals. They are intended to support clinical decision making and are widely available. However, their translation into clinical practice remains weak despite efforts to improve implementation such as through training clinicians and clinical staff in their use and benefits. In this paper, we argue that it is necessary to go beyond existing implementation efforts to understand and meaningfully integrate the clinician's perspective and tacit knowledge for

translating computational brain models in neurological practice. The empirical research draws on our collective seven-year engagement with the Human Brain Project as researchers of its 'Ethics and Society' subproject and includes analysis of published and grey literature, participant observation at workshops and conferences, and interviews with data scientists, neuroscientists, and neurologists in the UK and Europe developing computational tools for neurology. Our findings show that building trust in the relationships between clinicians and researchers (modelers, data scientists) through meaningful upstream collaboration, greater model transparency and integration of tacit knowledge play a salient role in translation processes with meaningful benefit for patients.

## 1. Introduction

At the 2018 Annual Summit meeting of the Human Brain Project (HBP), the chair of the short-lived Clinical Advisory Board pointed out that,

when you go see your doctor, you must feel confident in their diagnostics, be reassured, trust them. The paradox is, currently doctors don't or can't, capitalise on the previous hundreds of patients they and their colleagues may have seen, who may have displayed similar symptoms.

At the same time, the chair urged researchers to remember that when developing even the most wonderful of medical tools to fix this problem with the use of big data analytics, it is mandatory, always, to develop "the human side of these tools". Where data-intensive algorithms for brain disorders are concerned, this 'human side' represents a triangle of trust relationships: between patients and researchers<sup>1</sup>, between patients and clinicians, and between clinicians and researchers. In this article, we explore questions of trust between patients and clinicians and between clinicians and researchers. As we shall see, these questions have implications for the use of diagnostic technologies based on computational modelling and machine learning techniques for the analysis of the large quantities of patient data gathered by hospitals, through clinical research and elsewhere.

---

<sup>1</sup> In this paper, 'researchers' refer to modelers and data scientists involved with the development of computational models and data analysis tools.

In the HBP, work is underway to analyse large clinical and research datasets, using algorithms and machine learning. The aim is to search for patterns in data that can individuate the neurobiological correlates of a disorder in ways that could be used to aid diagnosis, to target treatments and hence to improve prognosis. Computational models are one among many types of approaches currently under development to integrate Artificial Intelligence (AI), broadly construed, in healthcare. In turn, these computational models draw on the vast amount of personal health data now available to develop disease-specific machine learning algorithms and statistical models that are expected to assist clinical decision making when offering a 'personalised' diagnosis or prognosis for individuals.

Recent surveys and mappings of the ethical and social questions raised by the use of AI in healthcare<sup>2</sup> place overwhelming emphasis on issues of *public trust*, hinged on questions of bias and explainability of algorithms (Watson et al. 2019), as well as transparent, fair, secure and equitable use of health data (Future Advocacy 2018; Joshi and Morley 2019). Epistemic concerns with AI applications for healthcare remain mostly related to evidence as inconclusive, inscrutable or misguided (Morley et al. 2020; Wessler et al. 2017). Stake holders engaged with improving implementation (and thus uptake) of AI applications remain focused on training as a key tool for overcoming clinician's assumed distrust (Liberati et al. 2017) of and reluctance to learn novel techniques (Future Advocacy 2018; Joshi and Morley 2019; Nuffield Council on Bioethics 2018).

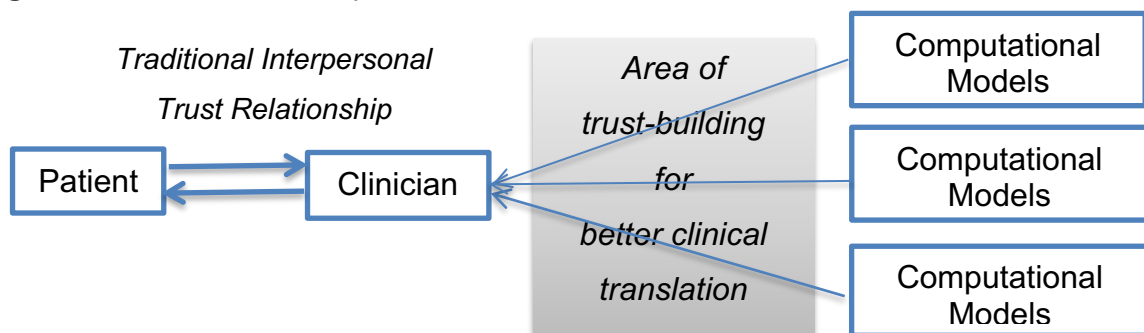
Yet, this focus on issues of public trust, evidence, and training misses and obscures the salient role of tacit knowledge in clinical diagnosis (Allegaert, Smits, and Johannes 2012) emblematic of the long, and ongoing, struggle of clinical practice to establish its epistemic value on firm and widely accepted grounds (Khushf 2013; Malterud 1995; Leblond 2013; Engel 2008). This struggle is inseparable from clinician and clinical staff's fears of job loss contingent on perceived threats to autonomy, authority and expertise, loosened relations with patients, deskilling, jobs displacement (Greenhalgh et al. 2017; Simonite 2016;

---

<sup>2</sup> Covering all areas of use from process optimisation to patient-facing applications and notably applications integrated in clinical pathways, which encompass among other techniques machine learning and data-driven computational models.

Rockoff 2016). We suggest understanding and integrating these entangled and contingent complexities of the clinician's perspective and the contribution of tacit knowledge towards building meaningful relationships of trust between clinicians and researchers as a first step in efforts to routinise computational brain models in neurological practice (see Figure 1). For patients (and thereby publics) need to trust clinicians as the custodians of their welfare, and by and large do so. Thus we argue that inadequate attention to the salience of clinician's trust in computational models risks weakening efforts to (re)gain public trust in AI-driven healthcare.

**Figure 1:** Trust Relationships in Clinical Translation of CPMs



We focus here not on trust from patients or publics, but on a different, but we suggest, an equally important dimension - the *clinician's trust*; a key factor in research-practice partnerships vital to the translation of machine learning and algorithm driven technologies. We reflect on the role of clinicians both as custodians of patient trust and welfare, and as end-users of computational models. In turn, this highlights the challenges and barriers to the responsible and successful adoption of such technologies; the clash between epistemic cultures and professional practices of data science and medicine; and the implications these have for how data is gathered and interpreted. Overcoming such challenges, we suggest, will depend on the trust of clinicians that their tacit, experiential, clinical knowledge is respected and integrated into data-driven technologies and that these technologies will meaningfully benefit patients.

We begin by briefly discussing some definitional aspects of trust in the context of the 'patient-clinician' relationship before considering the ways that clinician's trust in computational models in neurology shapes, and is shaped by, their traditional

interpersonal trust relationship with patients. This is followed by an analysis of the implications of clinician's relationship with researchers. This line of inquiry extends scholarship in healthcare and biomedicine around methodological concerns such as data anonymisation (Watson et al. 2019), consent (Larson 2013), platform standardisation (Shah, 2018) etc. At the same time, our inquiry provides a unique qualitative understanding of the ways in which trust relationships shape (or weaken) patient trust (Hall et al 2002; Fiscella et al. 2004; Thom 2001; Thorpe et al 2020; Klein et al. 2016). The empirical research for this article is grounded in our collective seven-year engagement with HBP as researchers of its 'Ethics and Society' subproject and includes analysis of published and grey literature, participant observation at workshops and conferences, and interviews with data scientists, neuroscientists, and neurologists in the UK and Europe developing computational tools for neurology.

## **2. Background**

### **2.1 Traditional clinician-patient trust relationship**

Trust in science, medicine and experts in general has been the subject of much academic and popular debate since at least the 1960s, although a general definition of trust remains elusive and contested (see exhaustive discussion in Mcknight and Chervany 1996). In contrast, specific 'trust' relationships like the patient-clinician trust relationship have been widely studied (Hall et al 2002; Fiscella et al. 2004; Thom 2001; Thorpe et al 2020; Klein et al. 2016). However, most studies have used large-scale survey data to identify objective measurable conditions that erode or enhance patient trust such as "perceived clinician financial conflicts of interest" has been shown to erode trust (Klein et al. 2016). While patient beliefs of clinician's honesty and competence to "act in their [the patient's] best interest, and preserve their confidentiality" (Fiscella et al. 2004), clinician's efforts to understand patient experiences and share power (Thom 2001) etc. have been found to enhance patient trust in clinicians.

Yet, whether measuring the erosion or enhancement of trust, these studies - perhaps more than anything else - implicitly emphasise the enduring nature of the interpersonal relationship of trust that patients have with clinicians as custodians of their welfare. A key reason for this, according to 70 percent respondents of a

nationwide survey of 3014 US adults by Pew Research, was because clinicians were "the central resource for information or support [for patients, carers and family] during serious episodes" and at other times (Fox and Duggan, 2013). As to "why do individuals trust their doctors the most?", the answer (according to another large survey conducted by PricewaterhouseCoopers 2012) was "human relationships". Hence, when novel clinical technologies are introduced, patient's (and thereby publics) decisions to adopt or reject them are overwhelmingly informed and guided by advice from clinicians. This is because "[patients (and publics)] want to trust and connect with the people providing [them] the care. ...it's easier to trust a person than an organization ...[and clinicians] have the ability to form human relationships and connections with their patients, which ultimately leads to increased trust" (Kathryn Armstrong, senior producer of web communications at Lehigh Valley Health Network, USA in PricewaterhouseCoopers 2012, 17). Indeed, the salience of this trust relationship is highlighted by PricewaterhouseCoopers' (2012, 17) survey of patient behaviors (aimed at understanding the nuances of health technology adoption) which concluded on a cautionary note that "to establish trust and credibility with consumers ...healthcare companies need[ed] to reconsider their approach to these [clinician-patient] relationships."

Within this complex interpersonal patient-clinician trust relationship, influenced to some extent by patient's institutional trust of the hospital or clinic where the clinician is embedded and from where they receive health(care) services (Gray 1997), computational models hope to gain a foothold as a trusted member of the clinician's diagnostic toolkit alongside stethoscopes and blood pressure monitors. Next we turn to the case of neurology where these models are being developed to aid neurologists to understand the challenges of how computational tools can gain a foothold - in other words, gain the clinician's trust.

## 2.2 Neuro-Diagnosis

Neurology has a long history of attempts to codify the diagnostic process but with limited success, especially around the interpretation of medical imaging, and formalising it in standard and computerised programmes (Doi 2007). Enormous research efforts over several decades, involving genetic, scanning and other

advanced neurotechnologies attempted to identify neurological 'biomarkers' for any of the current diagnostic categories used in clinical practice, for example those embodied in the successive editions of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (Rose 2013). However, with the exception of some forms of dementia, these attempts failed to identify clinically useful neurobiological or genetic biomarkers for diagnostic precision or treatment choice in the area of mental health. This led many psychiatric researchers, notably at the US National Institutes of Mental Health, to argue for a shift away from research based on diagnostic categories towards developing new approaches that would diagnose disorders on the basis of their biology.

The hope was that novel techniques using machine learning when used in the analysis of large data sets (containing information from genetic tests, brain scans and other physiological markers, together with data on clinical presentation, symptomatology, treatment and prognosis) might reveal previously hidden relations between neurobiology, symptomatology and treatment success. Computational models are emblematic of such attempts to formalise and standardise the diagnostic process. By codifying the range of data now available for neurodiagnosis from various sources (fMRI, MRI, PET, CT, EEG, MEG), models aim to develop 'objective' diagnoses, rather than person-driven analyses dependent on the interpretive skills of different clinicians. Currently, such approaches are being tried across a number of areas, to analyse case records, clinical and physiological data, test results and images from scans and link these to diagnosis and prognosis – in some cases producing results that are more accurate and reliable than those of even the most skilled diagnostician (Nuffield Council on Bioethics 2018). Not surprisingly, computational brain models have become an area of considerable commercial investment, bolstered to some extent - as in the United States - by the hope of automating clinical decisionmaking to reduce diagnostic errors held responsible for the rising cost of medical malpractice settlements<sup>3</sup>.

---

<sup>3</sup> see analysis by researchers at Johns Hopkins  
[https://www.hopkinsmedicine.org/news/media/releases/diagnostic\\_errors\\_more\\_common\\_costly\\_and\\_harmful\\_than\\_treatment\\_mistakes](https://www.hopkinsmedicine.org/news/media/releases/diagnostic_errors_more_common_costly_and_harmful_than_treatment_mistakes)

### 2.3 Tacit Knowledge

For many scholars, the challenge lies in trusting the 'reading' of data produced by these computational models. For instance, many have recognized that images (medical imaging data) do not speak for themselves, are not mere extensions of the naked eye and can only assist, not replace, expert opinion (Stone et al 2016). As clinicians such as Barry F Saunders (2008) demonstrated, the 'craft' practices involved in "learning to read" radiological images such as CT scans encapsulate complex institutional and hierarchical context within which doctors are trained to develop 'tacit knowledge' (see detailed discussion in Mahfoud, 2014). Anthropologist Andreas Roepstorff (2007) called this "skilled vision" - or ways of knowing that cannot easily be described, explained or put into words, but which enable us to do what we do (Polanyi 2009) – in this case to draw conclusions from visual and other evidence to make a diagnosis that will lead to a decision about action.

Unlike, earlier neurologists such as Kurt Goldstein (1878-1965) who had access only to the brains of deceased patients, and thus diagnosed living patients using case records and visual technologies such as films (Goldstein (1995[1939]) neurologists today have access to the living brain via a range of imaging technologies for diagnostic purposes - CT, PET, MRI, fMRI, EEG, etc. (Rose and Gainty, 2019). These images are generated by sophisticated technology and statistical procedures whose details are often unknown to those who make use of them. For instance, magnetic resonance imaging (MRI) uses magnetic field gradients that act upon certain atoms to generate detectable radio waves that are then processed using sophisticated algorithms to generate data on the distribution of water and fat in the body and then further processed to generate images of organs. However, these images embody many assumptions. For example, the hypothesis built into fMRI is that an increase in blood flow is a marker of an increase in brain activity. This assumption is embedded in the software embedded in the fMRI scanner that produces the images and despite longstanding critical evaluation, it is seldom questioned in the practices that utilise fMRI for research or diagnosis (Logothetis 2008). Indeed the standard fMRI paradigms have become controversial with the increasing recognition that measures of changes in blood oxygenation levels neglect the key role of highly distributed neural activity



– known as the ‘resting state’ - that is necessary for task performance, but is usually ‘subtracted’ from fMRI outputs as a result of the algorithms that are used to create the images (Gusnard and Raichle 2001).

These emerging data-driven technologies are thus far more than an aid to vision; they render some things visible at the expense of others, and frequently do so in ways that are ‘black boxed’ and not known or fully understood by those who use the results (Caruana et al 2020; Holzinger et al 2017). These issues are further complicated when large volumes of data produced by these technologies in many different clinics and research projects, using different research protocols, are linked using statistical devices to make them commensurable, and then analysed using machine learning to generate algorithms that are not known, let alone fully understood by potential users (the clinicians) who would have to decide whether and when to make use of the information provided in order to make a medical diagnosis.

Thus, many argue that as in previous diagnostic practices, ‘reading’ these images require the intervention of the trained eye of the expert (Mahfoud 2014), specifically trained in tacit practices typically taught outside of formalised education via apprenticeships (revealed in interviews with neurologists; see also Shah et al. 2018). As a result, the interpretive skills developed by neuroscientists (e.g. to interpret fMRI brain scans) involve both formal and informal education - one is required to know in order to see (Roepstorff 2007). In the clinical practice of neurology, ‘knowing’ (in order to read images) crucially draws on clinician’s tacit knowledge of individual patient’s physiologies and pathologies gained through interpersonal clinician-patient trust relationships. Indeed, the foundational salience of tacit knowledge in psychiatry is revealed by a recent survey of 791 psychiatrists across 22 countries (representing North and South America, Europe and Asia-Pacific) where a mere 3.8 percent “felt it was likely that future technology would make their jobs obsolete and only 17 % felt that future AI/ML was likely to replace a human clinician for providing empathetic care” (Doraiswamy, Blease and Bodner 2020) (see also Miner et al 2019). As Ferdinand Velasco, Texas Health’s chief medical information officer (op. cit 27) emphasises,

There is a lot of patient data—clinical and soon genomics as well. But what is really happening in our patients’ lives is missing to us and their

record— what’s happening in their lives is happening in the social space. ...If we understand the life factors that impact when and who they select for care and what challenges they face after receiving care, there is a lot of potential for merging analytics with the clinical side and improving care. The lack of access to various facets of human suffering that lies underneath the bare data is particularly acute for organisations and institutions engaged in translational research where researchers traditionally have no direct interaction with patients except during clinical trials (PricewaterhouseCoopers 2012, 17).

Thus, many argue that engagement is crucial between researchers and clinicians early on in technology development to bridge the widening “chasm in the understanding of end-users [clinicians] between [researcher's] imagination of a future user and users’ lived experiences” (Datta 2018, 354) (see also Epstein 1996; Smith et al. 2017). Without this early engagement, the expectation that once a technology is developed it will be readily adopted by clinicians (and patients) - at the researchers’ word of its power to improve their lives - is challenging. While engagement efforts after a technology is developed is widely considered a meaningless tick-box approach that is “largely ineffective in rebuilding public trust” (Wynne 2006, 217). Thus, from the clinician’s perspective, the computations involved in producing these images in order to ‘read’ and analyse brain scans hold little practical interest. However, as we shall see, clinicians do need to have confidence and trust in the process that has led from the initial data to the images that they have to interpret and utilise in clinical practice.

#### 2.4 Computational Brain Models

In contrast, from the perspective of the researchers seeking to interpret the range of information potentially available for diagnosis, machine learning promises to solve the problem of the masses of data from different sources now available - fMRI, MRI, PET, CT, EEG, MEG... - that are extremely difficult for human beings to integrate and analyse together. The hope is that machine learning tools can be used to analyse these large amounts of data from different sources, and further, that these computer-driven methods will be more objective than person-driven analyses, because they do not depend on the interpretive skills of different clinicians. Thus many argue that the digitization of data from patients’ clinical records and medical images combined with advanced data analytics can enable AI technologies such as machine learning and machine vision to distinguish between different potential diagnoses in a clinically meaningful way that can

enable clinicians to target specific treatments (Luo, et al. 2016). Data-driven technologies aimed at assisting healthcare practitioners in diagnostic processes have been approved by the US Food and Drug Administration (FDA) in recent years (Future Advocacy 2018). For example, the smartphone application 'Viz.AI' analyses CT images of the brains of patients admitted to hospitals with symptoms of stroke, identifies vessel blockages through these images, and sends this analysis via text to neurovascular specialists. This software was approved by the FDA as a "clinical decision support software" based on evidence submitted by the developers which demonstrated through a clinical trial that the software application more quickly identified the vessel blockages.<sup>4</sup> Development of several other software applications are underway (although not yet approved by the FDA), such as the collaboration between DeepMind Health and Moorfields Eye Hospital, London which uses neural networks to diagnose Age-related Macular Degeneration (AMD) through the analysis of Optical Coherence Tomography (OCT) (De Fauw et al. 2018).

In psychiatry, much of the research still begins from or utilises diagnoses according to current diagnostic categories despite the problems that many have identified with such classification (discussed earlier). Thus a collaboration between IBM and the University of Alberta uses neural networks to diagnose schizophrenia through an analysis of fMRI scans while patients undertake an audio-based exercise. The researchers claim to have identified "combinations of statistical features extracted from the data that can serve as reliable statistical (bio)markers of the disease, capable of accurately discriminating between schizophrenic patients and controls" (Gheiratmand 2017). These 'bio-markers' included an "abnormal" increase of connectivity between the thalamus and the primary motor/primary sensor cortex as well as "hyperconnectivity" in the fronto-parietal network. While the model was relatively successful at predicting a clinical diagnosis of schizophrenia (at above 70%), this research has not yet undergone the clinical trials needed for software regulation and approval. Further, it is being undertaken at a time when the very categories such as schizophrenia are contested by many experts in the field (Murray 2017). There is a widespread recognition that, across the whole spectrum of mental disorders, and particularly

---

<sup>4</sup> <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm>

in relation to psychoses, similar symptomatology may result from very different neurobiological pathways. There is thus a certain unhelpful circularity in seeking brain based biomarkers that correlate with symptom based diagnoses that are themselves contested and considered to lump together a variety of conditions that are developmentally, neurobiologically and prognostically distinct (Nature Biotechnology 2012).

Some six years ago, an Editorial in *Nature Biotechnology* (2012) entitled “What happened to personalized medicine?” reflected on the slow progress and unrealized hopes of those who predicted a revolution in medical diagnosis and treatment targeting based on biomarkers. The barriers they identified were less biological than social: a need “to broaden the concept of personalized medicine from the genetically reductionist version to one that includes other types of markers”; a need for more long term studies “linking specimens, sequence and other biomarker information to clinical outcomes”, a need for patients to be encouraged to share their data for research purposes, and a need to educate physicians “about the new diagnostics and how to integrate them with existing clinical information” which will require not only better education but also “the development of robust point-of-care devices and data-sharing technology and the establishment of trusted sources (e.g., medical association position statements on tests or the National Institutes of Health’s genetic testing registry). What perhaps stands out most in the editorial, is its emphasis on the salience of trust. For if clinicians and patients do not have legitimate trust in the accuracy, validity and utility of biomarkers, and that certainly includes brain based biomarkers, whatever the hopes of those who develop them, they will not ‘translate’ into clinical practice.

### **3. Gaining the clinician's trust**

So far, the literature on the dynamics of gaining the *clinician's trust* in clinical translation mostly derive from research focussed on other issues such as studies of computational technology adoption rates among clinical staff (Garland, Plemmons, and Koontz 2006), or critical scholarship identifying barriers to effective research-practice partnerships for better clinical -translation or -data collection (Mittelstadt and Floridi 2016). In turn, scholarship on barriers to the

adoption of computational models draws on these views of resistance to change in research-practice relationships to implicate this lack of trust as one among several 'technical and methodological' issues such as calibration, risk-sensitivity, data quality (Shah et al. 2018) and so forth. However some research does acknowledge the need to contextualise these understandings within the imperatives of big-data processing (Mittelstadt and Floridi 2016, 2-3].

### 3.1 Upstream clinician collaboration

Our research shows that collaborations with clinicians in the upstream research conceptualisation phase of technology development was crucial for gaining the clinicians' trust, as it provided the time and space to forge an interpersonal researcher-clinician trust relationship necessary to create bridges across the researcher-clinician divide. Typically, the relationship between clinicians and computational neuroscientists is based on give and take - clinic(ians) send researchers anonymised patient data, researchers perform analyses using this data in computational models under development and send back the results of the analyses to clinic(ians). However, our interviews revealed that trust was a defining element in this relationship between clinicians and researchers, particularly around data bias, model transparency and different epistemological traditions in neurology, the neurosciences, and computer sciences.

Importantly, researcher-developers of neuro-diagnostic tools highlighted the significance of close collaborations between clinicians, or "domain experts", and data analysts, not just for the sharing of medical data but more importantly for defining the research questions pursued. Computer scientists stated that research questions should be defined by clinicians, and only then can machine learning tools or methods be selected and decisions made about which types of data are needed. Similarly, there was acknowledgement among modellers that "domain-specific" training would be desirable to collaborate with clinicians, beyond the generalist traditions of computer science that require computational tools be developed for general-purpose and then adapted to specific use-cases. In reality, while some modellers do develop expertise in specific fields (such as neuroscience, or oncology), it is more common for computer scientists and engineers to move between biological domains of expertise.

Clinicians and biologists, on the other hand, found it concerning that modellers did not understand the biology and physiology of the conditions being explored, which they deemed necessary to develop clinically useful models. At the same time, researchers found it concerning that clinicians did not understand the modelling frameworks used - specifically the assumptions of the statistical, machine learning and other data analysis methods. Computational modellers suggested that future educational programmes for clinicians needs to include training in computational methods in order to adapt to changing clinical contexts where machine learning and other computational tools are likely to become more common place.

### 3.2 Greater model transparency

For clinicians, model transparency played a considerable role in deciding the extent of their trust in using a machine-learning based diagnostic tool. Some researchers argue for shorter-term technical solutions such as making machine learning tools more interpretable for clinicians by building in the ability for clinicians to trace back the way an algorithm has come to a certain conclusion, thus rendering the decision-making 'transparent'. Indeed, transparency is one of the key ethical principles in discussions around accountability in artificial intelligence (Mittelstadt and Floridi 2016). An association of researchers in Microsoft, Google, and others have, for example, proposed the principles of "Fairness, Accountability, and Transparency in Machine Learning" to address the "potentially discriminatory impact of machine learning" as well as the "dangers of inadvertently encoding bias into automated decisions".<sup>5</sup> Interviews with computational modellers suggested that to adhere to the principle of transparency, the use of supervised and semi-supervised learning algorithms was preferable to relying on unsupervised learning algorithms. This is because supervised and semi-supervised classification algorithms can be represented as decision-trees, which are more interpretable to collaborating clinicians than the 'black box' through which unsupervised learning algorithms produce their results.

---

<sup>5</sup> <http://www.fatml.org/>

Transparency and open science are core values of responsibility in scientific research and innovation (Von Schomberg 2013). These issues have become prominent in debates over algorithmics, with the demand for explainability and accountability of inscrutable systems getting stronger. This demand is gaining traction as it is increasingly believed to be a crucial, perhaps inescapable, step towards safety and trustworthiness of AI and machine learning systems, and it is seen as integral to ethically aligned design principles. Unsupervised learning algorithms, which aim is to discover inherent structures in data without using pre-existing categories, are under special examination for being inscrutable even to their designers. In this context, any proposals to use such unsupervised machine learning to discover ‘brain signatures’ that could bring about a complete revision of the classification of mental disorders appears highly problematic.

### 3.3 Integrating tacit knowledge

Even if algorithms can become less opaque, there remain other epistemological obstacles – that is to say, clashes of epistemologies - to collaborations between clinicians and researchers. Clinicians we interviewed talked about the importance of clinician-patient interaction for diagnosis. For example, in the diagnosis and treatment of epilepsy, clinicians carry the patients through from diagnosis to pre-surgical screening, surgery, and post-surgical rehabilitation. This is seen by some as already a highly personalised treatment since each patient is unique in terms of the symptoms exhibited and surgical treatment needed. Such ‘holistic’ treatment of the individual - common in neurology - is seen by computational neuroscientists as dependent on “subjective” and “biased” elements that need to be removed or reduced from clinical settings described as “low validity environments” with some citing psychologist Daniel Kahneman’s (2013) work on decision-making: “to maximize predictive accuracy... decisions should be left to algorithms in low validity environments” for support.

Despite such criticisms, ‘tacit knowledge’ of clinicians remains crucial in diagnostic processes. For example, it is clinicians who must use their formal and informal craft skills to interpret the various data sets from the patients to decide which part of the brain to remove during surgery or where to place SEEG (stereo electro encephalography) electrodes for an epileptic patient. In response to a

question about the receptivity among clinicians of computational brain modelling approaches for the diagnosis and treatment of neurological conditions, an engineer at a neuroscience laboratory in France said:

Clinicians do not explicitly state the idea behind why brain areas generate seizures, for example – it is an implicit model. I call it a model. It is what you learn when you study epileptology. Clinicians reason by saying that if this area affects this one, then if I remove this it will stop seizures. It is a model because it is a set of rules, but it is not entirely quantitative. There is a lot of experience needed to know, to have intuition. The goal of our quantitative model is to take into account the clinician's opinion but also other data to come up with results that bring those things together, hopefully in a robust fashion. The clinicians we work with are interested in having tools that will help them resolve the pathology for their patients in cases where they have no idea...They will try [the data analysis] out because they are curious, but will ignore the result. But as they get experience with tool, assuming it is good enough, it will become part of the workflow and part of their analysis.

For this researcher as well as others who we interviewed, such tacit or experiential knowledge is an implicit model as it involves a series of steps that result in a decision, or solution; they thus expressed concern that these steps were not made explicit. On the other hand, clinicians defended this 'intuition' as the result of years of diagnostic experience and surgical interventions that become internalised in their judgement processes. Clinicians insist that just because these were not fully articulated, or perhaps even capable of full articulation, does not mean they were not valid. While they did not use the term, it was clear here that clinicians considered that a 'skilled clinician' was one who had fully embodied the results of years of experience and immersion in a community of experts into the tacit knowledge that guided their decisions.

Not surprisingly, this epistemological conflict between the 'subjective' knowledge of the clinicians and the so-called 'objective' knowledge of the data analysis finds its way into discussions around data bias. Data scientists need what they call "good data distribution", standardised data, and comprehensive meta-data. As one computer scientist noted:

There are different cultures of how clinicians diagnose and treat – different clinicians with different expertise and knowledge can label patients differently. And many scores – like the Montreal Cognitive Assessment (MoCA) – are arbitrary, on a scale of one to five or



something. This kind of noise can be compensated for if you have large amounts of physicians involved who can make sure the data is distributed well and can check the quality of measurements, missing data.

On the one hand, this suggests that including clinician's tacit knowledge meaningfully in upstream technology development processes - rather than as part of a box-ticking exercise in post-development processes - is important for meaningful technology adaptation and its eventual adoption (Wong, Turner, and Yee 2008). This is consistent with Sullivan et al.'s (2005) conclusion in the domain of psychiatry research and practice, that a "bottom-up" approach in which services researchers assist frontline clinicians in testing interventions that clinicians themselves have devised ...[will] result in interventions that are more likely to be sustained over time." For instance, recent work on 'automatic speech recognition for psychotherapy' - whereby nuances of patient's speech in datasets used to develop a model vary significantly from patient populations it hopes to serve - highlight the acute need for clinician engagement in designing and guiding model development (Miner et al 2020) to reduce clinical bias in existing data and better reflect variability's in race, ethnicity, sex, gender, age etc. (Schwartz and Blankenship 2014).

On the other hand, modellers and clinicians alike acknowledge that clinicians do not have the required training to critically analyse the results of these new data analytics tools, and their distrust of these results may be caused by a lack of understanding of the models and the ways algorithms reach decisions. While unsupervised learning has been successful in diagnosis, especially in relation to the analysis of medical images (e.g. by Google DeepMind in De Fauw et al. 2018), these are far less interpretable than other machine learning methods. Yet, whether interpretable or not, clinicians we interviewed for this study suggested that if the computational models recommended a different diagnosis or different conclusions than their medical opinion, they were almost always likely to trust and pursue their own opinion.

For computational diagnostics tools aiming to become a part of the neurologist's toolkit, there is a need to move away from the negative 'quantitative' view that 'subjectivity' of clinicians' assessments are inherently 'biased' and acknowledge

the value of experiential and tacit knowledge in clinical reasoning that remains foundational to the relationships of trust between patients and clinicians. Incorporating Roepstorff's 'skilled vision' within algorithms (Roepstorff 2007) is the first step towards overcoming the difficulties of AI routinisation into practice for only those technologies that are perceived to add value to clinical reasoning rather than competing with it (like the new generation of AI-integrated or 'smart' computation diagnostics tools) are more likely to win clinician's trust.

Following this, the next step, as recent work in psychiatry suggests, is the need for careful research into the qualitative nuances (human impacts) of deploying "AI delivered, human [clinician]' supervised' psychotherapy" (Bhugra et al. 2017; Miner et al, 2019; Patel et al, 2019). A priority for such deployments of AI-enabled care must be to avoid situations where patient loss of trust in AI adversely reflects on patient trust in clinicians and in care provisioning (Miner et al, 2019) (see also Bhugra et al. 2017; Patel et al, 2019). This is especially likely if ever higher "expectations of benefit [from ever more sophisticated AI]" cause patients to "transition from feeling let down to feeling betrayed [by clinicians and by the systems of 'care']" (Miner et al 2019).

### 3.4. From clinician's trust to investor trust

For private investment to support the long process of clinical translation to the market, the technology end-user's (here the clinician) trust in and willingness to adopt the new technology is crucial (Greenhalgh et al. 2017). For instance, research shows that having clinicians in management positions with computational skills is positively correlated with "long-term commitment to the use of IT [in health care]" (Ingebrigtsen et al. 2014). The reverse - clinical leadership lacking a commitment to IT-enabled healthcare weakens *adoptability* (or end-user uptake) (Wong, Turner, and Yee. 2008) and ultimately investor confidence. Greenhalgh et al (2017) goes further to argue that when "the value proposition of the technology [is] unclear, in terms of a viable business venture for its developer [e.g. historical low CPM adoption rates] or in terms of a clear benefit for patients and an affordable real-world service model", the result is technology 'nonadoption and abandonment'. Our interviews with researchers reiterate this view that better

research-practice collaborations based on trust and transparency encourage sustained use of computational models and private investment. This is particularly crucial for the successful clinical translation of computational models developed by small and medium developers or public research institutions. For both need to build robust investor or public-funder's confidence in the commercial viability (end-user uptake) of their innovation to attract the substantial capital resources and regulatory expertise necessary to fund and drive the clinical evidence generation, evaluation, verification and validation processes needed to reach the market. At the same time, there is a need for stakeholders to acknowledge that widespread adoption of computational models should only be envisaged with particular attention to the risks of algorithmic bias - whereby AI applications normed on certain patient populations (e.g. whose data is easily available and widely used) may not be readily usable (without considerable adaptation, if at all) for diverse populations (see e.g. Miner et al 2020; Schwartz and Blankenship 2014).

## 4. Conclusion

In this paper, we have shown that the road from bench to bedside for computational brain models needs remains grounded in building better researcher-clinician trust relationships based on meaningful upstream collaboration that integrate model transparency and tacit knowledge. On the one hand, greater clinician trust contributes to the robustness of investor or public-funder's confidence in the viability (end-user uptake) of an innovation to reach and succeed in the market and increases the translation of computational brain models. On the other hand, the clinician, both as custodian of patient trust and welfare, and as the end-user of computational models, is uniquely placed to evaluate the patient benefit of allocating scarce time and capital resources to computational models instead of in other areas of patient care which may include low-tech investments such as employing more clinicians, updating aging equipment etc. Thus policies encouraging greater clinician engagement in deciding 'if a computational models is worth it?' will not only act as a check against faulty analytics but also lead to responsible adoption of computational models that do not merely encourage technology adoption for the sake of technological progress but for meaningful benefit to patient care.

## **Acknowledgements** (in alphabetical order)

We would like to convey our special thanks to Dr Edison Bicudo (University of Sussex, UK), Professor Alex Faulkner (University of Sussex, UK), and to all who participated in this research.

The authors acknowledge support from the European Union's Horizon 2020 Research and Innovation Programme funding for the *Human Brain Project* (Special Grant Agreement 2 number 785907).

## **Competing Interests Statement**

The authors report no conflicts of interest.

## **References**

- Allegaert, Karel, Smits, Anne and Johannes N. van den Anker. 2012. "Physiologically based pharmacokinetic modeling in pediatric drug development: a clinician's request for a more integrated approach." *Journal of Biomedicine and Biotechnology* 2012. doi.org/10.1155/2012/103763
- Bhugra, Dinesh, Tasman, Allan, Pathare, Soumitra, Priebe, Stefan, Smith, Shubulade, Torous, John, Melissa R. Arbuckle et al. 2017. "The WPA Lancet psychiatry commission on the future of psychiatry." *Lancet Psychiatry* 4(10):775–818. doi: 10.1016/S2215-0366(17)30333-4
- Caruana, Rich, Lundberg, Scott, Ribeiro, Marco Tulio, Nori, Harsha and Samuel Jenkins. 2020. "Intelligible and Explainable Machine Learning: Best Practices and Practical Challenges." In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (August): 3511-3512).
- Datta, Saheli. (2018). "Emerging dynamics of evidence and trust in online user-to-user engagement: the case of 'unproven' stem cell therapies'." *Critical Public Health* 28(3): 352-362. doi.org/10.1080/09581596.2018.1446509
- De Fauw, Jeffrey, Ledsam, Joseph R., Romera-Paredes, Bernardino, Nikolov, Stanislav, Tomasev, Nenad, Blackwell, Sam, Askham, Harry et al. 2018.

"Clinically applicable deep learning for diagnosis and referral in retinal disease." *Nature Medicine* **24** (9): 1342-1350. doi.org/10.1038/s41591-018-0107-6.

Doi, Kunio. 2007. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential." *Computerized Medical Imaging and Graphics* **31**(4): 198-211. doi.org/10.1016/j.compmedimag.2007.02.002

Doraiswamy, P. Murali, Blease, Charlotte and Kaylee Bodner. 2020. "Artificial intelligence and the future of psychiatry: Insights from a global physician survey." *Artificial Intelligence in Medicine* **102**: 101753.

Engel, Heiberg P.J. 2008. "Tacit knowledge and visual expertise in medical diagnostic reasoning: Implications for medical education." *Medical Teacher* **30** (7): e184-e188. doi.org/10.1080/01421590802144260

Epstein, Steven. 1996. *Impure science: AIDS, activism, and the politics of knowledge*. (Vol. 7). Berkeley: University of California Press.

Fiscella, Kevin, Meldrum, Sean, Franks, Peter, Shields, Cleveland G., Duberstein, Paul, McDaniel, Susan H. and Ronald M. Epstein. 2004. "Patient trust: is it related to patient-centered behavior of primary care physicians?." *Medical care*: 1049-1055. doi: 10.1097/00005650-200411000-00003.

Fox, Susannah and Maeve Duggan. 2013. "Health online 2013." *Pew Research Center*. <http://www.pewinternet.org/2013/01/15/health-online-2013/>

Future Advocacy. 2018. "Ethical, Social and Political Challenges of Artificial Intelligence in Health."

Garland, Ann F., Plemmons, Dena, & Leita Koontz. 2006. "Research-practice partnership in mental health: Lessons from participants." *Admin. and Policy in Mental Health and Mental Health Services Research* **33**(5): 517–528. DOI: 10.1007/s10488-006-0062-2

Gheiratmand, Mina, Rish, Irina, Cecchi, Guillermo A., Brown, Matthew RG, Greiner, Russell, Polosecki, Pablo I., Bashivan, Pouya, Greenshaw, Andrew J., Ramasubbu, Rajamannar and Serdar M. Dursun. 2017. "Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms." *NPJ Schizophrenia* **3**(22): 1-12. doi.org/10.1038/s41537-017-0022-8

Goldstein, Kurt. 1995 [1939]. *The organism: A holistic approach to biology derived from pathological data in man*. Zone Books, Brooklyn, NY.

- Gray, Bradford H. 1997. "Trust and trustworthy care in the managed care era." *Health Affairs* 16(1): 34-49. doi.org/10.1377/hlthaff.16.1.34
- Greenhalgh, Trisha, Wherton, Joseph, Papoutsis, Chrysanthi, Lynch, Jennifer, Hughes, Gemma, Hinder, Susan, Fahy, Nick, Procter, Rob and Sara Shaw. 2017. "Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies." *Journal of medical Internet research* 19(11): e367. DOI: 10.2196/jmir.8775
- Gusnard, Debra. A. and Marcus. E. Raichle. 2001. "Searching for a baseline: functional imaging and the resting human brain." *Nature Neuroscience Rev.* 2, 685–694. doi.org/10.1038/35094500
- Hall, Mark A., Camacho, Fabian, Dugan, Elizabeth and Rajesh Balkrishnan. 2002. "Trust in the medical profession: conceptual and measurement issues." *Health services research* 37(5): 1419-1439.
- Holzinger, Andreas, Biemann, Chris, Pattichis, Constantinos S. and Douglas B. Kell. 2017. "What do we need to build explainable AI systems for the medical domain?". *arXiv preprint arXiv:1712.09923*.
- Ingebrigtsen, Tor, Georgiou, Andrew, Clay-Williams, Robyn, Magrabi, Farah, Hordern, Antonia, Prgomet, Mirela, Li, Julie, Westbrook, Johanna and Jeffrey Braithwaite. 2014. "The impact of clinical leadership on health information technology adoption: systematic review." *International journal of medical informatics* 83(6): 393-405.
- Joshi, Indra and Jessical Morley. (eds). 2019. *Artificial Intelligence: How to get it right. Putting policy into practice for safe data-driven innovation in health and care*. London, United Kingdom: NHSX
- Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.
- Khushf, George. A. 2013. "Framework for Understanding Medical Epistemologies." *Journal of Medicine and Philosophy*, 38: 461–486. doi.org/10.1093/jmp/jht044
- Klein, E., Solomon, Andrew J., Corboy, John R. and James Bernat. 2016. "Physician compensation for industry-sponsored clinical trials in multiple sclerosis influences patient trust." *Multiple sclerosis and related*

*disorders* 8: 4-8. doi.org/10.1016/j.msard.2016.04.001

- Larson, Eric B. 2013. "Building trust in the power of "big data" research to serve the public good." *Jama* 309(23): 2443-2444. doi:10.1001/jama.2013.5914
- Leblond, Richard F. 2013. "An epistemology for clinical medicine: An argument for reflection on the ends of medical practice and ways of knowing with implications for the selection and training of physician." *Transactions of the American Clinical and Climatological Association* 124: 238-249. PMID: PMC3715941.
- Liberati, Elisa G., Ruggiero, Francesca, Galuppo, Laura, Gorli, Mara, González-Lorenzo, Marien, Maraldi, Marco, Pietro Ruggieri et al. 2017. "What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation." *Implementation Science* 12(1): 113.
- Logothetis, Nikos. K. 2008. "What we can do and what we cannot do with fMRI." *Nature* **453** (7197), 869-878. doi.org/10.1038/nature06976
- Luo, Jake, Wu, Min, Gopukumar, Deepika and Yiqing Zhao. 2016. "Big data application in biomedical research and health care: a literature review." *Biomedical informatics insights* 8: BII-S31559.
- Mahfoud, Tara. 2014. "Extending the mind: a review of ethnographies of neuroscience practice." *Frontiers in Human Neuroscience* 8 (359).
- Malterud, Kirsti. 1995. "The legitimacy of clinical knowledge: Towards a medical epistemology embracing the art of medicine." *Theoretical Medicine* 16: 183-198. doi.org/10.1007/BF00998544
- Mcknight, Harrison D. and Norman L. Chervany. 1996. "The Meanings of Trust." Technical Report 94-04, Carlson School of Management, University of Minnesota.  
[http://misrc.umn.edu/workingpapers/fullpapers/1996/9604\\_040100.pdf](http://misrc.umn.edu/workingpapers/fullpapers/1996/9604_040100.pdf)
- Miner, Adam S., Haque, Albert, Fries, Jason A., Fleming, Scott L., Wilfley, Denise E., Wilson, G. Terence, Arnold Milstein et al. 2020. "Assessing the accuracy of automatic speech recognition for psychotherapy." *npj Digital Medicine* 3(1): 1-8. <https://doi.org/10.1038/s41746-020-0285-8>
- Miner, Adam S., Shah, Nigam, Bullock, Kim D., Arnow, Bruce A., Bailenson, Jeremy and Jeff Hancock. 2019. "Key considerations for incorporating conversational AI in psychotherapy." *Frontiers in Psychiatry* 10: 746.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6813224/>

- Mittelstadt, Brent. D & Luciano Floridi. 2016. "The ethics of big data: current and foreseeable issues in biomedical contexts." *Science and Engineering Ethics* 22 (2), 303-341. doi.org/10.1007/s11948-015-9652-2
- Morley, Jessica, Machado, Caio CV, Burr, Christopher, Cows, Josh, Joshi, Indra, Taddeo, Mariarosaria and Luciano Floridi. 2020. "The Ethics of AI in Health Care: a Mapping Review." *Social Science & Medicine* (2020): 113172. doi.org/10.1016/j.socscimed.2020.113172
- Murray, Robin M. 2017. "Mistakes I Have Made in My Research Career." *Schizophrenia Bulletin* 43(2): 253–256. doi.org/10.1093/schbul/sbw165
- Nature Biotechnology. 2012. "What happened to personalized medicine?" *Nature Biotechnology* 30 (1). <https://doi.org/10.1038/nbt.2096>
- Nuffield Council on Bioethics. 2018. "Artificial intelligence (AI) in healthcare and research. Bioethics Briefing Note." *Nuffield Council on Bioethics*. May. <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf>.
- Patel, Vikram, Saxena, Shekhar, Lund, Crick, Thornicroft, Graham, Baingana, Florence, Bolton, Paul, Chisholm, Dan et al. 2018. "The Lancet Commission on global mental health and sustainable development". *Lancet* 392(10157): 1553–98. doi: 10.1016/S0140-6736(18)31612-X
- PricewaterhouseCoopers. 2012. "Social media 'likes' healthcare." PWC Health Research Institute." <https://www.pwc.com/us/en/health-industries/healthresearch-institute/publications/pdf/health-care-social-media-report.pdf>
- Polanyi, M. 2009. *The tacit dimension*. University of Chicago Press, Chicago & London.
- Roepstorff, Andreas. 2007. "Navigating the brainscape: When knowing becomes seeing." In Grasseni, C. (ed) *Skilled Vision* (Berghahn, Oxford, 191-206, 2007).
- Rockoff, J. D. 2016. "J & J to Stop Selling Automated Sedation System Sedasys." *Wall Street Journal*, 14 March. <https://www.wsj.com/articles/j-j-to-stop-selling-automated-sedation-system-sedasys-1457989723>
- Rose, Nikolas. 2013. "What is Diagnosis For?" *Talk given at the Institute of Psychiatry Conference on DSM-5 and the Future of Diagnosis*. 4 June.



- Rose, Nikolas. and Caitjan Gainty. 2019. "Neurovisions." *Advances in Clinical Neuroscience and Rehabilitation* **18** (2): 17-18. <https://www.acnr.co.uk/wp-content/uploads/2019/01/ACNR-NDJ19-low-rez-17-18.pdf>
- Saunders, Barry, F. 2008. *CT Suite: The Work of Diagnosis in the Age of Noninvasive Cutting*. Duke University Press, Durham & London.
- Shah, Nilay D., Steyerberg, Ewout W. and David M. Kent. 2018. "Big data and predictive analytics: recalibrating expectations." *Jama* 320(1): 27-28. doi:10.1001/jama.2018.5602.
- Schwartz, Robert C., and David M. Blankenship. 2014. "Racial disparities in psychotic disorder diagnosis: A review of empirical literature." *World journal of psychiatry* 4(4): 133. doi: 10.5498/wjp.v4.i4.133
- Simonite, T. 2016. "Automated Anesthesiologist Suffers a Painful Defeat." *MIT Technology Review*, <https://www.technologyreview.com/s/601141/automated-anesthesiologist-suffers-a-painful-defeat/>
- Smith, Rachel C, Bossen, Claus and Anne Marie Kanstrup. 2017. "Participatory design in an era of participation." *International Journal of CoCreation in Design and the Arts* 13(2): 65-69. doi.org/10.1080/15710882.2017.1310466
- Wynne, Brian. 2006. "Public engagement as a means of restoring public trust in science – Hitting the notes, but missing the music?" *Community Genetics*, 9(3): 211–220. doi.org/10.1159/000092659
- Stone, Peter, Brooks, Rodney, Brynjolfsson, Erik, Calo, Ryan, Etzioni, Oren, Hager, Greg, Hirschberg, Julia, et al. 2016. "Artificial Intelligence and Life in 2030. One hundred year study on artificial intelligence: Report of the 2015-2016 Study Panel." *Stanford University, Stanford, CA*. <http://ai100.stanford.edu/2016-report>.
- Sullivan, Greer, Duan, Naihua, Mukherjee, Snigdha, Kirchner, Joann, Perry, Dana and Kathy Henderson. 2005. "The role of services researchers in facilitating intervention research." *Psychiatric Services* 56(5): 537-542. doi.org/10.1176/appi.ps.56.5.537
- Thom, David. 2001. "Physician behaviors that predict patient trust. Stanford Trust SP." *Journal of Family Practice* **50**, 323–328. <https://go.gale.com/ps/anonymous?id=GALE%7CA74292253&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00943509&p=AONE&sw=w>

- Thorpe, Alistair, Sirota, Miroslav, Juanchich, Marie and Sheina Orbell. 2020. "Always take your doctor's advice': Does trust moderate the effect of information on inappropriate antibiotic prescribing expectations?" *British Journal of Health Psychology* 25(2): 358-376. <https://doi.org/10.1111/bjhp.12411>
- Von Schomberg, Rene, 2013. "A Vision of Responsible Research and Innovation." In R. Owen, M. Heintz and J Bessant (eds.) *Responsible Innovation Ch 3*. John Wiley, London, 51-74.
- Watson, David S., Krutzinna, Jenny, Bruce, Ian N., Griffiths, Christopher EM, McInnes, Iain B., Barnes, Michael R. and Luciano Floridi. 2019 "Clinical applications of machine learning algorithms: beyond the black box." *BMJ* **364** (1886): 1–9. [doi.org/10.1136/bmj.l886](https://doi.org/10.1136/bmj.l886)
- Wong, Ming C., Turner, Paul, & Kwang C Yee. 2008. "Involving clinicians in the development of an electronic clinical handover system-thinking systems not just technology." *Studies in health technology and informatics* **136**, 490-5. PMID: 18487779.
- Wessler, Benjamin S., Paulus, Jessica, Lundquist, Christine M., Ajlan, Muhammad, Natto, Zuhair, Janes, William A., Jethmalani, Nitin, Raman, Gowri, Lutz, Jennifer S., and David M. Kent. 2017. "Tufts PACE clinical predictive model registry: update 1990 through 2015." *Diagnostic and prognostic research* 1(1): 20. [doi.org/10.1186/s41512-017-0021-2](https://doi.org/10.1186/s41512-017-0021-2).

---