Published in final edited form as:

Am J Bioeth. 2020 November ; 20(11): 51–53. doi:10.1080/15265161.2020.1820108.

Respect and trustworthiness in the patient-provider-machine relationship: Applying a relational lens to machine learning healthcare applications

Stephanie A Kraft^{1,2}

¹Seattle Children's Research Institute.

²University of Washington School of Medicine.

Healthcare delivery is an interpersonal endeavor. In every clinical interaction, providers have an ethical obligation to show respect to their patients, and ideally over time these interactions lead to mutually respectful and trusting relationships. Examining healthcare through a relational lens recognizes patients and providers as socially embedded beings whose interactions and decisions are informed by their individual perspectives, experiences, and circumstances, as well as the broader systems that shape when, where, and how these interactions occur (Mackenzie & Stoljar 2000). As machine learning applications become increasingly intertwined in healthcare, they have the potential to alter these structures and change the contours of patient-provider relationships. These changes may sometimes be subtle, but their impact on healthcare relationships may nonetheless be significant and ethically relevant.

This commentary will argue that it is necessary to consider explicitly the relational implications of machine learning in healthcare as part of a thorough ethical analysis. Building on the framework that Char and colleagues lay out in their Target Article (2020), which maps key ethical considerations onto the development, implementation, evaluation, and oversight of machine learning healthcare applications, this commentary will examine how these applications can be designed and implemented to uphold and reinforce essential relational values including respect for persons and trustworthiness, with particular attention to any impacts on patient-provider relationships, and identify key questions to add to each stage of Char and colleagues' pipeline model framework.

Applying a relational lens to the pipeline model

Char and colleagues identify multiple important ethical considerations that should be addressed at each stage of development, implementation, evaluation, and oversight of machine learning healthcare applications. Their pipeline model incorporates key values-based questions that highlight considerations including transparency, social justice, informed consent, and accountability. While these questions do not *exclude* consideration of the relational implications of machine learning, they also do not explicitly recognize patients and providers as socially embedded beings nor address how these technologies may affect them from a relational perspective. Bringing questions about relational impact to the forefront at each stage of the pipeline model—conception, development, calibration, initial implementation evaluation, and subsequent evaluations and inspections—will encourage

developers and decision-makers to anticipate and mitigate any relational harms that these applications may create, while enhancing their potential benefits in ways that are most meaningful to patients.

Conception: Are choices about design and outcomes made in a way that respects patients as unique persons, with full consideration of their values and needs?

The Target Article rightly points out that application designers must select outcomes and goals that are relevant to key stakeholders. More specifically, patient values and needs should be a driving force in identifying goals for machine learning applications and determining where and how to invest development resources. One way of conceptualizing respect for persons is that it requires appreciating patients as unique individuals and acting in recognition of the demands that individuality places upon one's behavior (Dickert 2009). While choices regarding machine learning design and development occur on a systems level and may not have an immediate and direct impact on patients, the ethical obligation to act with respect for patients is no less important in this setting, as each choice has the potential to affect the patient experience. Centering patients in decision-making at an early stage—including ensuring the problems being addressed are meaningful and relevant to those the application purports to help—is a critical component of building respect for persons into every decision point.

Development: Do training datasets adequately reflect patients' lived experiences?

The Target Article describes the ways that bias, racism, and other structural injustices are built into many datasets and may be perpetuated or amplified by machine learning applications that are trained on these datasets. Expanding on the social justice considerations the authors identify, developers should ask whether training datasets include data that represent the full scope of the patient experience. That is, do the data that are able to be captured in a health record or other available dataset accurately reflect who a person is and the life they lead? To the extent that important qualitative, nuanced, and sometimes pervasive factors are not or cannot be systematically recorded on an individual basis, developers must find ways to account for these factors and their absence in training data. Any outputs of applications that do not reflect patients' lived experiences must be interpreted with caution, and serious questions raised as to whether they such applications can ethically be implemented.

Calibration: What impact do false positives and negatives have on trust between patients and providers and how can those potential harms be mitigated?

As Char and colleagues note, there is a potential for false positives and negatives from machine learning applications as they are calibrated for use in practice. To fully understand the scope of the potential harms of these errors and the tradeoffs inherent in the calibration process, it is necessary to look closely at the effects of each type of error on patients, providers, and the relationships between them. As a starting point, decision-makers need to understand how patients experience these errors, how patients perceive the responsibility of the application versus the provider, and whether errors cause the patient to question the provider's expertise, judgment, and/or commitment to their care. Further, if some of these errors are inevitable, providers must be prepared to respond appropriately and mitigate harm

to the patient-provider relationship. As tradeoffs are made in this context, it is important to reflect on the far-reaching implications that a seemingly small error may have, particular when it reflects a considered judgment about which harms are acceptable and which are not.

Initial implementation evaluation: Does the use of the application change the power dynamic between patient and provider or have other unintended consequences on the patient-provider relationship?

Error rates notwithstanding, an application may have unintended consequences on patient-provider relationships when implemented in a real world setting. Potential effects on the relationships and power dynamics between patients and providers should be anticipated and evaluated on an ongoing basis, as they are central to the patient's experience of healthcare. One such changing power dynamic could be observed if machine learning alters how sources of knowledge are prioritized; for example, if the output of a machine learning application is viewed as more reliable than a patient's own report, this may shift the already imbalanced patient-provider power dynamic in the clinical setting further away from the patient (Campelia and Feinsinger 2020). This sort of changing dynamic may have long-term consequences on patients' ability to effectively self-advocate and providers' inclination to fully engage with patients' perspectives.

Subsequent evaluations and inspections: What impact will the application have on long-term engagement between patients, providers, and their healthcare institutions?

As an application continues to be used, its impact across the healthcare system must be addressed. Harms that occur on the individual level, in aggregate, reflect upon the trustworthiness of a healthcare institution as a whole. If machine learning is implemented on an institutional basis without centering the principle of respect for persons, it could at best maintain, or at worst exacerbate, already-existing injustices that prevent people from being able to meaningfully engage with their providers and healthcare institutions, and thereby impede some patients' ability to access the care they need (Levesque et al. 2013). Ongoing evaluations of machine learning applications must be attuned to these systems-level considerations to ensure machine learning does not hinder patients' ability to form the mutually respectful, trusting relationships with their providers and institutions that support meaningful access to care over the long term.

Conclusion

Healthcare delivery cannot be divorced from interpersonal relationships, even with the increasing integration of machine learning applications. Therefore, ethical analyses of these applications must include explicit consideration of their relational impact. Incorporating relationship-focused questions like those identified here into the pipeline model laid out in the Target Article can strengthen its value as a framework to guide the ethical implementation of machine learning healthcare applications.

Acknowledgments

This work was supported by the National Human Genome Research Institute at the National Institutes of Health (grant number K01HG010361) and the Clinical Research Scholars Program at Seattle Children's Research Institute.

References

- Campelia GD, and Feinsinger A. 2020 Creating space for feminist ethics in medical school. *HEC Forum* 32(2):111–124, [PubMed: 32221818]
- Char DS, Abramoff MD, and Feudtner C. 2020 Identifying ethical considerations for machine learning healthcare applications. *American Journal of Bioethics* (in press).
- Dickert NW 2009 Re-examining respect for human research participants. *Kennedy Institute of Ethics Journal* 19(4):311–338. [PubMed: 20191947]
- Levesque JF, Harris MF, and Russell G. 2013 Patient-centred access to health care: conceptualising access at the interface of health systems and populations. *International Journal for Equity in Health*. 2013;12:18. [PubMed: 23496984]
- Mackenzie C, and Stoljar N. 2000 Introduction: autonomy refigured In *Relational autonomy: feminist perspectives on autonomy, agency, and the social self*, eds. Mackenzie C and Stoljar N, 3–31. New York: Oxford University Press.