Algorithms Advise, Humans Decide: the Evidential Role of the Patient Preference Predictor

Nicholas Makins

Forthcoming in *Journal of Medical Ethics*. Please cite published version.
DOI: 10.1136/jme-2024-110175

Abstract

An AI-based "patient preference predictor" (PPP) is a proposed method for guiding healthcare decisions for patients who lack decision-making capacity. The proposal is to use correlations between sociodemographic data and known healthcare preferences to construct a model that predicts the unknown preferences of a particular patient. In this paper, I highlight a distinction that has been largely overlooked so far in debates about the PPP–that between algorithmic prediction and decision-making–and argue that much of the recent philosophical disagreement stems from this oversight. I show how three prominent objections to the PPP only challenge its use as the sole determinant of a choice, and actually support its use as a source of evidence about patient preferences to inform human decision-making. The upshot is that we should adopt the evidential conception of the PPP and shift our evaluation of this technology towards the ethics of algorithmic prediction, rather than decision-making.

Keywords: patient preference predictor, artificial intelligence, capacity, medical decision-making.

1 Introduction

A well-worn aphorism in politics states that, "*advisers advise, ministers decide.*" We can think of advice here as taking the form of predictions like, "*policy A is likely to result in consequences x and y.*" The idea that we should distinguish between sources of advise and decisions is now well entrenched in many

areas of politics and beyond, including the growing field of AI ethics.[1-5] However, one debate in which this distinction has been somewhat overlooked concerns the design and implementation of the so-called patient preference predictor (PPP). This is a proposed AI system that would predict the healthcare preferences of individuals on the basis of known sociodemographic information about them, in order to guide medical decisions that they lack capacity to make for themselves.

The PPP promises to improve our ability to provide incapacitated individuals with healthcare that is aligned with their preferences, but it has proven controversial in the years since it was first proposed and a range of objections have been levelled against it. The aim of this paper is to show that much of this debate stems from a failure to pay proper attention to the distinction between algorithmic predictions and algorithmic decisions.[1] I will show that, once we conceive of the PPP as making probabilistic predictions that are used as evidence to inform human decision-making, rather than as making these decisions directly, a number of the most prominent objections to this sort of technology lose their bite. This means that the most promising interpretation of the proposal for a PPP is as a source of evidence about patients' preferences and our evaluation of it should therefore be focused on the ethics of algorithmic predictions.

The remainder of the paper will proceed as follows: Section 2 introduces the PPP and the motivation for developing it. Section 3 presents three types of objection to the proposed use of the PPP. Section 4 explains the distinction between predictions and decisions made by AI systems and applies this distinction to the PPP in particular. In Section 5, we return to the three objections to show how they no longer challenge the PPP once we think of it as having a predictive, evidential role. Section 6 concludes with some reflections on where this leaves the broader debate.

[1] Numerous advocates of the PPP already think of it as making predictions rather than decisions. See p.10 for a brief discussion and references.

2 Predicting Patients' Preferences

For situations in which individuals lack the capacity to make a particular decision about their own healthcare, there are two distinct guiding principles that predominate in medical ethics and law: the substituted judgement standard and the best interests standard. The substituted judgement standard states that one ought to make the decision that the incapacitated individual would have made if they had capacity to do so for themself. The best interests standard states that one ought to choose the option that will best promote the wellbeing of the individual on behalf of whom the choice is being made. On either view, the preferences of the individual are important. The substituted judgement standard can be thought of as the recommendation to choose the option that the patient would prefer. On the best interest standard, the satisfaction of an individual's preferences is typically taken to be at least good evidence for, and perhaps even constitutive of, an individual's wellbeing. So, whether guided by the substituted judgement or best interests standard, decision-makers have good reason to take a patient's preferences into consideration when making decisions about their healthcare.

It is troubling, therefore, that evidence appears to show that surrogate decision-makers are not very accurate in their predictions of patients' preferences.[6-13] Given the growing development of AI systems to make predictions in medical specialities including oncology,[14-18] cardiology,[19-23] and ophthalmology,[24-26] to name but a few, one might hope that similar technological solutions could be found to the problem of predicting the healthcare preferences of patients who lack capacity. Rid and Wendler have proposed and defended the use of an algorithm for this very purpose.[27-29] Their proposal is to develop a "patient preference predictor" (PPP) based on known correlations between individuals' healthcare preferences and sociodemographic data. This would then be used to predict the unknown preferences of incapacitated patients, on the basis of their own sociodemographic characteristics.[27,29-32] It is claimed that not only would such a system outperform the accuracy of surrogate decision-makers alone, but it would also

alleviate the distress that many surrogate decision-makers experience as a result of uncertainty about what the patient would want.

Despite the appealing features of this proposal, it faces a number of pressing objections. In what follows, I will describe three such objections, before moving on to show how their force can be undermined by a particular conception of the role of a PPP in the decision-making process.

3 Objections to the PPP

3.1 Bare Statistical Evidence

The first objection we will consider mirrors a well-known objection to the use of so-called "bare statistical evidence" in legal judgements. The apparent problem with bare statistical evidence in the law is illustrated by the following case. Suppose that a person's car is hit by a bus and damaged. This person is owed compensation by whichever bus company operated the bus that caused the collision. The case goes to court, but it is unknown which bus company was responsible. Moreover, the only available evidence that is relevant to the question of which bus company was responsible for the damage is the fact that 80% of the buses that drive on the road where the collision took place are operated by the Blue Bus Company and the remaining 20% are operated by the Red Bus Company. Although this makes it more likely than not that the Blue Bus Company was responsible, it is generally thought that they ought not to be found liable on the basis of this statistical evidence alone. This insufficiency of statistical evidence holds even though (1) the relevant standard of proof is the balance of probabilities (rather than the more demanding standard of beyond reasonable doubt), and (2) it would be deemed appropriate to find the Blue Bus Company liable on the basis of eye-witness testimony, even if eye-witness testimony was known to be accurate less than 80% of the time (i.e. even if the statistical evidence yielded a higher probability than the eye-witness testimony of it being a Blue Bus Company bus).

Sharadin suggests that the use of a PPP faces an analogous challenge to the problem of bare statistical evidence in legal proceedings.[33] Making decisions by using a PPP, so the argument goes, involves an objectionable reliance on statistical evidence in just the same way as in the bus company case. Pushing the analogy further, one can think of advance directives as playing a similar role to eyewitness testimony: there is something special about advance directives which means that they can form the basis of decisions that purely statistical evidence cannot, and this would hold even if we suppose that the PPP more accurately tracked patients' preferences at the time of treatment than advance directives.

There are various explanations of what is thought to be wrong with making legal decisions on the basis of bare statistical evidence, but one such explanation that carries over naturally to the context of medical decision making is that using statistical evidence treats people as though they are unable to diverge from statistical likelihoods and make autonomous choices for themselves. It is as though a person's membership of a particular group determines the decisions they would make, rather than their own autonomous agency. Sharadin suggests something along these lines when arguing that the problem may be treating people as though their preferences are caused by their demographic features.[33]

Finally, it is worth noting that Sharadin also offers a potential "debunking explanation" of our intuitions about bare statistical evidence, as a possible strategy for avoiding the objection. Ultimately, he argues for the conditional view that *if* there is a problem with bare statistical evidence in legal settings, *then* there is an equivalent problem with the PPP. We will return to this point in Section 5.

3.2 Advantages of surrogate decision-making

An alternative family of criticisms of the proposed use of the PPP stem from the claim that there are reasons to keep decision-making in the hands of surrogates that are not based on their ability to accurately identify what the patient themselves would have preferred. The central justification for the PPP is the claim that it would be able to predict individuals' preferences with greater accuracy than family, friends, and doctors, who are notoriously inaccurate in this regard. But if there are non-accuracy based considerations that support surrogate decision-making, then this justification may be undermined or outweighed.

For example, Um argues that people may trust their nearest and dearest to make decisions on their behalf not only because such people are likely to know their preferences, but because their relationship has produced a form of shared agency.[34] A surrogate's decisions may be a continued manifestation of this shared agency and therefore offer an important way of preserving the patient's autonomy when they are no longer able to decide for themselves. This argument reflects a broader set of views in medical ethics that challenge the "atomistic" conception of autonomy that separates a person from their deepest connections with other people.[35-36][2] These views are captured by Hardwig: "*There is no way to detach the lives of patients from the lives of those who are close to them. Indeed, the intertwining of lives is part of the very meaning of closeness.*"(p.5)[37] On this sort of view, involving the close friends and/or family of an individual is crucial for upholding that person's autonomy.

Even if one does not accept this picture of group agency and autonomy, the special status of surrogates can be justified on more theoretically modest grounds, given that entrusting a surrogate to make decisions on one's behalf is itself an expression of one's preferences. Therefore, an

[2] This literature often focuses on patients' families, but a person's nearest and dearest needn't be family and these arguments should not be thus constrained.

individual's autonomy can be respected once they are unable to decide for themselves by proceeding in accordance with their preference to have a particular individual make decisions on their behalf.[39-40]

3.3 Endorsed Reasons

The last challenge to the use of a PPP, presented by Stephen John, stems from the idea that respect for autonomy does not require the simple satisfaction of one's preference, but rather demands proper recognition of a person's capacity for rational deliberation.[41-42] This means that, when a choice must be made on behalf of an incapacitated patient, it should be decided on the basis of reasons that the patient would endorse as such in their own deliberation.

The problem for PPPs is that they would not do this, since they work with demographic facts, such as age and gender, that do not usually feature as reasons in people's own deliberation. The sorts of considerations that people usually consider when making significant healthcare decisions are things like what it will be like to experience the available options and their possible outcomes, and what effects these may have on one's daily life, work, and relationships. Conversely, these decisions are not usually made on the basis of what would be preferred by the majority of people to whom one is demographically similar. Therefore, if this is the correct conception of autonomy, then the claim that PPPs may support decision-making that respects patient autonomy faces a serious challenge. This is especially pressing given the central role that autonomy plays in motivating the case for a PPP.

It is important to note that John does not claim that this line of reasoning entirely rules out the use of PPPs. Rather, he suggests that it puts constraints on the kinds of category that a PPP can legitimately draw on, allowing only those that would be endorsed as reasons in the agent's own

deliberation. However, in the following sections I will show that we need not accept even this more cautious conclusion.

4 Predictions and Decisions

In the literature on AI ethics a simple, but important distinction is drawn between the predictions and the decisions that an algorithmic system may deliver. This distinction has often been overlooked, but is increasingly taking hold as relevant for understanding the normatively salient features of such systems.[1-5]

Beigang provides a precise characterisation of predictive models and decision functions, but for present purposes an informal explanation will suffice.[1] A predictive model takes specific values for a range of input variables and outputs a probability distribution over the possible values that the variable of interest could take. This can be interpreted as telling us the conditional probabilities of the possible values for the variable of interest, given the known input data. A decision function takes this probability distribution and applies a decision rule to select an element from the relevant option set. The decision rule may require an explicitly stated utility function, in order to apply a rule such as maximise expected utility, or it may just operate with a threshold rule. This very simple model is illustrated in Figure 1.

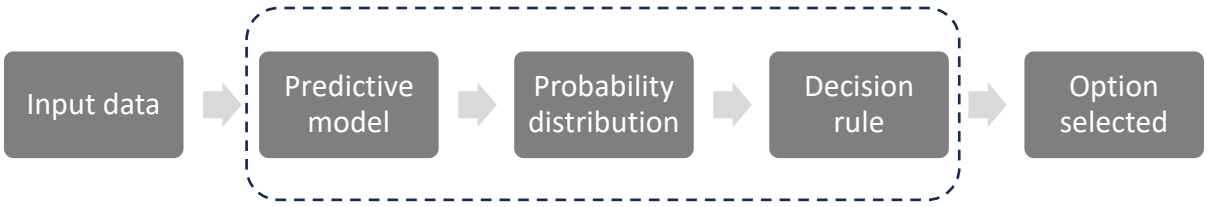

Figure 1. Model of a system with an in-built decision-function.

For example, an AI system may be used to read mammograms as part of a breast cancer screening programme. We can separate this system into a predictive model, which estimates the probability that the patient has breast cancer, and a decision rule, which recommends either "refer to specialist" or "do not refer" on the basis of that probability. Applying this to the PPP, a predictive model would take the relevant demographic information about the patient as input and provide a probability distribution over the possible preference orderings as output. This would tell us a conditional probability for each of the relevant possible preference orderings over treatment options that the patient could have, given the stated demographic information.[3] A decision rule would take this probability distribution and recommend a single treatment option to be pursued. For example, we could employ a threshold decision rule stating that the patient should be given any treatment option that has a probability greater than 0.5 of being what the patient would prefer to receive.

The point made by Beigang is that the conceptual distinction between predictive models and decision functions is crucial for the ethical evaluation of AI systems, because the sorts of ethical considerations that bear on predictions are different from those that bear on the decisions that are ultimately made.[1] However, we could go further than this and design systems that are purely predictive, without any in-built decision function. For example, the mammogram reading system could simply state the probability of breast cancer, and leave the decision of whether or not to refer to a specialist up to the doctor and patient. Birch et al., for example, argue that AI tools in healthcare should either take into account information about specific patients' values and risk attitudes, or should do without in-built decision functions altogether and just produce probabilistic

[3] An interesting further question, which cannot be explored in detail here, is: what interpretation of probability makes the most sense in this context? This is one instance of a more general question about whether the probabilities employed by different types of AI systems should be interpreted as more like credences, frequencies, propensities, or something else. In the case of preference prediction, it seems natural to think of the probabilities as expressing either frequencies in a reference population, or degrees of evidential support, but either view would require a full defence.

predictions that can inform human decision-making.[2] This alternative model is illustrated in Figure 2.

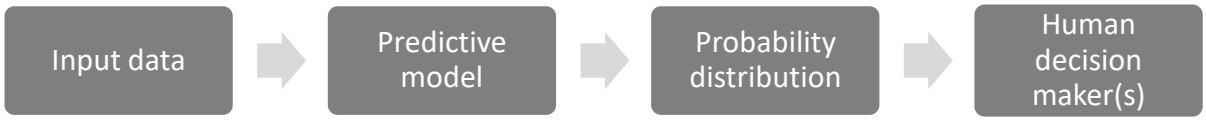

Figure 2. Model of a system without an in-built decision function.

What I propose here is that the PPP should be thought of as a purely predictive model, without any in-built decision function. That is to say, it should be designed to do what it says on the tin: predict, not decide. This probabilistic prediction can then be incorporated into a broader decision-making process, which may be sensitive to a wider range of relevant considerations.

This idea is not completely absent from the existing literature on PPPs. Rid and Wendler note that their proposed PPP should be thought of as only making a prediction, and that this does not settle the question of how such a prediction should be used in decision-making.[29] Rid and Wendler [29] and Jardas, Wasserman, and Wendler [32] mention a number of different models of decision-making informed by the prediction of a PPP, while Ferrario et al. provide a more detailed exploration of some ways in which such a prediction might be incorporated into clinical shared decision-making.[31] However, in their critiques of the PPP, some authors implicitly assume that it would deliver decisions, rather than predictions. Accordingly, adopting the purely predictive conception of the PPP provides a way of avoiding the objections presented in Section 3, to which we shall now return our attention.

5 Avoiding the Objections

5.1 Bare Statistical Evidence Revisited

To address this objection, we first need to clarify an ambiguity in the legal scholarship on the problem of bare statistical evidence. There are two versions of this problem, which arise from two distinct legal concepts: the admissibility and sufficiency of evidence. Admissibility is a matter of whether a type or specific piece of evidence should be allowed to be considered by a judge or jury. For example, witness testimony given as a result of torture, and physical evidence obtained by unlawful seizure are generally considered inadmissible. Sufficiency, on the other hand, is a matter of whether a piece of evidence is enough on its own to settle a particular legal question. If that piece of evidence were the only known fact relevant to the question, could the court reach the certitude required to answer it one way or the other? When we plug in these different legal concepts, we get different versions of the problem of bare statistical evidence: one says that bare statistical evidence ought not to be admissible, while the other says that it ought not to be considered sufficient. So which is the right way to think of the problem and, consequently, what is the precise nature of the analogous challenge to the PPP?

In the early literature on bare statistical evidence in the law, some authors argue that it should be generally inadmissible.[43-44] That is to say, it should be excluded altogether from legal proceedings. However, taking such a strong line on statistical evidence would rule out many types of evidence that neither should, nor in practice are, considered inadmissible. For example, Lewis Ross shows that DNA evidence is properly understood as a form of statistical evidence and that, while there may be some restrictions on the use of DNA profiling in legal proceedings, this form of evidence should not be ruled generally inadmissible on the basis that it is statistical in nature.[45] The more plausible view is that DNA profiling, along with other forms of statistical evidence, should be admissible (absent independent reasons for inadmissibility), but that it should not be considered sufficient on its own to settle a legal judgement. Accordingly, much of the debate about

bare statistical evidence now concerns when, if ever, it can provide sufficient grounds for a legal decision.[46]

This clarifies the problem for the PPP: if the analogy between healthcare and legal contexts holds, the most we should conclude is that the statistical evidence provided by a PPP is admissible, but not sufficient on its own to settle a choice. If the PPP was making decisions, then those decisions would necessarily be made on the basis of statistical evidence. And if bare statistical evidence ought not to be sufficient to settle a choice, then this would be a problem. However, if we think of the PPP as one source of evidence, as I have suggested, then it is entirely possible for a decision-maker to take other sources of evidence into consideration alongside the output of the PPP, thereby avoiding the problem with the sufficiency of bare statistical evidence. So, the analogy with this well-known problem in legal theory does not provide a general objection to the development and use of the PPP, but rather suggests that it should be thought of as one source of relevant information, which should be integrated with others in order to inform decisions.

A question remains about how to approach choices in which a PPP provides the only evidence about a patient's preferences, if they are completely unable to express their values and preferences, and there are no available family, friends, or advance directives to provide further information. The problem of bare statistical evidence, understood as a challenge to the sufficiency of the PPP for decision-making, surely has traction in these cases. However, these are also the cases in which the potential advantages of the PPP appear to be greatest. This is because the general lack of evidence about patient preferences is likely to compound the problem that the PPP was intended to solve: inaccurate prediction of patient preferences. It is natural to think that less evidence will lead to less accurate predictions, and studies bear this out: physicians' predictions alone are less accurate than those informed by evidence from advance directives [47] and family members.[47,48]

Indeed, this has led some authors to claim that cases in which no other evidence is available are precisely those in which the use of the PPP would be *least* problematic.[32,49]

I am inclined to be more cautious here and conclude that, when the PPP is the only available source of evidence, exactly how the pros and cons balance out may ultimately depend on the particular details of any given case. This further strengthens the analogy with the legal evidence. It has been suggested that bare statistical evidence is rightly considered sufficient in the types of legal case in which it has distinctive benefits, because a paucity of non-statistical evidence would otherwise lead to unjust decisions.[45,50] Accordingly, Ross argues that whether or not the use of bare statistical evidence to fill so-called "epistemic gaps" is justified, "*depends on weighing the relevant reasons rather than entertaining any categorical prohibition on statistical evidence*."[50, p.326]

The broader lesson here is that, although the problem of bare statistical evidence does tell us something relevant about the nature of the PPP as a source of evidence, it does not provide a general argument against its use in this way. Rather, it highlights some potential limitations to its use, of which those designing, regulating, and implementing such systems should be aware.

5.2 Advantages of surrogate decision-making revisited

In responding to the family of objections that highlight non-accuracy-based reasons to favour surrogate decision-making, the first thing to note is that these reasons do not hold true of all surrogates. We can divide cases into those in which there is an available surrogate who has the relevant feature (e.g. has entered into a form of shared agency with the patient, has been explicitly nominated as a surrogate by the patient, is uniquely placed to recognise the multifaceted nature of the patient etc.), and those in which there is not.

In the cases in which there is no surrogate available who would realise the relevant advantages, these objections have no force against the use of a PPP. In this context, the accuracy-based reasons in favour of using a PPP stand unopposed by non-accuracy-based reasons to the contrary. It is simply not the case that all people have formed the kind of relationship with their next of kin that generates a form of shared agency, or makes the latter uniquely placed to recognise their multifaceted nature. And it is clearly not true that all people have expressed a wish for some particular person or group of people to make choices on their behalf. So, this family of objections does not provide a general objection to all contexts in which a PPP may be used.

What about cases in which there is a surrogate available? The simple thing to note here is that surrogate decision-making and the evidential use of the PPP that I have suggested are not mutually exclusive. A surrogate decision-maker may take in a range of different pieces of information prior to making a decision, such as the perspective of other friends and family, and healthcare professionals' views about what people tend to prefer in similar situations. There seems to be no reason to think that they cannot also consider a prediction from the PPP as one piece of evidence alongside these others, and still secure the distinctive advantages of surrogate decision-making.

Might the use of the PPP have negative effects on surrogate decision-making, even if only considered as providing a piece of evidence? For example, Ferrario et al.,[31] Rid and Wendler,[28] and Tretter and Samhammer [38] all consider the possibility that surrogate's confidence may be undermined if their view conflicts with a prediction from the PPP, and this might plausibly limit some of the distinctive advantages of surrogate decision-making. This is a worry worth taking seriously, but it is an empirical question that should be investigated through trial and observation. Like any new healthcare technology, it should undergo rigorous testing before being used more widely. If it appears to have downsides in practice, we should attempt to understand the dynamics

that produce them, investigate whether changes could eliminate or mitigate them, and carefully consider how any unavoidable downsides weigh up against any benefits.

5.3 Endorsed reasons revisited

There are two points to be made in response to the objection from endorsed reasons. First, it is plausible that using the PPP might indirectly help surrogates to accurately identify the reasons that a person would endorse in their own deliberation. It can be very difficult to adopt a perspective different from one's own, and information from the PPP might help surrogates to do this. Once they are told the PPP's prediction, a surrogate can draw on their specific knowledge of the patient and undertake a kind of inference to the best explanation, in order to identify reasons that (1) would explain this preference ordering and (2) they think would be endorsed as reasons within the patient's own deliberation. Of course, there is no guarantee that this process would infallibly lead to decisions made solely on the basis of patient-endorsed reasons, but no such guarantee is available for surrogates without a PPP either, and it is hard to see why we should think that the evidence provided by a PPP would make it *harder* to identify such reasons.

Second, one can accept that it is important to act on the basis of endorsed reasons without thereby accepting the view that identifying the option of that an individual would have chosen is of no importance whatsoever. Not all of the reasons that a person endorses support the option for which they have an overall preference. I prefer to have a COVID-19 vaccination than not, due to the reduced risk of catching and spreading the virus, but I do still take the risks and side-effects of vaccination to be relevant considerations (reasons I endorse) in my deliberation. This means that one can make choices on behalf of another person on the basis of reasons that they would endorse, but end up making a choice that is not aligned with their preferences. Without further argument, it is far from obvious to me that this would be better than deliberating on the basis of non-endorsed reasons, but arriving at the option that the patient would actually prefer. And both

possibilities are clearly worse than choosing on the basis of reasons they would endorse *and* selecting the option that they would prefer. I am not claiming to show that accuracy about preferences is more important than acting for endorsed reasons. Rather, I am claiming that even if acting for endorsed reasons matters, so too does accuracy about preferences. And, by hypothesis, the PPP helps with the latter.

To see how the PPP might help surrogates to achieve the goal of attending to both endorsed reasons and overall preferences, consider a case in which a surrogate knows the sorts of reasons that a person would endorse, but does not know how these reasons balance out in a particular choice. For example, one might know that an individual towards the end of life cares about having the chance to live longer and that they care about avoiding time in hospital and invasive medical treatments, while remaining uncertain about whether, in a particular context, this person would prefer to be hospitalised for life-sustaining treatment, or receive palliative care in a hospice. We might think of this as the surrogate knowing that various conditionals are true of the patient, such as, "if they would prefer to be hospitalised, the reason for this is that they want to extend their life" and, "if they prefer not to be hospitalised, the reason for this is that they want to avoid invasive medical treatment." If the PPP can provide a reliable prediction of their preference, then the surrogate can make a decision that is justified in reference to reasons that they would endorse and is aligned with their overall preferences.

The key takeaway from these points is that even if we accept the view that acting on the basis of patient-endorsed reasons is important for proper respect for autonomy, this does not mean that there is no value in the PPP as a source of evidence about patient preferences.

6 Conclusion

The arguments of this paper do not provide a general defence of the PPP. Rather, the conclusions to be drawn are as follows. First and foremost, we should think of the PPP as playing an evidential role to support human decision-making, by making predictions about patient preferences, rather than making decisions about which options to pursue. Second, once we think in this way, we can see that the objections from bare statistical evidence, non-accuracy advantages of surrogates, and endorsed reasons do not undermine the proposal for a PPP.

Lastly, we should refocus our attention away from arguments that concern the PPP as a direct decision-making tool, and towards the considerations that are relevant to the evidential role of the PPP instead. Some of these considerations will be instances of more general questions within the ethics of predictive AI, such as whether certain protected characteristics should be excluded from the input data,[50-52] whether there is a significant degree of differential accuracy across salient groups,[3,52-53] and how these issues relate to existing ideas about the ethics of belief.[4] Some, however, will be more specific to this particular technology. For example, it will be important to investigate what effect this sort of technology may have on surrogate decision makers, and to consider new questions arising from this empirical research, about how to mitigate any evident problems through design, regulation, and implementation. There is plenty more work to be done to determine how the pros and cons of the PPP balance out, but I hope here to have clarified what the target of such work should be.

References

1. Beigang F. On the Advantages of Distinguishing Between Predictive and Allocative Fairness in Algorithmic Decision-Making. *Minds and Machines*. 2022;32(4):655–82, https://doi.org/10.1007/S11023-022-09615-9/FIGURES/3.
2. Birch J, Creel KA, Jha AK, Plutynski A. Clinical decisions using AI must consider patient values. *Nature Medicine*. 2022;28(2):229–32, https://doi.org/10.1038/s41591-021-01624-y.
3. Hedden B. On statistical criteria of algorithmic fairness. *Philosophy & Public Affairs*. 2021;49(2):209–31, https://doi.org/10.1111/papa.12189.
4. Lazar S, Stone J. On the Site of Predictive Justice. *Noûs*. 2023;1–25, https://doi.org/10.1111/nous.12477.
5. Corbett-Davies S, Gaebler JD, Nilforoshan H, Shroff R, Goel S. The Measure and Mismeasure of Fairness. *The Journal of Machine Learning Research*. 2024;24(1):14730–846, https://doi.org/10.48550/arXiv.1808.00023.
6. Spalding R, Edelstein B. Exploring variables related to medical surrogate decision-making accuracy during the COVID-19 pandemic. *Patient Education and Counseling*. 2022;105(2):311–21, https://doi.org/10.1016/J.PEC.2021.06.011.
7. Spalding R. Accuracy in Surrogate End-of-Life Medical Decision-Making: A Critical Review. *Applied Psychology: Health and Well-Being*. 2021;13(1):3–33, https://doi.org/10.1111/APHW.12221.
8. Batteux E, Ferguson E, Tunney RJ. A mixed methods investigation of end-of-life surrogate decisions among older adults. *BMC Palliative Care*. 2020;19(1):1–12, https://doi.org/10.1186/S12904-020-00553-W/TABLES/2.
9. Bryant J, Skolarus LE, Smith B, Adelman EE, Meurer WJ. The accuracy of surrogate decision makers: Informed consent in hypothetical acute stroke scenarios. *BMC Emergency Medicine*. 2013;13(1):1–6, https://doi.org/10.1186/1471-227X-13-18/TABLES/2.

10. Marks M, Arkes H. Patient and surrogate disagreement in end-of-life decisions: can surrogates accurately predict patients' preferences? *Medical Decision Making*. 2008;28(4):524–31, https://doi.org/10.1177/0272989X08315244.
11. Shalowitz D, Garrett-Mayer E, Wendler D. The accuracy of surrogate decision makers: a systematic review. *Archives of Internal Medicine*. 2006;166(5):493–7, https://doi.org/10.1001/ARCHINTE.166.5.493.
12. Coppolino M, Ackerson L. Do Surrogate Decision Makers Provide Accurate Consent for Intensive Care Research? *Chest*. 2001;119(2):603–12, https://doi.org/10.1378/CHEST.119.2.603.
13. Suhl J, Simons P, Reedy T, Garrick T. Myth of Substituted Judgment: Surrogate Decision Making Regarding Life Support Is Unreliable. *Archives of Internal Medicine*. 1994;154(1):90–6, https://doi.org/10.1001/ARCHINTE.1994.00420010122014.
14. Placido D, Yuan B, Hjaltelin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nature Medicine*. 2023;29(5):1113–22, https://doi.org/10.1038/s41591-023-02332-5.
15. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. *Nature Medicine*. 2020;26(8):1229–34, https://doi.org/10.1038/s41591-020-0942-0.
16. Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine*. 2019;25(10):1519–25, https://doi.org/10.1038/s41591-019-0583-3.
17. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature Medicine*. 2019;25(7):1054–6, https://doi.org/10.1038/s41591-019-0462-y.
18. Huang P, Lin CT, Li Y, Tammemagi MC, Brock M V, Atkar-Khattra S, et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a

deep learning method. *Lancet Digit Health*. 2019;1(7):e353–62, https://doi.org/10.1016/S2589-7500(19)30159-1.

19. Chiarito M, Luceri L, Oliva A, Stefanini G, Condorelli G. Artificial Intelligence and Cardiovascular Risk Prediction: All That Glitters is not Gold. *European Cardiology Review*. 2022;17, https://doi.org/10.15420/ecr.2022.11
20. Cai Y, Cai YQ, Tang LY, Wang YH, Gong M, Jing TC, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review. *BMC Medicine*. 2024;22(1):56, https://doi.org/10.1186/s12916-024-03273-7.
21. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*. 2019;25(1):65–9, https://doi.org/10.1038/s41591-018-0268-3.
22. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*. 2019;394(10201):861–7, https://doi.org/10.1016/S0140-6736(19)31721-0.
23. Baeßler B, Götz M, Antoniades C, Heidenreich JF, Leiner T, Beer M. Artificial intelligence in coronary computed tomography angiography: Demands and solutions from a clinical perspective. *Frontiers in Cardiovascular Medicine*. 2023;10, https://doi.org/10.3389/fcvm.2023.1120361.
24. Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z, Prunotto M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digital Medicine*. 2019;2(1):92, https://doi.org/10.1038/s41746-019-0172-3.
25. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in

Eye Clinics: A Multicentre Randomized Controlled Trial. *eClinicalMedicine*. 2019;9:52–9, https://doi.org/10.1016/j.eclinm.2019.03.001.

26. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22):2402, https://doi.org/10.1001/jama.2016.17216.
27. Rid A, Wendler D. Can We Improve Treatment Decision-Making for Incapacitated Patients? *Hastings Center Report*. 2010;40(5):36–45, https://doi.org/10.1353/hcr.2010.0001.
28. Rid A, Wendler D. Treatment Decision Making for Incapacitated Patients: Is Development and Use of a Patient Preference Predictor Feasible? *Journal of Medicine and Philosophy*. 2014;39(2):130–52, https://doi.org/10.1093/jmp/jhu006.
29. Rid A, Wendler D. Use of a Patient Preference Predictor to Help Make Medical Decisions for Incapacitated Patients. *Journal of Medicine and Philosophy*. 2014;39(2):104–29, https://doi.org/10.1093/JMP/JHU001.
30. Wendler D, Wesley B, Pavlick M, Rid A. A new method for making treatment decisions for incapacitated patients: what do patients think about the use of a patient preference predictor? *Journal of Medical Ethics*. 2016;42(4):235–41, https://doi.org/10.1136/MEDETHICS-2015-103001.
31. Ferrario A, Gloeckler S, Biller-Andorno N. Ethics of the algorithmic prediction of goal of care preferences: from theory to practice. *Journal of Medical Ethics*. 2023;49(3):165–74, https://doi.org/10.1136/jme-2022-108371.
32. Jardas E, Wasserman D, Wendler D. Autonomy-based criticisms of the patient preference predictor. *Journal of Medical Ethics*. 2022 Dec;48:304–10, https://doi.org/10.1136/medethics-2021-107629.

33. Sharadin NP. Patient preference predictors and the problem of naked statistical evidence. *Journal of Medical Ethics*. 2018;44(12):857–62, https://doi.org/10.1136/medethics-2017-104509.

34. Um S. Autonomy, shared agency and prediction. *Journal of Medical Ethics*. 2022;48(5):313–4, https://doi.org/10.1136/medethics-2022-108289.

35. Berger JT, DeRenzo EG, Schwartz J. Surrogate Decision Making: Reconciling Ethical Theory and Clinical Practice. *Annals of Internal Medicine*. 2008;149(1):48, https://doi.org/10.7326/0003-4819-149-1-200807010-00010.

36. Mappes TA, Zembaty JS. Patient Choices, Family Interests, and Physician Obligations. *Kennedy Institute of Ethics Journal*. 1994;4(1):27–46, https://doi.org/10.1353/ken.0.0065.

37. Hardwig J. What about the Family? *Hastings Center Report*. 1990;20(2):5, https://doi.org/10.2307/3562603.

38. Tretter M, Samhammer D. For the sake of multifacetedness. Why artificial intelligence patient preference prediction systems shouldn't be for next of kin. *Journal of Medical Ethics*. 2023;49(3):175–6, https://doi.org/10.1136/jme-2022-108775.

39. Kim SYH. Improving Medical Decisions for Incapacitated Persons: Does Focusing on "Accurate Predictions" Lead to an Inaccurate Picture? *Journal of Medicine and Philosophy*. 2014;39(2):187–95, https://doi.org/10.1093/jmp/jhu010.

40. Brock DW. What Is the Moral Authority of Family Members to Act as Surrogates for Incompetent Patients? *The Milbank Quarterly*. 1996;74(4):599, https://doi.org/10.2307/3350394.

41. John S. Patient Preference Predictors, Apt Categorization, and Respect for Autonomy. *Journal of Medicine and Philosophy*. 2014;39(2):169–77, https://doi.org/10.1093/JMP/JHU008.

42. John S. Messy autonomy: Commentary on Patient preference predictors and the problem of naked statistical evidence. *Journal of Medical Ethics*. 2018;44(12):864–864, https://doi.org/10.1136/medethics-2018-104941.

43. Blome-Tillmann M. 'More Likely Than Not' - Knowledge First and the Role of Bare Statistical Evidence in Courts of Law. In: Carter A, Gordon E, Jarvis B, editors. *Knowledge First - Approaches in Epistemology and Mind*. Oxford: Oxford University Press; 2017. p. 278–92, https://doi.org/10.1093/oso/9780198716310.003.0014.

44. Enoch D, Spectre L, Fisher T. Statistical Evidence, Sensitivity, and the Legal Value of Knowledge. *Philosophy & Public Affairs*. 2012;40(3):197–224, https://doi.org/10.1111/papa.12000.

45. Ross L. Rehabilitating Statistical Evidence. *Philosophy and Phenomenological Research*. 2021;102(1):3–23, https://doi.org/10.1111/phpr.12622.

46. Ross L. Recent work on the proof paradox. *Philosophy Compass*. 2020;15(6):1–11, https://doi.org/10.1111/phc3.12667.

47. Coppola KM, Ditto PH, Danks JH, Smucker WD. Accuracy of Primary Care and Hospital-Based Physicians' Predictions of Elderly Outpatients' Treatment Preferences With and Without Advance Directives. *Archives of Internal Medicine*. 2001;161(3):431-440, https://doi.org/10.1001/archinte.161.3.431.

48. Uhlmann RF, Pearlman RA, Cain KC. Physicians' and Spouses' Predictions of Elderly Patients' Resuscitation Preferences. *Journal of Gerontology*. 1988;43(5):M115–21, https://doi.org/10.1093/geronj/43.5.M115.

49. Varma S, Wendler D. Medical Decision Making for Patients Without Surrogates. *Archives of Internal Medicine*. 2007;167(16):1711, https://doi.org/10.1001/archinte.167.16.1711.

50. Ross L. Justice in epistemic gaps: The 'proof paradox' revisited. *Philosophical Issues*. 2021;31(1):315–33, https://doi.org/10.1111/phis.12193.

51. Johnson GM. Algorithmic bias: on the implicit biases of social technology. *Synthese*. 2021;198(10):9941–61, https://doi.org/10.1007/s11229-020-02696-y.

52. Fazelpour S, Danks D. Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*. 2021;16(8), https://doi.org/10.1111/phc3.12760.

53. Beigang F. Reconciling Algorithmic Fairness Criteria. *Philosophy & Public Affairs*. 2023;51(2):166–90, https://doi.org/10.1111/papa.12233.