Taylor & Francis
Taylor & Francis Group

Should the Use of Adaptive Machine Learning Systems in Medicine be Classified as Research?

Robert Sparrow, Joshua Hatherley, Justin Oakley & Chris Bain

Published online: 25 Apr 2024.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

OPEN ACCESS

Check for updates

Should the Use of Adaptive Machine Learning Systems in Medicine be Classified as Research?

Robert Sparrow (iD), Joshua Hatherley (iD), Justin Oakley (iD) and Chris Bain (iD)

Monash University

ABSTRACT
A novel advantage of the use of machine learning (ML) systems in medicine is their potential to continue learning from new data after implementation in clinical practice. To date, considerations of the ethical questions raised by the design and use of adaptive machine learning systems in medicine have, for the most part, been confined to discussion of the so-called "update problem," which concerns how regulators should approach systems whose performance and parameters continue to change even after they have received regulatory approval. In this paper, we draw attention to a prior ethical question: whether the continuous learning that will occur in such systems after their initial deployment should be classified, and regulated, as medical research? We argue that there is a strong *prima facie* case that the use of continuous learning in medical ML systems should be categorized, and regulated, as research and that individuals whose treatment involves such systems should be treated as research subjects.

INTRODUCTION

Machine learning has the potential to generate exciting advances in medicine. It also raises many ethical issues (Char, Shah, and Magnus 2018; Grote and Berens 2020; Sparrow and Hatherley 2019; Svensson and Jotterand 2022; Vayena, Blasimme, and Cohen 2018). In this paper we identify an ethical—and regulatory—question regarding the use of *adaptive* machine learning in medicine that has not, to our knowledge, previously been discussed. An advantage of the use of machine learning (ML) systems in medicine is their potential to continue learning from data gathered in clinical practice. To date, considerations of the ethical questions raised by the use of adaptive machine learning systems in medicine have, for the most part, been confined to discussion of the "update problem," which concerns how regulators should approach systems whose performance and parameters continue to change even after they have received regulatory approval. We draw attention to a prior ethical question: whether the continuous learning that will occur in such systems after their initial deployment should be classified, and regulated, as medical research? We argue that there is a strong *prima facie* case that the use of continuous

learning in medical ML systems should be categorized as research and that individuals whose treatment involves such systems should be regarded as research subjects. We also discuss some implications of, and possible responses to, this conclusion.

ADAPTIVE MACHINE LEARNING IN MEDICINE

Machine learning is a subdiscipline of Artificial Intelligence (AI) research that "addresses the question of how to build computers that improve automatically through experience" (Jordan and Mitchell 2015, 255). It is widely believed that ML has the potential to revolutionize medicine (Esteva et al. 2019; Rajkomar, Dean, and Kohane 2019; Rajpurkar et al. 2022; Sparrow and Hatherley 2020). ML systems are being developed for a diverse range of clinical tasks, including diagnosis, prognostication, and patient monitoring. They are being pursued with especial vigor in radiology, pathology, and oncology, which make extensive use of medical imaging technologies (Rajkomar, Dean, and Kohane 2019). The pace of regulatory approvals for such devices has accelerated significantly in recent years (Lyell et al. 2021).

ML for medical use can be divided into two kinds—"locked" and "adaptive" systems. Locked systems have parameters and functions that are fixed prior to their clinical application: they "provide the same result each time the same input is provided" [Food and Drug Administration (FDA) 2019, 5]. Recently, however, researchers have become increasingly interested in the use of "adaptive" ML systems in medicine (Li et al. 2023; Vokinger, Feuerriegel, and Kesselheim 2021). An adaptive ML system has the capacity to change and improve in performance over time through exposure to the new data that it gathers after it is deployed and as it is used in practice: it engages in "continuous learning". We shall refer to such (Medical) Adaptive Machine Learning System(s) as "MAMLS."

A virtue of MAMLS is that they can be progressively "tuned" to the physiology of an individual patient or to patient demographics at a clinical site.[1] ML is already being combined with a variety of other emerging technologies (e.g. wearables, implantables, and microfluidics) to develop forms of "precision medicine" based on the data of particular individuals. For instance, ML-enabled devices are being investigated for personalized monitoring and detection of hypoglycemic events in diabetic patients (Porumb et al. 2020), for personalized detection of ventricular arrythmias (Jia et al. 2020), and to predict seizures in patients with drug resistant epilepsy (Cook et al. 2013; Pinto et al. 2021). Permitting such devices to learn continuously would allow them to adjust to changes in the patient's condition over their lifetime and might also make it easier to personalize them to patients in the first place. MAMLS could also be applied to improve knowledge of, and outcomes at, particular sites or institutions, by training them on data collected from the relevant cohort (Ong et al. 2021). For instance, they might be used to predict which patients might require readmission within some period if they are discharged from a particular hospital, or to identify individuals at high risk of suffering a heart-attack within a particular community (Yu et al. 2015).

THE UPDATE PROBLEM

Discussion of the ethical challenges associated with adaptive ML systems has centered around what Babic and coauthors (2019) refer to as the "update problem." Existing regulatory processes for medical products are ill-suited for regulating the manufacture and use of devices, such as MAMLS, the operations, and capacities, of which may change over time in the course of their clinical use. Left to continue learning post-deployment, MAMLS may adopt erroneous associations from new data that could jeopardize patient health. Indeed, adaptive medical ML systems are susceptible to the phenomenon of "catastrophic forgetting," in which a model completely overwrites itself during an update (Kirkpatrick et al. 2017; van de Ven and Tolias 2019). For this reason, regulatory approvals of medical ML have, for the most part, been restricted to locked systems.

The US Food and Drug Administration (FDA) grappled with the update problem in its 2019 *Proposed Regulatory Framework for Modifications to ML-Based Software as a Medical Device (SaMD)* and subsequent *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device Action Plan* (FDA 2021). Under the FDA's proposed framework, manufacturers would be required to submit Algorithmic Change Protocols (ACPs) for review, as part of their application for pre-market approval of any medical AI device.[2] The function of ACPs is to delineate how an AI can be expected to change and learn over time, and how manufacturers plan to mitigate the risks associated with these changes.

The FDA's proposal has itself been subject to criticism on a number of grounds (Babic et al. 2019; Gerke et al. 2020). In particular, the FDA offers little information as to how they propose to monitor the performance and use of these systems post-deployment: the FDA even suggests that such monitoring be performed by the manufacturers themselves. We are sympathetic to many of these criticisms. However, we believe that there is an important prior ethical and regulatory question, which has so far received little attention: whether the post-deployment learning that will proceed in these systems should be classified, and regulated, as medical research?[3]

---

[1]Our argument below implies that individuals whose treatment involves such devices should be understood to be primarily research subjects rather than patients. However, for ease of expression in what follows we will continue to refer to them as patients until we have presented the argument for this conclusion. Similarly, in what follows, we shall refer to MAMLS as being used in the "treatment" of patients although in many cases such systems are likely to be used primarily for diagnosis or prognosis: this usage is justified by the fact that the (for instance) diagnostic use of MAMLS will inevitably play an important role in shaping the course of the patient's treatment.

[2]The FDA document *Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff* (FDA 2022) is also relevant here insofar as it provides guidance as to the scope of FDA regulation of Clinical Decision Support Software as a medical device.

[3]The nearest thing to a discussion of this question that we have been able to identify in the literature is a discussion of what appropriate

THE MORAL SIGNIFICANCE OF THE DISTINCTION BETWEEN CLINICAL PRACTICE AND RESEARCH

Medical activities are traditionally categorized as either clinical practice or clinical research. According to the authoritative Belmont Report, clinical practice

> refers to interventions that are designed solely to enhance the wellbeing of an individual patient or client and that have a reasonable expectation of success. The purpose of medical or behavioral practice is to provide diagnosis, preventive treatment or therapy to particular individuals [National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (NCPHSBBR) 1978: Part A].

By contrast, "research"

> designates an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships). [NCPHSBBR 1978: Part A; see Council for International Organizations of Medical Sciences (CIOMS) 2016, xii].

This difference between the purposes of clinical research and clinical practice explains why this distinction is morally significant. Undertaking an activity aimed at developing generalizable medical knowledge involves adopting a role the governing values of which differ significantly from the values that govern the role of a medical doctor in clinical practice (Churchill 1980; Litton and Miller 2005; Oakley 2019). Indeed, the pursuit of knowledge sometimes requires individuals who are both physicians and clinical researchers to act in ways that they would not if motivated by a concern for—and that are perhaps even contrary to—the best interests of their own patients.[4]

It is precisely because the purpose of research is something other than the best interests of the research subject that such activity requires ethical oversight by, for example, an Institutional Research Board (IRB) (NCPHSBBR 1978: Part A). Research can often generate risks and non-clinical burdens to research subjects that are not compensated for by expected clinical benefits. The nature and dynamics of the relationship between researchers and researcher subjects differ from those between clinicians and patients. The benefits and burdens associated with research can also be, and often historically have been, distributed unequally and inequitably. For all these reasons, the conduct of research deserves especial ethical scrutiny and the classification of an activity as medical research rather than clinical practice brings into play a distinct schema of intuitions, institutions, and regulations.

IS CONTINUOUS LEARNING "RESEARCH"?

There has been a flurry of interest in the last 5 years in research governance of activities in biomedicine involving AI and the ethical issues that arise in the course of them (See, for instance: Angus 2020; Cruz Rivera et al. 2020; Genin and Grote 2021; Grote 2022; Liu et al. 2020; Park et al. 2020; Topol 2020). Writing in this journal, Melissa McCradden and coauthors have recently put forward a research ethics framework for the clinical evaluation of medical ML systems (McCradden et al. 2022; see also McCradden, Stephenson, and Anderson 2020).[5] Although McCradden and coauthors' research ethics framework provides support for our argument here, insofar as it assumes that the initial training of ML systems should be understood as research, their primary focus is on the question of whether human efforts to clinically validate ML systems should be classified as research and how they should be regulated. Our concern is quite different and relates to how we should understand the learning that goes on in—or, better, via—the ML system itself as it continues to gather data and refine its algorithm or model as people use it.

Prima facie grounds for thinking that *this* activity should be understood, and regulated, as research is provided by the fact that the post-deployment processes of adaptive learning in such systems are essentially the same as the processes involved in training them in the first place, which, as McCradden et al. argue, should be categorized as research (McCradden,

---

oversight of clinical decision support systems might look like (Evans and Whicher 2018). However, this paper does not discuss the issues associated with continuous learning with which we are concerned here.

[4]For example, it is important for physician-researchers to avoid compromising the scientific integrity of a clinical trial through 'selection bias', by selectively enrolling their own patient in a clinical trial likely to benefit that patient despite the fact that the patient does not meet the inclusion criteria of the study.

[5]Their framework consists of three phases. First, the exploratory ML research phase involves applying ML techniques to retrospective datasets with the aim of developing models that can identify and predict health-related patterns and events. Second, the silent evaluation phase involves the non-interventional evaluation of a model's performance on prospective data in a real-world clinical setting. Third, the prospective clinical evaluation phase involves the evaluation of an ML model's influence upon patient health outcomes in a real-world clinical environment through observational, quasi-interventional, and/or interventional studies, with the aim of determining whether using an ML system generates superior outcomes to an existing standard of care.

Stephenson, and Anderson 2020; McCradden et al. 2022). If collecting and using medical data to train a model is research, then continuing to collect data generated during the application of the MAMLS in clinical practice and training the model on that data should also be considered research. Moreover, there is a clear sense in which the continuous learning of an MAMLS *is* "an activity designed to test a hypothesis" and to "permit conclusions to be drawn": the patient's data is not being used solely to inform his/her treatment; it is also being used to refine the model upon which the MAMLS relies. In most cases, although the patient stands to benefit from using a system informed by machine learning, they would get the same benefit out of a system that did not use their data to facilitate continuous learning. The aim of the continuous learning is therefore to produce generalizable knowledge, which might benefit *other* patients in the future, rather than to serve the best interests of the particular patient.

Granted, it can be difficult to identify the hypotheses that ML systems are testing as they gather more data or the improvement in understanding of the world that is achieved when they revise their algorithms or models, especially where the systems are essentially "black boxes" by virtue of the involvement of deep learning (Burrell 2016). The difficulty in doing so is one reason to be cautious about the use of such systems. The popular, and influential, demand that AI should be "explainable" is justified, in part, by the concern that unless we understand how and why systems are producing the outputs they do, we will not be justified in relying on them (Shortliffe and Sepúlveda 2018; Vayena, Blasimme, and Cohen 2018). This demand is also revealing, we believe, insofar as it implies that AI—including ML—systems generate and contain knowledge about the world, albeit in a form that can be hard for human beings to access (Bjerring and Busch 2021, 349–351). Indeed, the very description of what goes on in ML systems as machine *learning* implies that such machines produce knowledge about the world. The process whereby ML systems develop and expand this knowledge is, *prima facie*, a continuation of a program of research initiated by the designers of the MAMLS.

There is an important class of exceptions to this claim, which consists in those devices wherein continuous learning is used to personalize a device to an individual over an extended period. For instance, continuous learning might be employed in implantable medical devices to predict the onset of epileptic seizures or ventricular arrythmias in a particular patient and to revise and refine the capacity to do so over the life-course of that patient. In such cases, it is plausible to hold that the aim of the continuous learning *is* to serve the interests of the individual patient and that the use of the MAMLS would therefore constitute clinical practice—or perhaps quality improvement—rather than research: there would be no attempt to discover generalizable knowledge. In order to remain securely within this category though, it must not be the case that the results of the continuous learning are also used to improve the functioning of systems used by people elsewhere or in the future. This would, for instance, exclude devices that contribute to "federated learning" for the sake of improving other, or future, devices (Rieke et al. 2020).

Another reason for thinking that continuous learning is research is that individuals engaging with MAMLS will be entering into a morally significant relationship with people who are not directly involved in their care. The goal of the designers of the system—or of those overseeing the continuous learning—is to improve the model of the world instantiated in the ML system, rather than to treat or diagnose the individual patient, and the user is being enlisted in this project.

These relationships generate moral hazards (Rowell and Connelly 2012). That is, they create a situation in which one party (the researcher) has incentives to disregard risks that they impose on another (the patient/subject).

Manufacturers of MAMLS have a financial interest in their products improving via continuous learning. The more people use their product, the better it will get. There is, therefore, a risk that researchers will encourage patients to use an MAMLS when it is not clinically indicated, for the sake of being able to use their data to train the MAMLS. This may be especially tempting where a patient is from a rare demographic or has an unusual disease progression and/or set of symptoms, such that their data would be especially useful for training the system. This risk is arguably higher with ML systems than with other medical products due to the larger role played by "network effects" (Katz and Shapiro 1985) in the market for the former. Particularly when a MAMLS is first being deployed, then, designers and manufacturers have a powerful incentive to try to increase the number of people using it.

Conversely, the goal of improving the performance of the MAMLS via continuous learning may sometimes be served by excluding particular patients, or classes of patients, from using the system even if they might stand to benefit from doing so. More generally,

as with medical research elsewhere, the pursuit of generalizable knowledge via continuous learning may require acting in ways that are not in the best interests of the individual patient.

It might be objected that the continuous learning of MAMLS—or at least continuous learning that will only lead to changes that the FDA is willing to approve—does not generate risk of harms to patients and that, for this reason, there is no need to ensure that patients are protected by being classified as research subjects.

However, this objection neglects the fact that the presence of risk is not the only grounds for distinguishing research from clinical practice: the nature of the relationship between the researcher and the research subject also matters. For instance, where a clinician is required to have the best interests of the patient at the forefront of their mind, it is appropriate for the researcher to be primarily motivated by the desire to generate new findings. The different roles of clinicians and researchers mean that there are different characteristic virtues and vices associated with these roles: it also means that the relations between clinicians and patients and researchers and research subjects tend to flourish or go wrong in different ways (Oakley 2019).

Moreover, the claim that continuous learning does not involve risk to patients is false. As we noted above, "continuous learning" does not necessarily mean continuous improvement: there is a chance that MAMLS may learn erroneous—and therefore dangerous—associations in the course of this learning. Even if it is possible to be confident that the performance of the MAMLS will evolve only as stipulated in the (approved) APC, the existence of risk is why there is an APC, and risk management plan, in the first place. It would be disingenuous to suggest that the fact that efforts had been taken to reduce the risk meant that there was not a risk.

Indeed, the mere fact that the operations of these systems will change as they learn does generate risks to those whose care is impacted by their use. Clinicians' understanding of how the system operates, acquired when they first encounter the system, and how the system actually operates may come apart over time as the system continues to evolve in response to data gathered in the course of its use, which in turn may lead to medical errors (Hatherley, Sparrow, and Howard 2022). This risk is especially high if the ML is a "black box"—if the details of its internal operations are not available to users—and remains even if the performance of the MAMLS itself only improves as it accumulates more data. For instance, if clinicians do not understand the extent of this improvement, they may over-trust or under-trust the outputs of the MAMLS, at the expense of the best interests of their patients.

Finally, the use of MAMLS, as opposed to ML systems that do not engage in continuous learning, may generate extra clinical, and non-clinical, burdens for patients. For instance, doctors or administrators will often need to collect information about patients and their health outcomes in order to provide and/or label new data to facilitate the MAMLS' continuous learning. The collection and storage of this data may pose an extra risk to the privacy of patients (Price and Cohen 2019). Facilitating the system's continuous learning may also require that patients undergo tests and examinations beyond that which would typically be required to serve their individual medical interests, which may expose them to risk of iatrogenic harms (for instance, as a result of the generation of incidental findings); it may also require them to attend clinic, or even incur extra expenses, where they would not otherwise need to do so. We acknowledge that this concern is somewhat speculative, but deny that it is excessively so: it is hardly unprecedented for researchers to require that research subjects undergo extra tests or procedures for the sake of gathering data for research purposes.

Both separately and together, these considerations establish a strong case that the use of MAMLS is research rather than treatment, by virtue of the continuous learning that will go on in these systems, and that the individuals whose data is being used to improve these systems are research subjects rather than patients.

AN ALTERNATIVE FRAMING? LEARNING HEALTHCARE SYSTEMS

As we discuss further below, acknowledging that the post-deployment training that occurs in MAMLS is research would have large ramifications for how the use of these systems would need to be regulated. For this reason, manufacturers of MAMLS, as well as clinicians who are keen to use them, have a strong incentive to resist the classification of the use of MAMLS as research. Moreover, both the distinction between research and clinical practice, and the ethical significance of this distinction, are more-and-more contested as a result of increasing awareness of the ways in which existing and emerging technologies and institutional practices elide or blur the line between them (Kass et al. 2013). In particular, some authors now defend the pursuit of knowledge in the context

of "learning healthcare systems", or—more narrowly—quality improvement, and/or surgical innovation, and argue that such should not be categorized as "research". In this section we therefore consider whether MAMLS might be better conceptualized via one of these alternative framings.

According to the Institute of Medicine, a learning healthcare system (LHCS) is one "in which knowledge generation is so embedded into the core of the practice of medicine that it is a natural outgrowth and product of the healthcare delivery process and leads to continual improvement in care" (Olsen, Aisner, and McGinnis 2007).

In a LHCS framework, a "learning activity" "(1) involves the delivery of health care services or uses individual health information, and (2) has a targeted objective of learning how to improve clinical practice or the value, quality, or efficiency of the systems, institutions, and modalities through which health care services are provided" (Faden et al. 2013, S19). This broad category includes clinical activities that lie in between the traditional categories of clinical research and clinical practice, such as quality assurance and improvement, auditing, and surgical innovation. According to Faden et al. (2013), the need for third-party ethical oversight should be evaluated on the basis of the risks and benefits of the learning activity in question, rather than the classification of the activity as either "research" or "practice" (See also Kass et al. 2013).

It might therefore be suggested that the continuous learning that occurs in MAMLS should be adduced to the idea a learning healthcare system (LHCS). Where MAMLS train continuously on the data of an individual user with the sole purpose of improving outcomes for that user—and that user alone—then it may well be appropriate to view these as an instance of a learning healthcare system and, more specifically, as a form of quality improvement (see below). However, as we have observed, many MAMLS will learn for the sake of improving the treatment of *other* patients *by refining the model instantiated in the ML system*. That is, it is the generalizable knowledge encoded in the model that allows the MAMLS to serve the interests of patients: the purpose of continuous learning is the pursuit of this generalizable knowledge. Again, we would emphasize that if the initial training of an ML system counts as research, as McCradden et al. suggest (and we agree), then further training of the system after it is deployed should also be classified as research.

In any case, classifying the use of MAMLS as a learning activity would do little to resolve all the problems with their use that we highlight here. A number of authors have argued that, in at least some contexts, learning activities themselves should be governed by research protocols, or something similar, especially when participation in the activity places extra burdens on patients (Finkelstein et al. 2015; Largent, Miller, and Joffe 2013). We have suggested that this will often be the case with MAMLS. Second, as Brody and Miller (2013) have argued, even under a LHCS framework, it remains important to pay attention to the distinction between research and clinical practice in order to protect patient-subjects from exploitation or manipulation, since the activities associated with each entail significantly different relationships between patient-subjects and clinician-investigators. Again, we have suggested that the relationship between the designer of a MAMLS and the user differs significantly from that between clinician and patient. Finally, though it is possible that classifying the use of adaptive ML as a learning activity would reduce the administrative burden involved in using MAMLS, it does not resolve the problem that, as we discuss further below, such systems will sometimes impose burdens on particular patients, or classes of patients, that will be difficult to justify given the likelihood that a consequence of the continuous learning will be that the patient, or people relevantly like them, will actually get worse treatment in the future.

Quality Improvement?

One set of practices that sit with the broader category of learning healthcare systems are those associated with "quality improvement" (QI). QI refers to "systematic, data guided activities designed to bring about immediate, positive changes in the delivery of health care in particular settings" (Baily et al. 2006, S5). Unlike research, QI does not seek to produce generalizable findings. For this reason, QI activities are often understood to occupy a methodological and ethical "grey zone" between research and clinical practice. In many institutional contexts, classifying an activity as QI enables clinician-investigators to bypass research oversight mechanisms in order to carry out research-adjacent tasks more easily. It might therefore be suggested that MAMLS should be understood through the lens of a quality improvement framework.

However, again, except in the special case where a MAMLS is intended to train on the data of a single patient for the sake of improving that individual's care, the learning that occurs as a result of the continuing refinement of the model instantiated in the

MAMLS is both more fraught in terms of risk and more generalizable than that which occurs in QI. In some cases, the learning that goes on in MAMLS will help shape the treatment of all the patients that engage with it, or a similar system, in the future. Admittedly, where adaptive learning is used to train a system on data from a particular site, or cohort of patients, in order to improve their care, a QI improvement framework looks more plausible. However, even in these cases, the use of adaptive learning entails some risk as well as ethically significant changes in the relationships between patient-subjects and clinician-investigators that we ignore at our (and the users) peril. For this reason, we believe that any broad classification of the use of MAMLS as QI rather than research should be strenuously resisted.

Surgical Innovation?

Finally, it might be suggested that the use of MAMLS could be understood as analogous to the practice of surgical innovation. Surgical innovation is also typically understood to occupy a "grey area" between clinical research and clinical practice (Rogers, Hutchison, and McNair 2019). Specifically, surgical innovation is distinguished from clinical research on the grounds that the goal of surgical innovation is to provide better outcomes to individual patients, rather than to produce generalizable knowledge. Surgical exceptionalism is further defended on the grounds that randomized controlled trials and standard IRB oversight of surgical innovations generate serious ethical concerns. For instance, in order to generate a control group in a surgical trial, some patients may need to receive "sham" surgeries. Surgical innovations also typically lack clinical equipoise (Angelos 2010). Furthermore, randomized controlled trials of surgical innovation are often held to be methodologically impractical due to challenges associated with small sample sizes, measuring surgical outcomes, and standardizing procedures (Angelos 2010; Broekman, Carrière, and Bredenoord 2016).

However, few, if any, of these reasons for exempting surgical innovation from the category of research apply to MAMLS. First, as noted earlier, although using MAMLS is likely to contribute to patient care, the use of these systems also produces generalizable knowledge insofar as "continuous learning" involves improving the model instantiated in the MAMLS in order to benefit other users. Second, while it may be reasonable to expect—or at least hope—that an innovative surgery will provide benefits to the particular patient compared to the existing standard of surgical care, in the vast majority of cases the use of a patient's data for continuous learning does not offer them a clear benefit relative to a "locked" system. Nor do randomized controlled trials of MAMLS generate methodological and ethical concerns analogous to those generated by surgical innovation. Thus, the ethical obstacles used to justify "surgical exceptionalism" cannot be used to justify rejecting the claim that using MAMLS ought to be classified as research.

IF MAMLS ARE RESEARCH…

The conclusion that the process whereby MAMLS learn from data gathered during their use should be conceptualized as research is disconcerting for a number of reasons.

First, it implies that the use of these systems should be regulated by the institutions, and under the legislative and regulatory frameworks, that exist to protect the interests of human research subjects. In particular, as with other medical research, it should be subject to scrutiny by IRBs, or the local equivalents thereof. As the Belmont Report suggests,

> the general rule is that if there is any element of research in an activity, that activity should undergo review for the protection of human subjects (NCPHSBBR 1978: Part A).

IRBs are the mechanisms that most nations have settled on to scrutinize and thereby regulate medical research in order to protect human subjects and so if the use of MAMLS is medical research then it will fall under their purview as *per* other research involving ML (McCradden et al. 2022). This may impose significant administrative burdens on clinicians and institutions that wish to use them.

Second, and relatedly, there is a strong *prima facie* case that patients would need to provide written (informed) consent to participation in such research. Given that it is likely that many adaptive ML devices will be diagnostic systems, which are often employed in the absence of formal procedures to ensure and document patient consent, this will significantly increase the costs of using these systems, perhaps even to the point of rendering some impractical or uneconomic. However, depending on the level of risk judged to be involved to patients/research subjects, it is possible that it would be appropriate for IRBs to waive the need for consent, to permit clinicians to secure only oral—rather than written—consent to participation in research, or to assume that the consent to use the MAMLS in a clinical context includes

consent to participate in the research that is conducted while the device continues to gather data and to learn from it.[6] These mechanisms would make it more plausible to use MAMLS by reducing the administrative burden associated with the continuous learning that goes on in them being classified as research. It is, though, worth re-emphasizing that, as we observed above, the risks involved in the use of systems that learn continuously are not necessarily insignificant. We strongly doubt that any general, or *a priori,* conclusion as to the relative weight of the competing imperatives to permit patients to access the improvements in care facilitated by continuous learning and to protect the interests of those who become research subjects by virtue of the continuous learning that goes on in MAMLS is possible, at least in the short-to-medium term. Until we have much more experience with the use of MAMLS, IRBs will need to resolve the question of whether, or how, to solicit and record consent on a case-to-case basis in the context of a larger deliberation about the relative weight of these imperatives, which itself will need to be informed by consideration of the risks and benefits involved in the use of the particular MAMLS.

Finally, understanding continuous learning as research will have dramatic implications for the feasibility of certain otherwise desirable choices when it comes to the design of these systems.

Although the literature on the ethics of MAMLS is cognizant that such systems may evolve over time ("diachronic" *evolution*), it is less often recognized that such evolution is also likely to generate variation across space ("synchronic" *variation)* (Hatherley and Sparrow 2023). Small variations across training datasets can result in significant differences between the end states of different instantiations of the same ML system. MAMLS will be used by, or implanted in, individuals with different physiologies and/or disease progressions and deployed in clinical settings with different data collection policies and patient demographics, which will affect the datasets upon which these systems learn. For this reason, copies of the same base-level MAMLS deployed at different sites may eventually come to differ significantly in their operations and in their accuracy.

The potential of MAMLS to lead to synchronic variation generates a range of pressing ethical issues, which two of the authors have discussed elsewhere (Hatherley and Sparrow 2023). For current purposes, it will suffice to observe that some of these issues are so troubling that, we anticipate, designers of MAMLS may prefer to take steps to prevent synchronic variation from arising. In many—although perhaps not all—cases they will be able to do this by instituting a collective learning approach, wherein the systems would "pool" their data so that they all train on the same data and evolve in step.

However, as Futoma et al. (2020, e489) note, "the demand for universal rules—generalizability—often results in [AI] systems that sacrifice strong performance at a single site for systems with mediocre or poor performance at many sites". Indeed, the application of a one-size-fits-all model across different subpopulations can result in a model that is sub-optimal for all groups, or optimal only for the dominant subpopulation—a phenomenon known as "aggregation bias" (Suresh and Guttag 2019).

Consequently, a significant proportion of the people who contribute to the training of MAMLS that use collective learning will stand to gain nothing when the system "improves". Some research subjects will receive *worse* treatment once the MAMLS with which they engage updates (Hatherley and Sparrow 2023). This will also be true of people relevantly like them. Despite a long history of IRBs devoting concerted attention to the "risk-benefit" ratio present in research (King and Churchill 2011; Rajczi 2004), recent scholarship tends to reject the idea that there need be any expectation of personal benefit from participation in research or that there is a threshold of risk that it is unethical to ask research subjects to incur (Miller and Brody 2007). Nevertheless, it will be very hard to justify asking research subjects to incur risks and burdens in the course of research that not only holds out no prospect of benefitting them but is likely to mean that people like them get worse treatment in the future: to do so is to contravene the fundamental ethical prohibition on using people as mere means (Hatherley and Sparrow 2023). If MAMLS are research, then, this will place severe ethical limits on the use of collective, including federated, learning in such systems.

CONCLUSION

We have argued that there is a strong *prima facie* case that the use of continuous learning in medical ML systems should be categorized, and regulated, as

---

[6]In the US, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) permits the use of de-identified patient data for research without the explicit authorization of the patient. Importantly, however, whether it will be possible to achieve continuous learning with de-identified data and what counts as de-identified data in the context of the training of a MAMLS are likely to differ between different MAMLS.

research, and that individuals whose treatment involves MAMLS should be recognized to be research subjects, except where continuous learning is used solely to improve the care of an individual patient over time. The extra costs and burdens imposed by the need to conduct ethical review of this research and to secure consent to participation in research from those using the systems may significantly reduce the incentives healthcare providers have to adopt MAMLS, that clinicians have to utilize them, and, thus, that manufacturers have to develop these systems. Given the potential benefits associated with the use of machine learning in medicine it is worth considering how designers—and regulators—might respond to this conundrum.

One option would be to give up on the project of *continuous* learning in medical ML systems entirely. Alternatively, the use of continuous learning in medical ML might be abandoned except where it will be used solely to improve the care of an individual patient over time. Patients might still benefit from the use of ML but would miss out on many of the benefits associated with continuous learning. Given that the benefits of continuous learning might be substantial, this would be disappointing. However, it is possible that other ethical issues arising from the use of continuous learning in medicine, beyond those we have discussed here, may force the same conclusion.

Another option would be to carve out a regulatory exemption for the continuous learning of ML systems or adduce them to the category of learning healthcare systems (or some subset of learning healthcare systems such as quality improvement or surgical innovation). To do the former seems unprincipled: it neglects the distinctive features of the relationship between the designers and the research subjects whose data is being used to train these systems and the not-insignificant risks and burdens associated with continuous learning. A case can always be made for the benefits of research, but this is insufficient to establish that it should not be regulated. Adducing MAMLS to learning healthcare systems, or some subset thereof, seems more defensible, although the significance of doing so remains unclear while the debate about how best to regulate learning activities, including quality improvement, and surgical innovation, continues. Moreover, many MAMLS are likely to be an uneasy "fit" for this description: not every practice that involves learning is properly described as a "learning healthcare system". If this option is to be pursued, care will need to be taken to ensure that it does not lead to a general lowering of standards of regulation of research.

Finally, designers and regulators might choose to permit continuous learning to proceed in a subset of devices—and treat it as research—before rolling out the results to other devices. This would allow users to receive some—if not all—of the benefits of continuous learning but also ensure that the research was conducted ethically, with adequate protections for research subjects in place. It would, however, mean that the developers of MAMLS would need to address the challenges associated with the conduct of research that we have identified here whenever they seek to improve their products by means of continuous learning.

Carving out an exemption for MAMLS from the regulation of research, or allowing research to proceed in a subset of devices, would leave the original update problem intact. A full reckoning of the ethical and regulatory implications of continuous learning awaits a more extended philosophical engagement with the update problem and with other ethical issues that are likely to be raised by this technology. Our investigation here, then, is only a part of what is required before we can be confident that we can use MAMLS ethically. However, we hope that by drawing attention to the question of whether the use of MAMLS should be categorized—and regulated—as research, this paper has made a useful contribution to this larger project.

AUTHORS' CONTRIBUTIONS

All four authors contributed to the design of the research. Sparrow and Hatherley wrote the bulk of the first draft, with Oakley contributing. Bain provided comments, edits, and further input. All four authors discussed, substantially revised, and approved the publication of the final text.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

ORCID

Robert Sparrow   http://orcid.org/0000-0001-6093-9456
Joshua Hatherley   http://orcid.org/0000-0002-8581-9669
Justin Oakley   http://orcid.org/0000-0003-3609-0020
Chris Bain   http://orcid.org/0000-0002-5281-2364

DATA AVAILABILITY STATEMENT

This research did not generate empirical data.

REFERENCES

Angelos, P. 2010. The ethical challenges of surgical innovation for patient care. *The Lancet* 376 (9746):1046–7. doi: 10.1016/s0140-6736(10)61474-2.

Angus, D. C. 2020. Randomized clinical trials of artificial intelligence. *JAMA* 323 (11):1043–5. doi: 10.1001/jama.2020.1039.

Babic, B., S. Gerke, T. Evgeniou, and I. G. Cohen. 2019. Algorithms on regulatory lockdown in medicine. *Science* 366 (6470):1202–4. doi: 10.1126/science.aay9547.

Baily, M. A., M. Bottrell, J. Lynn, and B. Jennings. 2006. The ethics of using QI methods to improve health care quality and safety. *The Hastings Center Report* 36 (4):S1–S40. doi: 10.1353/hcr.2006.0054.

Bjerring, J. C., and J. Busch. 2021. Artificial intelligence and patient-centered decision-making. *Philosophy & Technology* 34 (2):349–71. doi: 10.1007/s13347-019-00391-6.

Brody, H., and F. G. Miller. 2013. The research-clinical practice distinction, learning health systems, and relationships. *The Hastings Center Report* 43 (5):41–7. doi: 10.1002/hast.199.

Broekman, M. L., M. E. Carrière, and A. L. Bredenoord. 2016. Surgical innovation: The ethical agenda. *Medicine* 95 (25):e37980. doi: 10.1097/MD.0000000000003790.

Burrell, J. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3 (1). doi: 10.1177/2053951715622512.

Char, D. S., N. H. Shah, and D. Magnus. 2018. Implementing machine learning in health care—addressing ethical challenges. *The New England Journal of Medicine* 378 (11):981–3. doi: 10.1056/NEJMp1714229.

Churchill, L. R. 1980. Physician-investigator/patient-subject: Exploring the logic and the tension. *The Journal of Medicine and Philosophy* 5 (3):215–24. doi: 10.1093/jmp/5.3.215.

Cook, M. J., T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka, et al. 2013. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study. *The Lancet Neurology* 12 (6):563–71. doi: 10.1016/S1474-4422(13)70075-9.

Council for International Organizations of Medical Sciences (CIOMS). 2016. *International ethical guidelines for health-related research involving humans.* 4th ed. Geneva: Council for International Organizations of Medical Sciences (CIOMS).

Cruz Rivera, S., X. Liu, A.-W. Chan, A. K. Denniston, and M. J. Calvert, The SPIRIT-AI and CONSORT-AI Working Group. 2020. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *The Lancet Digital Health* 2 (10):e549–e560. doi: 10.1016/S2589-7500(20)30219-3.

Esteva, A., A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. S. Corrado, S. Thrun, and J. Dean. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25 (1):24–9. doi: 10.1038/s41591-018-0316-z.

Evans, E. L., and D. Whicher. 2018. What should oversight of clinical decision support systems look like? *AMA Journal of Ethics* 20 (9):E857–E863.

Faden, R. R., N. E. Kass, S. N. Goodman, P. Pronovost, S. Tunis, and T. L. Beauchamp. 2013. An ethics framework for a learning health care system: A departure from traditional research ethics and clinical ethics. *The Hastings Center Report* 43 (s1):S16–S27. doi: 10.1002/hast.134.

Finkelstein, J. A., A. L. Brickman, A. Capron, D. E. Ford, A. Gombosev, S. M. Greene, R. P. Iafrate, L. Kolaczkowski, S. C. Pallin, M. J. Pletcher, et al. 2015. Oversight on the borderline: Quality improvement and pragmatic research. *Clinical Trials* 12 (5):457–66. doi: 10.1177/1740774515597682.

Food and Drug Administration (FDA). 2019. *Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) – discussion paper and request for feedback.* Silver Spring, MD: US Food & Drug Administration.

Food and Drug Administration (FDA). 2021. *Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan.* Silver Spring, MD: US Food & Drug Administration.

Food and Drug Administration (FDA). 2022. *Clinical decision support software: Guidance for industry and Food and Drug Administration staff.* Silver Spring, MD: US Food & Drug Administration.

Futoma, J., M. Simons, T. Panch, F. Doshi-Velez, and L. A. Celi. 2020. The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health* 2 (9):e489–e492. doi: 10.1016/S2589-7500(20)30186-2.

Genin, K., and T. Grote. 2021. Randomized controlled trials in medical AI: A methodological critique. *Philosophy of Medicine* 2 (1):1–15. doi: 10.5195/pom.2021.27.

Gerke, S., B. Babic, T. Evgeniou, and I. G. Cohen. 2020. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digital Medicine* 3 (1):53. doi: 10.1038/s41746-020-0262-2.

Grote, T. 2022. Randomised controlled trials in medical AI: Ethical considerations. *Journal of Medical Ethics* 48 (11):899–906. doi: 10.1136/medethics-2020-107166.

Grote, T., and P. Berens. 2020. On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics* 46 (3):205–11. doi: 10.1136/medethics-2019-105586.

Hatherley, J., and R. Sparrow. 2023. Diachronic and synchronic variation in the performance of adaptive machine learning systems: The ethical challenges. *Journal of the American Medical Informatics Association* 30 (2):361–6. doi: 10.1093/jamia/ocac218.

Hatherley, J., R. Sparrow, and M. Howard. 2022. The virtues of interpretable medical artificial intelligence. *Cambridge Quarterly of Healthcare Ethics*. Published online, 16 December 2022. doi: 10.1017/S0963180122000305.

Jia, Z., Z. Wang, F. Hong, L. Ping, Y. Shi, and J. Hu. 2020. Personalized deep learning for ventricular arrhythmias detection on medical IoT systems. In: *ICCAD '20: Proceedings of the 39th International Conference on Computer-Aided Design,* November 2-5, 2020, Virtual Event, USA, 1–9. New York, NY: Association for Computing Machinery. doi: 10.1145/3400302.3415774.

Jordan, M. I., and T. M. Mitchell. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349 (6245):255–60. doi: 10.1126/science.aaa8415.

Kass, N. E., R. R. Faden, S. N. Goodman, P. Pronovost, S. Tunis, and T. L. Beauchamp. 2013. The research-treatment distinction: A problematic approach for determining which activities should have ethical oversight. *The Hastings Center Report* 43 (s1):S4–S15. doi: 10.1002/hast.133.

Katz, M. L., and C. Shapiro. 1985. Network externalities, competition, and compatibility. *The American Economic Review* 75 (3):424–40.

King, N. M. P., and L. R. Churchill. 2011. Assessing and comparing potential benefits and risks of harm. In *The Oxford textbook of clinical research ethics*, eds. E. J. Emanuel, C. C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller, and D. D. Wendler, 514–526. Oxford: Oxford University Press.

Kirkpatrick, J., R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America* 114 (13):3521–6. doi: 10.1073/pnas.1611835114.

Largent, E. A., F. G. Miller, and S. Joffe. 2013. A prescription for ethical learning. *The Hastings Center Report* 43 (s1):S28–S29. doi: 10.1002/hast.135.

Li, J., L. Jin, Z. Wang, Q. Peng, Y. Wang, J. Luo, J. Zhou, Y. Cao, Y. Zhang, M. Zhang, et al. 2023. Towards precision medicine based on a continuous deep learning optimization and ensemble approach. *NPJ Digital Medicine* 6 (1):18. doi: 10.1038/s41746-023-00759-1.

Litton, P., and F. Miller. 2005. A normative justification for distinguishing the ethics of clinical research from the ethics of medical care. *The Journal of Law, Medicine & Ethics* 33 (3):566–74. doi: 10.1111/j.1748-720x.2005.tb00519.x.

Liu, X., S. Cruz Rivera, D. Moher, M. J. Calvert, and A. K. Denniston, the SPIRIT-AI and CONSORT-AI Working Group. 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *The Lancet Digital Health* 2 (10):e537–e548. doi: 10.1016/S2589-7500(20)30218-1.

Lyell, D., E. Coiera, J. Chen, P. Shah, and F. Magrabi. 2021. How machine learning is embedded to support clinician decision making: An analysis of FDA-approved medical devices. *BMJ Health & Care Informatics* 28 (1):e100301. doi: 10.1136/bmjhci-2020-100301.

McCradden, M., J. A. Anderson, E. A. Stephenson, E. Drysdale, L. Erdman, A. Goldenberg, and R. Z. Shaul. 2022. A research ethics framework for the clinical translation of healthcare machine learning. *The American Journal of Bioethics* 22 (5):8–22. doi: 10.1080/15265161.2021.2013977.

McCradden, M. D., E. A. Stephenson, and J. A. Anderson. 2020. Clinical research underlies ethical integration of healthcare artificial intelligence. *Nature Medicine* 26 (9):1325–6. doi: 10.1038/s41591-020-1035-9.

Miller, F. G., and H. Brody. 2007. Clinical equipoise and the incoherence of research ethics. *The Journal of Medicine and Philosophy* 32 (2):151–65. doi: 10.1080/03605310701255750.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (NCPHSBBR). 1978. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Washington: Department of Health, Education, and Welfare.

Oakley, J. 2019. Virtues in research ethics: Developing an empirically-informed account of virtues in biomedical research practice. In *Beyond autonomy: Limits and alternatives to informed consent in research ethics and law*, ed. D. G. Kirchhoffer and B. J. Richards, 133–149. Cambridge: Cambridge University Press.

Olsen, L. A., D. Aisner, and J. M. McGinnis. 2007. *The learning healthcare system*. Washington, DC: National Academies Press.

Ong, C. S., E. Reinertsen, H. Sun, P. Moonsamy, N. Mohan, M. Funamoto, T. Kaneko, P. S. Shekar, S. Schena, J. S. Lawton, et al. 2021. Prediction of operative mortality for patients undergoing cardiac surgical procedures without established risk scores. *The Journal of Thoracic and Cardiovascular Surgery* 165 (4):1449.e15–1459.e15. doi: 10.1016/j.jtcvs.2021.09.010.

Park, Y., G. Purcell Jackson, M. A. Foreman, D. Gruen, J. Hu, and A. K. Das. 2020. Evaluating artificial intelligence in medicine: Phases of clinical research. *JAMIA Open* 3 (3):326–31. doi: 10.1093/jamiaopen/ooaa033.

Pinto, M. F., A. Leal, F. Lopes, A. Dourado, P. Martins, and C. A. Teixeira. 2021. A personalized and evolutionary al-

gorithm for interpretable EEG epilepsy seizure prediction. *Scientific Reports* 11 (1):3415. doi: 10.1038/s41598-021-82828-7.

Porumb, M., S. Stranges, A. Pescapè, and L. Pecchia. 2020. Precision medicine and artificial intelligence: A pilot study on deep learning for hypoglycemic events detection based on ECG. *Scientific Reports* 10 (1):170. doi: 10.1038/s41598-019-56927-5.

Price, W. N., and I. G. Cohen. 2019. Privacy in the age of medical big data. *Nature Medicine* 25 (1):37–43. doi: 10.1038/s41591-018-0272-7.

Rajczi, A. 2004. Making risk-benefit assessments of medical research protocols. *The Journal of Law, Medicine & Ethics* 32 (2):338–48, 192. doi: 10.1111/j.1748-720x.2004.tb00480.x.

Rajkomar, A., J. Dean, and I. Kohane. 2019. Machine learning in medicine. *The New England Journal of Medicine* 380 (14):1347–58. doi: 10.1056/NEJMra1814259.

Rajpurkar, P., E. Chen, O. Banerjee, and E. J. Topol. 2022. AI in health and medicine. *Nature Medicine* 28 (1):31–8. doi: 10.1038/s41591-021-01614-0.

Rieke, N., J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, et al. 2020. The future of digital health with federated learning. *NPJ Digital Medicine* 3 (1):119. doi: 10.1038/s41746-020-00323-1.

Rogers, W., K. Hutchison, and A. McNair. 2019. Ethical issues across the IDEAL stages of surgical innovation. *Annals of Surgery* 269 (2):229–33. doi: 10.1097/SLA.0000000000003106.

Rowell, D., and L. B. Connelly. 2012. A history of the term 'moral hazard'. *Journal of Risk and Insurance* 79 (4):1051–75. doi: 10.1111/j.1539-6975.2011.01448.x.

Shortliffe, E. H., and M. J. Sepúlveda. 2018. Clinical decision support in the era of artificial intelligence. *JAMA* 320 (21):2199–200. doi: 10.1001/jama.2018.17163.

Sparrow, R., and J. Hatherley. 2019. The promise and perils of AI in medicine. *International Journal of Chinese & Comparative Philosophy of Medicine* 17 (2):79–109. doi: 10.24112/ijccpm.171678.

Sparrow, R., and J. Hatherley. 2020. High hopes for 'deep medicine'? AI, economics, and the future of care. *The Hastings Center Report* 50 (1):14–7. doi: 10.1002/hast.1079.

Suresh, H., and J. V. Guttag. 2019. A framework for understanding unintended consequences of machine learning. Accessed March 1, 2022. https://arxiv.org/abs/1901.10002.

Svensson, A. M., and F. Jotterand. 2022. Doctor ex machina: A critical assessment of the use of artificial intelligence in health care. *The Journal of Medicine and Philosophy* 47 (1):155–78. doi: 10.1093/jmp/jhab036.

Topol, E. J. 2020. Welcoming new guidelines for AI clinical research. *Nature Medicine* 26 (9):1318–20. doi: 10.1038/s41591-020-1042-x.

van de Ven, G. M., and A. S. Tolias. 2019. Three scenarios for continual learning. Accessed March 1, 2022. https://arxiv.org/abs/1904.07734.

Vayena, E., A. Blasimme, and I. G. Cohen. 2018. Machine learning in medicine: addressing ethical challenges. *PLoS Medicine* 15 (11):e1002689. doi: 10.1371/journal.pmed.1002689.

Vokinger, K. N., S. Feuerriegel, and A. S. Kesselheim. 2021. Continual learning in medical devices: FDA's action plan and beyond. *The Lancet Digital Health* 3 (6):e337–e338. doi: 10.1016/S2589-7500(21)00076-5.

Yu, S., F. Farooq, A. van Esbroeck, G. Fung, V. Anand, and B. Krishnapuram. 2015. Predicting readmission risk with institution-specific prediction models. *Artificial Intelligence in Medicine* 65 (2):89–96. doi: 10.1016/j.artmed.2015.08.005.