AMA Journal of Ethics®

September 2018, Volume 20, Number 9: E902-910

VIEWPOINT: PEER-REVIEWED ARTICLE

Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm

Camillo Lamanna, MMathPhil, MBBS and Lauren Byrne, MBBS

To claim one AMA PRA Category 1 Credit™ *for the CME activity associated with this article, you must do the following: (1) read this article in its entirety, (2) answer at least 80% of the quiz questions correctly, and (3) complete an evaluation. The quiz, evaluation, and form for claiming* AMA PRA Category 1 Credit™ *are available through the* AMA Education Center.

Abstract

A significant proportion of elderly and psychiatric patients do not have the capacity to make health care decisions. We suggest that machine learning technologies could be harnessed to integrate data mined from electronic health records (EHRs) and social media in order to estimate the confidence of the prediction that a patient would consent to a given treatment. We call this process, which takes data about patients as input and derives a confidence estimate for a particular patient's predicted health care-related decision as an output, the *autonomy algorithm*. We suggest that the proposed algorithm would result in more accurate predictions than existing methods, which are resource intensive and consider only small patient cohorts. This algorithm could become a valuable tool in medical decision-making processes, augmenting the capacity of all people to make health care decisions in difficult situations.

The Case for an AI-Assisted Autonomy Algorithm

In this article, we argue that artificial intelligence (AI) can be used to mine data from electronic health records (EHRs) and social media in order to predict an incapacitated person's preferences regarding health care decisions. The argument proceeds in three steps.

We first show that a significant proportion of patients do not have the capacity to make health care decisions and motivate the search for a reliable mechanism to predict patient preferences. We describe the triple burden that incapacity creates: the ethical burden upon health care systems to respect the wishes of these patients; the emotional burden upon surrogates to make difficult decisions; and the economic burden upon society to fund investigations and treatments that the incapacitated patient would have declined.

The second part of the argument concerns existing tools to identify patient preferences. We discuss the literature on identifying patient factors that predict patient treatment preferences and then suggest that AI will lead to a step change in our power to predict these preferences. We sketch how existing AI technologies could integrate data mined from EHRs and social media in order to estimate the confidence of the prediction that a patient would consent to a given treatment. We call this computational process—which takes data about patients as input and derives a confidence estimate for a particular patient's predicted health care-related decision as an output—the *autonomy algorithm*.

In the third section, we consider some ethical issues raised by this approach. First, an autonomy algorithm must be interpreted with caution: simply because we can be confident that a person would choose treatment X, it does not follow that this person *should* choose X. The second point is more hypothetical: if increasingly massive data sets enable the autonomy algorithm to offer very high levels of predictive accuracy, should AI replace human decision makers, regardless of a patient's decision-making capacity?

It is concluded that an AI-assisted autonomy algorithm, if thoughtfully implemented and judiciously used, could offer some relief from the aforementioned triple burden posed by incapacitated patients: it could lead to improved respect for autonomy, reduced burnout of surrogates, and economic gains for society. However, we must tread carefully in the implementation of the proposed technology and remember that algorithms function as decision aids, not dictates.

Decision-Making Capacity and Surrogate Decision Making

Decision-making capacity consists of the ability to understand the information related to a decision, to appreciate its significance, to reason about the costs and benefits of different courses of action, and to communicate the decision one has made. Although thinkers use terms such as "understand," "appreciate," and "reason" in a variety of ways, in broad terms this is the definition accepted by the medical community.[1]

Incapacity is no small problem: estimates suggest that more than one-third of elderly and psychiatric hospital inpatients lack decision-making capacity.[2,3] Moreover, in one study, health care professionals failed to identify incapacity in 42% of cases.[4] When clinicians do correctly identify a patient without decision-making capacity, the evidence suggests that they often fail to match their treatment plan to the patient's preferences.[5] Reasons for this disconnect are multifactorial and include clinicians' difficulty in synthesizing information about the patient and cognitive biases at work in the hospital environment.[6,7]

Making life-and-death decisions for incapacitated patients takes a considerable toll upon clinicians, as studies indicate an association between end-of-life decision-making and health care professional burnout.[8,9] Involving family members or patient surrogates in

the decision-making process, however, is no panacea. Surrogates predict patients' preferences incorrectly in roughly one-third of cases, typically projecting their own wishes onto the patient concerned.[10,11] Moreover, many surrogates experience subsequent stress and mental health problems, with the effects sometimes persisting for years.[12] One proposed solution to this problem is the advance directive or advance care plan. The ethical and practical issues with these tools have been discussed elsewhere; for the purposes of this paper, we consider only patients who have not indicated advance preferences for their care.

One corollary of the difficulty in predicting an incapacitated patient's preferences is overtreatment. Every day, patients without decision-making capacity are subjected to investigations and treatments to which they would not have consented. Indeed, unnecessary investigations and treatments are not only ethically troubling but also place undue economic strain upon already-stretched health care systems.[13]

We suggest that just as AI algorithms enable online vendors to predict which products a customer is most likely to buy or which films they are most likely to enjoy, so AI could be harnessed to predict which health care choices a patient would make.

Using Data to Make Predictions

According to some studies, using only the base rate (ie, the proportion of *all* patients favoring treatment X over treatment Y) to predict a given patient's preferences is as accurate as using a surrogate.[14-16] Provided there are data sets that contain the relevant information, it follows that creating a patient preference predictor that is more accurate than a surrogate would require minimal fine tuning.[17,18]

One area that has been well researched is the treatment choices made by patients with localized prostate cancer. In particular, it has been shown that younger patients tend to prefer more aggressive treatment,[19] a finding echoed by other studies on preferences for surgery.[20,21] Furthermore, men who are married are more likely to opt for aggressive treatment,[22] and those who are more prone to risk taking prefer a watch-and-wait approach.[23] Thus, for this example, one could create a regression model that takes age and marital status as input variables and yields a probability that a given patient would opt for surgery. As surrogates are no more accurate than the base rate (ie, population) preference in predicting a given patient's preference, a model that is trained on a data set that contains the two additional features of age and marital status will likely be more accurate than surrogates. However, deriving such a model for treatment preferences for localized prostate cancer requires significant time, manpower (eg, investigators, data collectors, and statisticians), and funding because potential determinants of preference (eg, age, marital status) need to be identified, health records need to be read and coded, and statistical analyses need to be performed. Perhaps more importantly, traditional regression analysis only allows for a handful of preselected determinants to be analyzed

in one study. As a result, important predictors may be overlooked if researchers do not expect them: regression can only predict treatment choices based on the input variables given.

We propose that AI would be able to revolutionize both the availability and accuracy of predictions regarding health care decisions. Two strong assumptions, however, are required: AI must have access to population-wide electronic health records (EHRs) and these EHRs must be interpretable by AI.

Suppose a clinician wants to know if a patient would wish to undergo risky surgery that might restore his or her power of speech, which was lost due to brain cancer. A machine learning algorithm would be trained on the EHRs of patients who faced a decision about a similarly risky surgery for brain cancer but, due to the location of their tumor, were able to communicate their decision. The input vector, therefore, would include demographic indicators (eg, age, marital status, ethnicity) as well as detailed information from the EHR regarding prior health care consultations, treatments, side effects, investigations, previously expressed preferences and desires, and antecedent choices in other health-related decisions. The output would be a probability estimate that the patient would choose to have surgery.

In this way, algorithmic analyses of EHRs would be able to perform a predictive function similar to human-run studies but complete them in a much shorter timeframe, handling much larger sets of observations and analyzing a wider range of predictors.[24] Whereas a human-run study would incur the aforementioned costs for each treatment choice that one wished to predict, an AI algorithm would only need to be developed once: each time it is given a new preference to predict (eg, type of treatment in localized prostate cancer), it uses the same logic to derive its prediction model. Already, by applying machine learning techniques to EHRs, it is possible to predict outcomes after cardiac surgery more accurately than using traditional regression analyses.[25] Accordingly, it seems reasonable to assume that one would see the same increase in accuracy when using machine learning tools to predict patient preferences, provided the relevant data sets exists and are machine-readable.

However, the machine learning approach need not be confined to EHRs. Examining a person's social media profile can already reliably predict his or her religious and political preferences, propensity for risk-taking behavior, and happiness.[26] There is evidence that an algorithm analyzing only Facebook "likes" outperforms spouses in predicting a person's personality traits.[27] Given that personality traits also appear to predict one's preferences regarding end-of-life treatment decisions,[28] it follows that using data from social media in addition to data from EHRs could lead to more precise predictions regarding health care decisions than using data from EHRs alone. Suppose, for example, that machine learning detected a robust connection between "liking" the organization

Death with Dignity National Center and an expressed preference for comfort-focused end-of-life care in the general population. Then, even if an index patient made no explicit statement regarding her end-of-life treatment preferences, if she "liked" Death with Dignity, the probability would increase that she would prefer comfort-based care. What we call the autonomy algorithm takes patients' EHR and social media footprint as input and generates a confidence estimate for a particular patient's predicted treatment preference as an output.

Ethical Issues

There would certainly be benefits to an effective autonomy algorithm. In addition to increased accuracy, a computerized approach could alleviate some of the weight of making life-and-death decisions. An algorithm will not lose sleep if it predicts with a high degree of confidence that a person would wish for a life-support machine to be turned off. The surrogate who ends life-support may rest a little easier knowing that the autonomy algorithm has also concluded that this is likely what the patient would have wanted. Moreover, the autonomy algorithm is truly patient centered. While it can be trained on population-wide data sets, ultimately, it does not receive explicit input from doctors or family members regarding their thoughts on the correct medical decision; rather, it examines data provided by the patient themselves, be it implicitly through the investigations, treatments, diagnoses, and choices recorded on their EHR or more explicitly through social media activity. However, the use of an autonomy algorithm to estimate confidence of predicted treatment decisions raises some practical and ethical questions.

The first question is practical: the use of the aforementioned machine learning tools can simply reflect existing biases. In the research regarding treatment for prostate cancer outlined above, one study found that the most significant predictor of treatment choice was the specialty of the consulting doctor; patients referred to urologists were most likely to choose surgery and those referred to radiation oncologists were most likely to choose radiotherapy.[19] An algorithm trained on this data set would therefore generate a high confidence estimate for the prediction that a patient seeing a urologist would choose surgery. While this might be true, the association (we assume) is not due to genuine patient preference but reflective of the fact that patients are prone to being talked into a certain therapy by their clinician; it would be a bug and not a feature of the proposed autonomy algorithm to reinforce this fact.

Indeed, algorithms can propagate even more insidious associations. Supposing those with lower health literacy are more disposed to choose the (less effective) treatment X, then an algorithm trained on this data set might generate a high confidence estimate for the prediction that a patient with low health literacy would choose X. Of course, this does not mean that patients *should* choose X. There are numerous examples of algorithms in other fields "learning" prejudice; there is no reason to assume health care would be any

different.[29] Therefore, the autonomy algorithm's confidence estimates must be examined critically by patients and health care professionals: incorrectly applied, the autonomy algorithm might just reinforce an undesirable status quo.[30]

This potential for bias leads us to ask to what extent we should be prepared to accept the outputs of the autonomy algorithm. We should recall that surrogates predict preferences of incapacitated patients roughly a third of the time. It would appear reasonable, therefore, to use the output of the autonomy algorithm to help refine one's decision in the context of surrogate decision making. But what if a patient with full capacity was faced with a decision regarding surgery for localized prostate cancer? Health care decisions are inherently stressful and increasingly involve a reasonably sophisticated understanding of probability and uncertainty.[31] It is well known that decision-making processes in these contexts are subject to bias and error.[32] Suppose our hypothetical cancer patient was told that the autonomy algorithm had analyzed the data of millions of patients in similar situations and found that the patients most similar to him opted for a watch-and-wait approach 90% of the time and that, moreover, the rate of decisional regret was higher in the 10% who opted for active treatment. This would be useful, patient-centered information. However, as outlined above, one must guard against unreflexively deferring to the output of the algorithm.

Conclusions

In this essay, we have made the case that it should be possible to construct an autonomy algorithm to estimate confidence for predicted preferences of incapacitated patients by using machine learning technologies to analyze population-wide data sets, including EHRs and social media profiles. The proposed algorithm would result in more accurate predictions than existing methods, which are resource intensive and examine only small patient cohorts.

It was noted that this tool would both help incapacitated patients realize their preferences in spite of being unable to express them and reduce the significant burdens of patients' incapacity by lowering the emotional strain on proxies and reducing the economic costs of unwanted tests and treatments. Moreover, it was suggested that the algorithm could function as a decision aid to patients with decision-making capacity who are facing complex decisions regarding their own health care.

We also highlighted that the proposed autonomy algorithm could potentially propagate established yet erroneous decision-making practices and hence insidiously reinforce health inequalities. In particular, we noted that if less health literate patients typically chose an inferior treatment X in the algorithm's data set, the algorithm would generate a high confidence estimate for the prediction that a less health literate patient would choose the inferior treatment X. If the algorithm was blindly applied with patients automatically opting to choose treatment X, it would strengthen the association

between low health literacy and treatment X in the data set and thereby propagate health inequality. The outputs of the autonomy algorithm need to be carefully interpreted by both clinicians and patients in order to avoid this trap.

In conclusion, we submit that it is the process of making a decision that is humanizing and autonomy affirming. Therefore, it would be dehumanizing to automate this process and defer to algorithmic outputs as a matter of course. Nonetheless, it appears the autonomy algorithm should form part of the decision-making process. If correctly implemented, it would not be liable to the varied biases, projections, and misapprehensions of human decision makers; rather, it would make reliable estimates based on a wealth of real-world data. In this way, the autonomy algorithm could become a valuable tool in the stressful medical decision-making process, augmenting the capacity of all people to make decisions in difficult situations.

References

1. Grisso T, Appelbaum PS, eds. *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. New York, NY: Oxford University Press; 1998.
2. Fitten LJ, Waite MS. Impact of medical hospitalization on treatment decision-making capacity in the elderly. *Arch Intern Med*. 1990;150(8):1717-1721.
3. Lepping P, Stanly T, Turner J. Systematic review on the prevalence of lack of capacity in medical and psychiatric settings. *Clin Med (Lond)*. 2015;15(4):337-343.
4. Sessums LL, Zembrzuska H, Jackson JL. Does this patient have medical decision-making capacity? *JAMA*. 2011;306(4):420-427.
5. Teno JM, Fisher E, Hamel MB, et al. Decision-making and outcomes of prolonged ICU stays in seriously ill patients. *J Am Geriatr Soc*. 2000;48(5)(suppl):S70-S74.
6. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16:138. doi:10.1186/s12911-016-0377-1.
7. Janz NK, Wren PA, Copeland LA, et al. Patient-physician concordance: preferences, perceptions, and factors influencing the breast cancer surgical decision. *J Clin Oncol*. 2004;22(15):3091-3098.
8. Embriaco N, Papazian L, Kentish-Barnes N, Pochard F, Azoulay E. Burnout syndrome among critical care healthcare workers. *Curr Opin Crit Care*.2007;13(5):482-488.
9. Chuang CH, Tseng PC, Lin CY, Lin KH, Chen YY. Burnout in the intensive care unit professionals: a systematic review. *Medicine (Baltimore)*. 2016;95(50):e5629. doi:10.1097/MD.0000000000005629.
10. Shalowitz DI, Garrett-Mayer E, Wendler D. The accuracy of surrogate decision makers: a systematic review. *Arch Intern Med*. 2006;166(5):493-497.

11. Marks MA, Arkes HR. Patient and surrogate disagreement in end-of-life decisions: can surrogates accurately predict patients' preferences? *Med Decis Making*. 2008;28(4):524-531.
12. Wendler D, Rid A. Systematic review: the effect on surrogates of making treatment decisions for others. *Ann Intern Med*. 2011;154(5):336-346.
13. Cardona-Morrell M, Kim JCH, Turner RM, Anstey M, Mitchel IA, Hillman K. Non-beneficial treatments in hospital at the end of life: a systematic review on extent of the problem. *Int J Qual Health Care*. 2016;28(4):456-446.
14. Smucker WD, Houts RM, Danks JH, et al. Modal preferences predict elderly patients' life-sustaining treatment choices as well as patients' chosen surrogates do. *Med Decis Making*. 2000;20(3):271-280.
15. Houts RM, Smucker WD, Jacobson JA, Ditto PH, Danks JH. Predicting elderly outpatients' life-sustaining treatment preferences over time: the majority rules. *Med Decis Making*. 2002;22(1):39-52.
16. Shalowitz DI, Garrett-Mayer E, Wendler D. How should treatment decisions be made for incapacitated patients, and why? *PLoS Med*. 2007;4(3):e35. doi:10.1371/journal.pmed.0040035.
17. Rid A, Wendler D. Treatment decision making for incapacitated patients: is development and use of a patient preference predictor feasible? *J Med Philos*. 2014;39(2):130-152.
18. Rid A, Wendler D. Can we improve treatment decision-making for incapacitated patients? *Hastings Cent Rep*. 2010;40(5):36-45.
19. Sommers BD, Beard CJ, D'Amico AV, Kaplan I, Richie JP, Zeckhauser RJ. Predictors of patient preferences and treatment choices for localized prostate cancer. *Cancer*. 2008;113(8):2058-2067.
20. Kurd MF, Lurie JD, Zhao W, et al. Predictors of treatment choice in lumbar spinal stenosis: a spine patient outcomes research trial study. *Spine (Phila Pa 1976)*. 2012;37(19):1702-1707.
21. Heit M, Rosenquist C, Culligan P, Graham C, Murphy M, Shott S. Predicting treatment choice for patients with pelvic organ prolapse. *Obstet Gynecol*. 2003;101(6):1279-1284.
22. Denberg TD, Beaty BL, Kim FJ, Steiner JF. Marriage and ethnicity predict treatment in localized prostate carcinoma. *Cancer*. 2005;103(9):1819-1825.
23. López-Pérez B, Barnes A, Frosch DL, Hanoch Y. Predicting prostate cancer treatment choices: the role of numeracy, time discounting, and risk attitudes. *J Health Psychol*. 2017;22(6):788-797.
24. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Eng J Med*. 2016;375(13):1216-1219.
25. Allyn J, Allou N, Augustin P, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS One*. 2017;12(1):e0169772. doi:10.1371/journal.pone.0169772.

26. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *PNAS*. 2013;110(15):5802-5805.
27. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *PNAS*. 2015;112(4):1036-1040.
28. Lattie EG, Asvat Y, Shivpuri S, et al. Associations between personality and end-of-life care preferences among men with prostate cancer: a clustering approach. *J Pain Symptom Manage*. 2016;51(1):52-59.
29. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Penguin Random House; 2016.
30. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378(11):981-983.
31. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull*. 2009;135(6):943-973.
32. Kahneman D, Slavic P, Tversky A, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press; 1982.

Camillo Lamanna, MMathPhil, MBBS is an internal medicine physician affiliated with the University of Cape Town in South Africa. His research interests include the ethics of acute and emergency care and the use of emerging technologies in medicine in the developing world.

Lauren Byrne, MBBS is an emergency department physician affiliated with the University of Sydney in Australia. Her professional interests include health care economics with a particular focus on resource allocation in critical care medicine.

Citation
AMA J Ethics. 2018;20(9):E902-910.

DOI
10.1001/amajethics.2018.902.

Conflict of Interest Disclosure
The author(s) had no conflicts of interest to disclose.

The viewpoints expressed in this article are those of the author(s) and do not necessarily reflect the views and policies of the AMA.

ISSN 2376-6980