

The near-term ethical challenges of digital twins

In 'Represent me: please! Towards an ethics of digital twins in medicine'¹, Mattias Braun analyses the potential for simulations of human organs and bodies, or "digital twins," to faithfully represent a person. Drawing from several French philosophers, he introduces "first conditions for digital twins to take on an ethically justifiable form of representation".¹

The analysis predominantly focuses on challenges that arise in terms of representation, embodiment, control by the patients after which the twin is modelled. Many challenges are posited on the existence of digital twins as autonomous technologies capable of acting on behalf of the patient. One may be reasonably sceptical about the potential to achieve "digital twins" in the near future; the author, to his credit, acknowledges this fact in the section 'Simulating a Precise Efficacy' while still providing a faithful overview of a nascent field of development. Given the state of development described in the paper, the risk of autonomous digital twins in the near future seems remote.

There are two clear ethical risks of digital twin which, to my reading, deserve more attention than they receive in Braun's analysis. Let us refer to these as the 'near-term' challenges of digital twins, in the sense that their significance does not turn on the degree of agency or embodiment of digital twins, but rather their mere existence and potential lack of control by their 'hosts'.

The first concerns medical paternalism. According to the author, the distinct characteristic of a digital twin compared to 'basic simulations' of individual organs is the shift from static to dynamic simulation powered by a novel feedback loop between the simulation and its object. Of course, any new form of collecting sensitive health data carries risks, but active monitoring is certainly not a novel feature of digital twins. Rather, any risks associated with the new data collection and feedback loops created by digital twins mirror the impact of other novel health technologies, most immediately the health monitors and sensing networks which would, presumably, provide the dynamic health data needed to keep digital twins synchronised with the patient's actual health.²

If treated as a reliable source of information about the current (and future) health of the patient, digital twins will undermine a patient's authority and autonomy in the doctor-patient relationship by providing an alternative, more 'objective' account of the patient's health. Moreover, as digital twins will be powered by AI and inferential analytics,^{1,3} this account will not be limited to the individual patient's data but rather models built from population-level inferences and data. In their predictive capacity twins will seemingly breach the notion of individualised or personalised care by introducing predictions drawn from population-level insights as a reliable indicator of the future course of the patient's health. Each of these effects have the potential to drastically impact the doctor-patient relationship,⁴ and deserve consideration prior to deploying digital twins for individualised health predictions.

All of this, though, misses the core risk posed by digital twins to the health of their 'host'. The decision to use digital twins in the care of patients, and to treat them as a reliable source of information and predictions about the course and complications of health conditions at an individual level, is normatively significant. To use the author's words, the significant risks of digital twins arise from the willingness to treat the simulation "as a surrogate representing the body of the represented person."¹

Twins are designed to predict major health events and recommend major interventions at an early stage. Given the inherent risks of these interventions, the decision to treat digital twins as reliable surrogates immediately raises serious questions of epistemic responsibility for clinicians, data scientists, and model developers.^{5,6} What are, or should be, the standards of model reliability that must be met to act on predictions produced by manipulation of a digital twin? How, if at all, do these

standards change for twins created using population level data and inferences? When is a predictive model sufficiently accurate to justify preventative interventions, such as a mastectomy?⁷ Such standards account for model accuracy and opacity in AI,⁸ safety and reliability standards in medical devices and interventions, and balance interests in patient safety against preventative care. The risk here is of creating a new wave of overdiagnosis in pursuit of personalised care based on 'black box' predictions.

Technical feasibility and reliability are central components of the ethics of digital twins. As with any type of simulation in a high-risk environment, data quality and model accuracy are paramount concerns to ensure predictions are, at a minimum, being made using robust methods on a reliable evidence base. Individual organs and physical systems in the human body are incredibly difficult to simulate reliably as static objects.⁹ Simulating the dependencies between the organs and systems in a human body is exponentially more difficult.

More fundamentally, models and simulations are only as good as the data from which they are built. In the context of healthcare, patients routinely lie or at least fail to disclose all relevant information about their health and health-related behaviours in clinical encounters. Errors and inaccuracies in the data will perpetuate through the system and make predictions highly unreliable. As a result, the foundation of a personalised twin is unreliable.

The obvious 'fix' for this flaw is to only use data from 'objective' or reliable sources; for example, inferring alcohol consumption through data collected via a health monitor. While this may be more plausible in terms of collecting accurate data, its implications in terms of medical paternalism and violation of privacy and data protection legislation merit serious independent discussion.^{2,4} Usage of health monitoring could of course be incentivized through other means, but this does not resolve the inherent risks of collecting health-related data, and the paternalism in care it enables by undermining the patient's credibility as a 'more objective' data source.

Despite these significant challenges facing the development of digital twins, the current discussion of their ethics presumes the existence of reliable simulations updated dynamically and personalised to the organs, multi-organ systems, and or indeed whole body of an individual. I fear we may be discussing the ethical implications of a technology far too early in its development, when its future course, technical feasibility, features, and impact remain highly uncertain, and thus when we are at greatest risk of extrapolating into science fiction.

References

1. Braun, M. Represent me: please! Towards an ethics of digital twins in medicine. *J Med Ethics* (2021) doi:10.1136/medethics-2020-106134.
2. Mittelstadt, B., Fairweather, N. B., Shaw, M. & McBride, N. The Ethical Implications of Personal Health Monitoring. *International Journal of Technoethics* **5**, 37–60 (2014).
3. Wachter, S. & Mittelstadt, B. D. A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. *Columbia Business Law Review* **2019**, (2019).
4. Pellegrino, E. D. & Thomasma, D. C. *The virtues in medical practice*. (Oxford University Press, 1993).
5. Feynman, R. 'Cargo Cult Science' - by Richard Feynman. http://neurotheory.columbia.edu/~ken/cargo_cult.html (1974).
6. Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data & Society* **3**, (2016).

7. Welch, H. G. & Black, W. C. Overdiagnosis in cancer. *Journal of the National Cancer Institute* **102**, 605–613 (2010).
8. Burrell, J. How the Machine ‘Thinks:’ Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* (2016) doi:10.1177/2053951715622512.
9. Margaritis, K. N. & Black, R. A. Modelling the lymphatic system: challenges and opportunities. *Journal of the Royal Society Interface* **9**, 601–612 (2012).