nature machine intelligence

PERSPECTIVE

https://doi.org/10.1038/s42256-021-00373-4

Machine learning and algorithmic fairness in public and population health

Vishwali Mhasawade[1], Yuan Zhao[2] and Rumi Chunara[1,3] ✉

Until now, much of the work on machine learning and health has focused on processes inside the hospital or clinic. However, this represents only a narrow set of tasks and challenges related to health; there is greater potential for impact by leveraging machine learning in health tasks more broadly. In this Perspective we aim to highlight potential opportunities and challenges for machine learning within a holistic view of health and its influences. To do so, we build on research in population and public health that focuses on the mechanisms between different cultural, social and environmental factors and their effect on the health of individuals and communities. We present a brief introduction to research in these fields, data sources and types of tasks, and use these to identify settings where machine learning is relevant and can contribute to new knowledge. Given the key foci of health equity and disparities within public and population health, we juxtapose these topics with the machine learning subfield of algorithmic fairness to highlight specific opportunities where machine learning, public and population health may synergize to achieve health equity.

Decades of work in population and public health has informed research and practice that aim to understand what makes and keeps people and populations healthy[1]. The major underpinning principle is that of health equity, defined as "minimizing avoidable disparities in health and its determinants—including but not limited to health care—between groups of people who have different levels of underlying social advantage or privilege, that is, different levels of power, wealth, or prestige due to their positions in society relative to other groups"[2]. Inequality at a societal level is itself harmful across the population as a whole[3]. To capture the complex interplay between individual, community and other structural factors such as racism, which affect and are leverage points for health, the social ecological model was developed[4]. This framework outlines how the health of an individual is affected by multiple factors operating at different levels in a hierarchy (Fig. 1). Indeed, an understanding of and focus on the macro-level properties (for example, levels beyond the individual in Fig. 1) is critical to put individuals in the best context to leverage interventions, without increasing disparities[5]. A focus on determinants, antecedents and other factors related to health outside the hospital are imperative to not only address specific challenges for high-risk individuals, but also determine what policies would benefit communities as a whole.

An understanding of the importance of macro-level properties can be useful, for example, in examining the impact of introducing healthier food options in a neighbourhood to help people in a community improve their diet[6] or how maternal health policies impact child mortality[7]. These are pressing challenges; in the United States social determinants (broadly defined as the conditions in which people are born, grow, live, work and age)[8] account for 25–60% of deaths in any given year according to results from meta-analyses[9]. Moreover, 80% of the growing burden of non-communicable diseases worldwide could be prevented through modifying behaviours such as reducing tobacco, alcohol, fat and/or salt consumption, promoting physical activity and improving environmental conditions such as air quality and urban planning[10].

In the past decade, the development of statistical and machine learning approaches with a focus on clinical tasks, such as predicting disease prognosis and identifying phenotypes, has greatly matured with some demonstrations of benefit to patients[11,12]. Echoing research in population and public health, the recent COVID-19 pandemic has highlighted how multi-sectoral factors outside of the clinic such as community, social networks and environment are also critical with respect to health[13–19]. Accordingly, this Perspective illustrates where and how machine learning has been shown to be useful in a holistic set of tasks related to health. We summarize existing data and methods used in public and population health, and use this synthesis to present directions for future work to leverage synergies in machine learning and population and public health.

Data in public health

Before elaborating on machine learning efforts in public and population health and gaps, an outline of the types of data that are commonly used is pertinent. Data commonly used in public health can be broadly categorized into (1) surveys conducted by public health and governmental organizations that aggregate individual-level information and (2) person-generated data that provide information at a finer resolution[20]. These types of data each provide complementary types of information relevant to public health; each also associated with their own challenges. In the following subsections we outline each of these data and associated methodological challenges with respect to their use in models of health.

Data from surveys and government reports. Traditional approaches to data collection in public health involve the aggregation of data via local officials and channels. When health providers report notifiable diseases on a case-by-case basis, it is known as passive surveillance; often useful during disease outbreaks and to gain a baseline view of disease burden in a specific location. Conversely, active surveillance is when a health department proactively contacts health care providers to request information about diseases, which

[1]Department of Computer Science and Engineering, Tandon School of Engineering, New York University, New York, NY, USA. [2]Department of Epidemiology, School of Global Public Health, New York University, New York, NY, USA. [3]Department of Biostatistics, School of Global Public Health, New York University, New York, NY, USA. ✉e-mail: rumi.chunara@nyu.edu

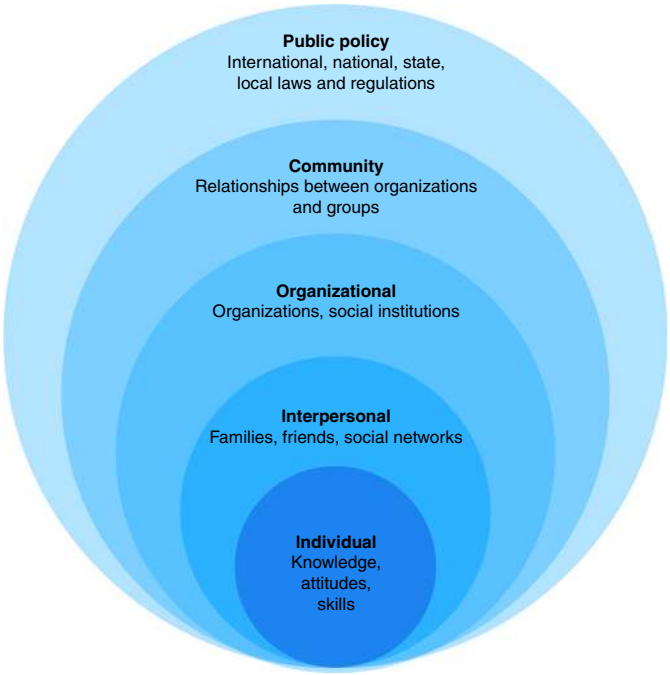

Fig. 1 | The socio-ecological model of health. This model (adapted from Bronfenbrenner et al.[4]) can be used to understand and identify leverage points for the multifaceted and interactive effects of the multi-level individual and environmental factors that determine health. Macro-level properties (those above the individual level) are also key to understanding inequities and interventions that can reduce inequity.

can be catered to specific needs and easier to validate, but more labour intensive. This surveillance information is then forwarded to national health ministries and international organizations like the World Health Organization. Such collection procedures result in robust, denominator-based public data that are often made publicly available by public health and governmental organizations. Examples include the National Health and Nutrition Examination Survey (NHANES) or the Demographic and Health Surveys programme (DHS). These types of data system also aim to capture different indicators of health, designed via specific constructs. For example, housing quality can be measured via rental status, sanitation status, crowding, indoor air quality and so on[21]. Despite the robust nature of the constructs and denominator-based data collection processes[22], loss of information at an individual level due to aggregation, privacy concerns and temporal delays in the process are challenges in the use of data for all situations—such as the need for rapid policy-making, which was exemplified during the COVID-19 pandemic.

Person-generated data. The rising ubiquity of technologies such as smartphones and physical activity trackers, as well as data from social media sites such as Twitter and Instagram and Internet surveys, have made it possible to access high-resolution and geo-linked data in near-real time. The nature of this data can help evade recall or information biases associated with traditional surveys. The often-linked geographic information and time stamps further enable the capture of hyper-local, daily and sub-daily health-related information from behaviours to exposures and other macro-level properties, as well as health outcomes[23,24]. Accordingly, such ubiquitous technologies can provide opportunities to better measure the social determinants of health in a targeted way, by person, location and/or time[21–25]. These attributes of person-generated data can complement denominator-based survey and report data that are not available at such high granularity. However, the opt-in nature of and resources required for the tools used to produce person-generated data (that is, an individual chooses to use a certain app, and there can be cost associated with access) alongside the unstructured nature of the data (not in the form of specific measures or constructs from the outset) bring new computational challenges to the fore if we intend to use the data to make inferences within and across populations. These challenges have been outlined in detail previously[22] and are used to motivate discussion in subsequent sections.

Machine learning in population and public health

Building on the focus of public and population health, here we outline a taxonomy to organize and illustrate machine learning efforts linked to priorities of these fields. We use this summary and juxtaposition with current research to identify gaps in the application of machine learning in public and population health.

- **Identification of factors and their relation to health outcomes**. Learning what contributes, at all levels of the socio-ecological framework and their interactions, to health outcomes is a significant part of public and population health research. Machine learning has so far played a role in identification in a broad range of studies from learning biological mechanisms[26] to establishing the multivariate empirical relationship between the probability of disease outbreak and environmental conditions[27]. Given the complex relationships and possible mediations between multi-level factors[28], by leveraging new data sources there is the opportunity to use and develop machine learning methods for the interpretable identification and assessment of the source and magnitude of a wide variety of multi-level factors in health outcomes.
- **Design of interventions**. Besides multi-level factors related to disease, a socio-ecological framework also indicates the utility of interventions at multiple levels. Besides exacerbating inequities, targeting individuals directly can be highly stigmatizing, aggravating health-related behaviours that may be the target of intervention[29]. Thus, while a large body of work in machine learning has focused on targeting the individual, for example towards depression management[30], self-efficacy for weight loss[31], smoking cessation[32] and personalized nutrition based on glycaemic response[33], the possibility of leveraging data and machine learning in efforts that consider the multi-level nature of influence around an individual is an open investigation area. For example, group-based intervention programmes are one of the means to reduce substance abuse by reinforcing positive behaviour. System dynamics modelling approaches or agent-based models (which involve simulation of the history, location in space and time, and interaction between individual agents) can be useful to design and evaluate the effectiveness of interventions at multiple levels[34,35].
- **Prediction of outcomes**. Predicting mortality risk [36], hospital readmission[37] and disease prognoses from pathology, imaging or other clinical data[38] are well-studied tasks using probabilistic machine learning and deep learning methods. Prediction has also been leveraged for population-level questions, largely in geographic disease risk mapping[39]. Conversely, mitigating health disparities and the prediction of outside-hospital events are crucial challenges that have received less attention from the machine learning community. Although new data provide the potential to incorporate social, environmental and other multi-level determinants in (and thus improve) prediction models, there is a need to expand on research, which may use similar methods to the vast preponderance of research on clinical prediction models to incorporate these new data sources. Such data can also better capture factors that are represented by proxies such as race at present[40,41]. Besides the addition of data to

prediction models, a critical focus on the types of prediction task and how they are used in practice is also essential. For example, although the incorporation of patient socio-economics can improve risk assessment, and epidemiological evidence shows the relation of these to several important outcomes[42], concerns regarding their use being utilized to justify lower standards of care for poor patients[43] have been voiced. Such important risks illustrate the importance of the development and use of prediction models that are closely linked to clinical and public health practices and priorities. In the United States, for example, accountable health initiatives and tax codes have encouraged the activation of risk mitigation actions beyond primary care that could leverage risk information from social factors, such as through improving health care teams' ability to understand the 'upstream' factors impacting their patients' health and the ability to act on care recommendations, informing clinical care decisions[44].

- **Allocation of resources**. The use of machine learning and artificial intelligence for resource allocation has been promoted in several types of health care and public health tasks, typically with a focus on allowing for estimation under uncertainty, computing treatment effects or alternate scenarios to aid in decision-making, augmenting decision rule approaches by incorporating more information and handling missing data. In health care, this has been applied at the person level (for example, for learning personalized management and treatment plans by modelling the temporal evolution of patient data[45]). Machine learning approaches have also been incorporated in resource allocation for measurement; for example deciding what laboratory measurements or psychosocial measures from mobile phones should be measured, when and on whom, trading off the value of information against the cost of acquisition[46,47]. With a public and population health lens, the propagation and enhancement of disparities in resource allocation should also be considered via a multi-level perspective. Examples of such efforts to account for and address inequity could be the development of methods that include relevant information beyond the individual level, such as the patient's geographic location, in the optimization procedure[48].

Current challenges

From the elemental look at the fields of population and public health and current data and machine learning efforts, we now shift our attention to future enterprise. What are the significant challenges that must be surmounted and for which there is room for machine learning innovation? We consider challenges across data, problem selection and formulation.

Privacy and health data. In light of a multi-level perspective, it should first be clarified that privacy can be viewed at collective and individual levels. This is important because collective rights are not necessarily a large-scale representation of individual rights and related issues[49]. At a collective level, for example, people may wish to avoid being stigmatized via certain assessments. At an individual level, people can also be concerned about their own privacy. The COVID-19 pandemic brightly illuminated and surfaced these issues and several trade-offs specific to person-generated data and public health based on the rapid use of data to inform policy efforts and models during the crisis. While focused literature has comprehensively summarized challenges, proposed recommendations and discussed these in detail[50,51], we summarize main concepts here.

To make the use of person-generated data feasible, and more broadly any digital data in public health research and practice efforts at scale, regulatory frameworks for appropriate guidance for the health sector and other end-users have been emphasized.

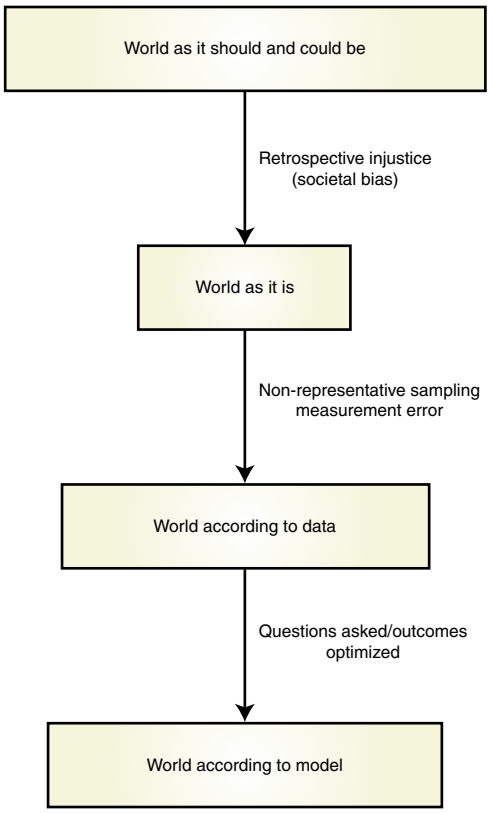

Fig. 2 | Illustration of sources of bias at different stages of data and algorithm use. The figure, adapted from Mitchell et al.[100], shows where disparities can manifest in the process of generating and using data. Where machine learning is applied (what questions are asked) can impact disparities and equity, and is added to the pipeline. Algorithmic fairness comes into play at the end of this pipeline (via questions asked/outcomes optimized).

Regulation has an important role; indeed, current procedural mechanisms are lacking in their uniformity, which is needed to promote predictability and trust in the public at individual and collective levels. Regulatory guidance, drawn from human rights legislation, must be supported by broad audit and enforcement powers[52]. Beyond this, institutions that use data should foster robust data stewardship and standardize their practices to international best practices (such as ensuring the team developing the framework is appropriately diverse, in technical specialization as well as in terms of demographics and connections with appropriate communities). The latest recommendations on data privacy foster a dynamic approach, open to iteration. Approaches must also go beyond anonymization, which in itself may decrease utility and remains vulnerable to de-anonymization via aggregation of disparate pieces of information[53,54].

A comprehensive approach to privacy is also imperative in lieu of relying only on regulation, which can be slow to change, or be a source of structural discrimination[55]. Research has also shown that within current regulatory frameworks it is common that individuals who share data publicly may not be cognizant that their data may be used by external parties for public health monitoring or research purposes[56]. Thus a discussion on data privacy is inextricable from efforts to empower the public to make their decisions and weigh in on what risks are appropriate for them. Best practices in health communication can be leveraged for this purpose. Data sharing can be a form of public health intervention and can increase the likelihood of using the data further and translating results into action.

Reporting exposure data back to study participants is increasingly critical and can increase self-efficacy, particularly when working with underserved communities[57]. Personal actions as well as attitudes and trust of the use of data are essential in response to public health crises[58]. Accordingly, there is room for the development of such approaches in effective communication of data, which could entail summarizing and communicating one's own data in a manner that is interpretable, and also aggregating or contrasting it with data from others in a privacy-preserving manner. Machine learning is starting to be used to automate patient education and information communication efforts[59]. Such approaches can be informed by the health communication literature, which shows that instead of treating health literacy as a patient problem that needs to be fixed or circumvented, health literacy interventions can integrate the principles of socio-ecology to develop interventions that build capacity and empower individuals and communities via factors significant to them[60,61].

Assessing external validity. External validity, the validity of applying the conclusions of a scientific study outside the context of that study is an important consideration in all statistical and machine learning in health endeavours. Given the comprehensive consideration of multi-level attributes in population and public health, spanning populations and their context, external validity brings new considerations for data and algorithms. Whereas large denominator-based survey and government reports typically used in public health efforts aim to provide the information to mitigate these challenges (that is, by including a representative population, or at least information about which group is represented), the localized information offered through person-generated data comes with several external validity challenges. First, the data often do not come with linked information regarding attributes of the persons sharing the data (for example gender, age and so on), making it difficult to understand who has contributed the data[22]. Second, as the data from such sources are not organized into specific constructs (that is, the free-form text or images have to be processed to form features), variables from one dataset or environment may not be comparable to another. Finally, given the opt-in nature of (and resources required for) the tools used to produce person-generated data, measurement of an outcome may be affected by selection bias. Accordingly, it is important to understand the factors that lead to the data being contributed; developing algorithms on data in a new context can result in biased results if the mechanisms by which the algorithm worked in the original population are not well understood. For example, in one location people may tend to share more information due to different social norms[23].

There are several machine learning avenues for increasing standardization across environments (where an environment is a data source, hospital, location and so on) or for analysing data from multiple environments. Data augmentation is an approach to fill gaps in non-representative samples[62]. Domain adaptation methods can be developed to address distribution shifts that may occur across different environment, both based on different populations and data generation mechanisms.[25,63,64] In particular, thinking about data distribution shifts and differences from a causal perspective has been utilized to inform the empirical learning processes [23,65].

Measuring and integrating social determinants of health. New data provide a needed way to capture data from daily life outside of the hospital, yet still critical for a comprehensive understanding of health. However, social determinants encompass a broader set of information than just that at the individual level. There is a need for better measures of upstream factors that are important social determinants. This spans environmental, policy and other social factors such as racism[55]. Some work leveraging new data and machine learning methods has focused on using natural language processing, computer vision or other approaches to identify and generate relevant features for a specific task, which can be leveraged to generate features of the built and social environments from unstructured data in scalable ways (that is, for more environments and communities)[66–69]. Second, although social determinants have been studied extensively in the epidemiology literature, the findings from these studies have underscored the need for methods that can better capture flexible and complex relationships between social determinants and health outcomes[70,71]. This need also indicates an opportunity for machine learning; the flexibility of modern machine learning methods may help us model these relationships[72,73]. Third, the use of social variables in causal models is often restricted, under the premise that they are non-manipulable or not intervenable[74]. A causal perspective (for example, by the use of directed acyclic graphs) is important to systematically assess the role of different variables to model them in a manner that will also enable identification of where interventions can and should occur. At the same time, causal methods often assume stable unit treatment value, which implies that there is no interference and only one version of treatment; this is often non-tractable with the complex nature of social determinants and other methods must be investigated. In general, a full specification of social determinants and the pathways by which they operate is important. This requires identification of the variables and mediator or moderator effects[75]. An understanding of distal and proximal determinants and their relations is also relevant to be able to examine and consider the effect of different forms of interventions and to identify and work towards structural changes at the root cause of disparities[55].

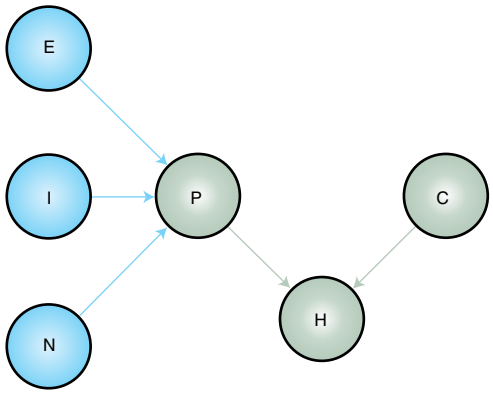

Fig. 3 | Abstract illustration of challenge in algorithmic fairness due to unexplained variance or proxy variables. Consider a given *P* that is composed of several factors *E*, *I* and *N*. *H* is to be predicted using *C* and *P* while ensuring that the model prediction is fair with respect to *P*. However, *E*, *I* and *N* (shaded blue) are typically not accounted for, though doing so would better represent variance in *P*.

Health disparities. The description and explanation of racial and ethnic health disparities, which are differences in health status or in the distribution of health resources between different population groups, arising from the social conditions in which people are born, grow, live, work and age are major focuses of public and population health research. These disparities often manifest in specific gender, income level and race/ethnicity groups experiencing greater health risks[2]. Algorithmic fairness has recently emerged as a field of machine learning, with the goal of mitigating differences in machine learning outcomes across social groups. Broadly, algorithmic fairness approaches have consisted of statistical notions that ensure some form of parity for members of different protected groups (for example by race, gender and so on) and individual notions that aim to ensure that people who are 'similar' with respect to the classification

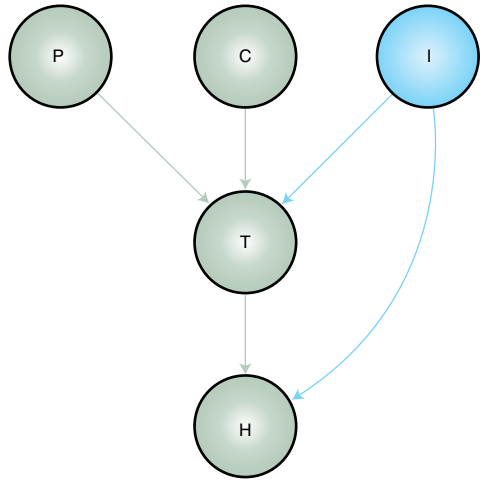

Fig. 4 | Abstract illustration of challenge in algorithmic fairness owing to who the data represents. A given *P* and *C*, which is used to determine the treatment *T*, are ultimately used to assess *H*. If high-risk populations such as the uninsured are not included (shaded blue), because *I* affects *T*, inferred relationships and treatment effects would not be relevant to those most vulnerable. Furthermore, if such populations continue to be excluded from algorithmic efforts, overall disparities can increase.

task receive similar outcomes. We refer the reader to detailed summaries of algorithmic fairness elsewhere[76].

Although it is becoming clear that algorithms are an important place to search for bias, as algorithms become embedded in many societal efforts, algorithms can incorporate and augment existing biases through many means, not only the outcomes they optimize for, which are a main focus of algorithmic fairness efforts (Fig. 2). Indeed, one risk is that clinicians and others may trust that issues of bias are sufficiently managed via algorithmic fairness efforts[77]. Accordingly, here we place algorithmic fairness in relation to the broader literature in health disparities within public and population health to identify challenges and opportunities in algorithmic fairness work with respect to advancing health.

Health disparities are reflective of social oppression and its influence on the health of individuals that identify with such marginalized communities. It is essential to identify what leads to disparate health outcomes to design interventions to mitigate disparities and improve the health of high-risk populations. This involves multiple types of task, such as measuring health outcomes[78,79] and disparities across social groups[80–82], as well as designing policies to mitigate the disparities[82]. Figure 2 provides a framework for analysing how societal bias can result in biased predictions and where algorithmic fairness contributes (bottom two boxes in Fig. 2). Essentially, data are always sourced via some perspective[83]; an important consideration in any use of data to understand disparities. Issues of data representation in training data have also been well documented in terms of their importance in biased algorithm outcomes[84]. Linked to this, how algorithms make inference from underrepresented features[85,86] can also contribute to biased outcomes and disparate performance. In the following subsections, current gaps in machine learning and algorithmic fairness work with respect to health disparities are made explicit, and recommendations are provided for ensuring that health equity remains an inherent goal in the design of machine learning algorithms in health settings.

Algorithmic design. Obermeyer and Mullainathan[79] recently presented a commonly used clinical risk score that considered financial cost expenditure as a proxy for health care needs. Owing to unequal access to care, less money is spent on care for Black patients compared with white patients, and thus although health care cost appeared to be an effective proxy for health by some measures of predictive accuracy, large racial biases resulted. As such cost-based proxy objectives are not uncommon, an essential step is to be aware of task goals and outcomes and interrogate them with respect to health equity. It should be noted that several of the constructs considered in today's algorithmic fairness measures are socially determined (for example race, gender) and thus consideration of them without broader attention to the systemic processes involved in their determination shifts focus away from the root causes of inequity[55]. Historically, algorithmic fairness has not accounted for the complex causal relationships between the biological, environmental and social factors that give rise to differences in medical conditions across protected identities. These missing factors can result in misalignment in equity and algorithmic fairness notions as described above[77]. Moreover, the deployment of algorithms could also perpetuate or augment disparities even with 'algorithmically fair' efforts[87]. In the following sections, challenges related to the use of such variables (as well as those related to algorithmic fairness that are exterior to the algorithm) are elaborated.

Pitfalls with 'proxies' in modelling social variables. Although accounting for factors such as 'race' may be important in specific analyses, it is often unknown what the comprising factors of such social constructs are, how they interact and how to model them[88]. Indeed, poor representation of variations within and between groups, along with the difficulty in attaining the appropriate factors, have led to the use of social constructs such as race as proxies for unknown and/or unmeasured biological and social factors (including racism). A recent study highlighted several clinical risk estimation tools across cardiology, nephrology, obstetrics and many other specialties that all use race, and how this use of a simple race variable severely compromises the health of marginalized individuals[41]. Accordingly, how relevant biologic variation is to be assessed and reported without stratifying populations based on factors such as race and ethnicity is still a challenge to be addressed[89]. Overall, the use of variables such as race, even when considering them as markers to be fair with respect to, such as in algorithmic fairness efforts, obscures variation within and between individuals, impeding equity goals.

Multiple axes of disparities and intersectionality. Health disparities are often measured by considering averages across individuals who identify with a certain attribute, such as a race category. However, measuring health disparities as averages by category does not fully represent or describe the multifaceted and interwoven effects of the forces shaping disparities. Moreover, gradients within specific categories are also neglected. For example, disparities have continued across income groups of childbearing women even after the introduction of policies to improve the health of pregnant women in California[2]. Indeed, lived experiences are frequently the product of intersecting patterns of social forces such as racism and sexism[90], and modelling them as simple additive or multiplicative effects will not capture the full complexity. Research in public health has aimed to address this challenge of modelling the non-additive and dynamic nature of multiple social factors beyond simple interaction terms[91]. Recently, a statistical multi-level method for capturing social factors and their intersections, known as multi-level analysis of individual heterogeneity and discrimination accuracy has been described[92]. This method involves decomposing total variance into (1) between-strata variance, which allows the identification and assessment of disadvantaged groups, and (2) within-strata variance, which allows the identification of individuals within a social group that are at added disadvantage compared with other members of the group. The approach presents several advantages over fixed-effect models that include interaction terms for multiple sensitive attributes, including restricting parameter growth to linear (as opposed

to geometric) forms and adjusting for the sample sizes of the intersectionalities. Overall, the dynamic nature of intersectionality highlights the need for the development of new machine learning approaches that can handle multiple protected attributes and their intersections, including at higher dimensions.

Algorithmic fairness in deployment and health disparity dissonances. Going beyond the data used and methods employed, there can be further equity impediments based on the types of questions that are the focus of machine learning efforts. Indeed, as work interrogating the limits of algorithmic fairness has discussed, the possible veneer of technical neutrality may break down once the full systems the technology is embedded in are considered[93]. In particular, although algorithms themselves may be deemed 'fair', they can result in unfairness when considered in the context of the systems they are deployed in. We illustrate such dissonances via two scenarios.

First, we highlight unfairness that can result from the use of protected attributes that are proxies. Consider the scenario represented in Fig. 3; predicting a health outcome H (for example, risk of cardiovascular disease) based on individual-level information including the perceived protected attribute, P, and clinical conditions, C. Even if successful in ensuring that the risk is fair using a group metric such as demographic parity (equal decision rates across groups regardless of outcome)[94] with respect to racial identity as the protected group attribute, by only considering individual-level attributes, we are still left with unexplained variance for aspects that race can be acting as a proxy for, such as education, E, income levels, I, and neighbourhood characteristics N[95]. Indeed, equal risk scores for Black and non-Black patients would not eliminate disparities across lower-income Black patients versus higher-income Black patients. In health efforts, considering sensitive variables at macro[96] and individual levels simultaneously is one route to a more holistic consideration of fairness[97].

Another challenge in applying algorithmic fairness efforts that can exacerbate disparities is posed by the population to which the focus of algorithmic development and fairness efforts are devoted. Figure 4 represents a health care example wherein individual-level factors are included in an algorithm used to make a decision on a treatment, T, for patients based on C. An algorithm ensuring that model outcomes are fair with respect to P could still perpetuate health inequities in the population if insurance status, represented by I, and necessarily populations for which this varies are not accounted for (because insurance status can be related to health outcomes[98]). A continued focus on efforts that improve care for only the top tier of patients will advance disparities and can be detrimental to all. Machine learning approaches such as transportation of causal effects and domain adaptation may be used to focus questions and studies on improving efforts for neglected populations[99]. At minimum, population representation (who is being included in fairness efforts) should be identified and tallied to draw focus to the development of innovations for underrepresented groups.

Conclusions

Through a discussion of the principles of population and public health, as well as current machine learning efforts, we synthesize and summarize areas where machine learning innovation may synergize with, advance and build on research and practice in these fields. We distil major areas of challenge spanning the data used, methods developed and questions asked, which are all important in equity considerations. These challenges include: the relevance of multi-level factors in health, privacy considerations with relation to health data, external validity concerns specific to public health data and questions, as well as the measurement of and inclusion of social determinants in causal models. We identify how algorithmic fairness efforts must be considered in the context of the data and systems in which they are applied, showing how they could otherwise perpetuate or advance health disparities. Speaking of health equity when only referring to clinical decision-making and fair AI in health care limits equity considerations with respect to health, ignoring the multi-level influences on (and possible interventions in) our health. These principles to shape data, measures and questions of machine learning efforts in health should be leveraged from domains such as public and population health, in which the study of inequity and health is rooted. This grounding is desirable as we work towards improving health for all populations amidst shifting climates, priorities and data. In sum, this Perspective aims to open the imaginary and activate the machine learning community regarding the types of data and questions we consider when thinking about machine learning and AI in health.

Received: 21 October 2020; Accepted: 21 June 2021;
Published online: 29 July 2021

References

1. Rose, G. Sick individuals and sick populations. *Int. J. Epidemiol.* **14**, 427–432 (1985).
2. Braveman, P. Health disparities and health equity: concepts and measurement. *Annu. Rev. Public Health* **27**, 167–194 (2006).
3. Woolf, S. H., Johnson, R. E., Fryer Jr, G. E., Rust, G. & Satcher, D. The health impact of resolving racial disparities: an analysis of US mortality data. *Am. J. Public Health* **94**, 2078–2081 (2004).
4. Bronfenbrenner, U. Toward an experimental ecology of human development. *Am. Psychol.* **32**, 513 (1977).
5. Veinot, T. C., Mitchell, H. & Ancker, J. S. Good intentions are not enough: how informatics interventions can worsen inequality. *J. Am. Med. Inform. Assoc.* **25**, 1080–1088 (2018).
6. Barrientos-Gutierrez, T. et al. Neighborhood physical environment and changes in body mass index: results from the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* **186**, 1237–1245 (2017).
7. Creanga, A. A. et al. Maternal mortality and morbidity in the United States: where are we now? *J. Women's Health* **23**, 3–9 (2014).
8. *Social Determinants of Health* (WHO Regional Office for South-East Asia, 2008).
9. Heiman, H. J. & Artiga, S. Beyond health care: the role of social determinants in promoting health and health equity. *Health* **20**, 1–10 (2015).
10. *2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases: Prevent and Control Cardiovascular Diseases, Cancers, Chronic Respiratory Diseases and Diabetes* (World Health Organization, 2009).
11. Saria, S., Rajani, A. K., Gould, J., Koller, D. & Penn, A. A. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci. Transl. Med.* **2**, 48ra65 (2010).
12. Sweatt, A. J. et al. Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ. Res.* **124**, 904–919 (2019).
13. Gatto, M. et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10484–10491 (2020).
14. Smit, A. J. et al. Winter is coming: a southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19. *Int. J. Environ. Res. Public Health* **17**, 5634 (2020).
15. Sajadi, M. M. et al. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). *JAMA Netw. Open* **3**, e2011834 (2020).
16. Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko, J. & Riazi, S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine* **25**, 100464 (2020).
17. Bann, D. et al. Changes in the behavioural determinants of health during the COVID-19 pandemic: gender, socioeconomic and ethnic inequalities in five British cohort studies. *J. Epidemiol. Commun. Health* https://doi.org/10.1136/jech-2020-215664 (2021).
18. Laurencin, C. T. & McClinton, A. The COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. *J. Racial Ethnic Health Dispar.* **7**, 398–402 (2020).
19. Abedi, V. et al. Racial, economic, and health inequality and COVID-19 infection in the United States. *J. Racial Ethnic Health Dispar.* **8**, 732–742 (2021).
20. Chunara, R., Smolinski, M. S. & Brownstein, J. S. Why we need crowdsourced data in infectious disease surveillance. *Curr. Infect. Dis. Rep.* **15**, 316–319 (2013).
21. Kusnoor, S. V. et al. Collection of social determinants of health in the community clinic setting: a cross-sectional study. *BMC Public Health* **18**, 550 (2018).

22. Chunara, R., Wisk, L. E. & Weitzman, E. R. Denominator issues for personally generated data in population health monitoring. *Am. J. Prevent. Med.* **52**, 549–553 (2017).
23. Mhasawade, V., Elghafari, A., Duncan, D. T. & Chunara, R. Role of the built and online social environments on expression of dining on instagram. *Int. J. Environ. Res. Public Health* **17**, 735 (2020).
24. Zhan, A. et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. *JAMA Neurol.* **75**, 876–880 (2018).
25. Mhasawade, V., Rehman, N. A. & Chunara, R. Population-aware hierarchical Bayesian domain adaptation via multi-component invariant learning. In *Proc. ACM Conference on Health, Inference, and Learning* 182–192 (ACM, 2020).
26. Burgess, S., Foley, C. N. & Zuber, V. Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu. Rev. Genom. Hum. Genet.* **19**, 303–327 (2018).
27. Bhatt, S. et al. The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
28. Zhao, Y. et al. Machine learning for integrating social determinants in cardiovascular disease prediction models: a systematic review. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.11.20192989 (2020).
29. Goldberg, D. S. Social justice, health inequalities and methodological individualism in US health promotion. *Public Health Ethics* **5**, 104–115 (2012).
30. Burns, M. N. et al. Harnessing context sensing to develop a mobile intervention for depression. *J. Med. Internet Res.* **13**, e55 (2011).
31. Manuvinakurike, R., Velicer, W. F. & Bickmore, T. W. Automated indexing of internet stories for health behavior change: weight loss attitude pilot study. *J. Med. Internet Res.* **16**, e285 (2014).
32. Ahsan, G. T. et al. Toward an mhealth intervention for smoking cessation. In *Proc. 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops* 345–350 (IEEE, 2013).
33. Triantafyllidis, A. K. & Tsanas, A. Applications of machine learning in real-life digital health interventions: review of the literature. *J. Med. Internet Res.* **21**, e12286 (2019).
34. Mahamoud, A., Roche, B. & Homer, J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. *Soc. Sci. Med.* **93**, 247–255 (2013).
35. Kouser, H. N., Barnard-Mayers, R. & Murray, E. Complex systems models for causal inference in social epidemiology. *J. Epidemiol. Commun. Health* **75**, 702–708 (2021).
36. Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. *npj Digit. Med.* **1**, 18 (2018).
37. Shameer, K. et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai heart failure cohort. In *Pacific Symposium on Biocomputing 2017* 276–287 (World Scientific, 2017).
38. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
39. Bhatt, S. et al. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *J. R. Soc. Interface* **14**, 20170520 (2017).
40. Galiatsatos, P. et al. The association between neighborhood socioeconomic disadvantage and readmissions for patients hospitalized with sepsis. In *C94: The Impact of Social Determinants in Pulmonary and Critical Care* A5569 (American Thoracic Society, 2019).
41. Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight-reconsidering the use of race correction in clinical algorithms. *N. Engl. J. Med.* **383**, 874–882 (2020).
42. Hamad, R., Nguyen, T. T., Bhattacharya, J., Glymour, M. M. & Rehkopf, D. H. Educational attainment and cardiovascular disease in the united states: a quasi-experimental instrumental variables analysis. *PLoS Med.* **16**, e1002834 (2019).
43. Bynum, J. & Lewis, V. Value-based payments and inaccurate risk adjustment-who is harmed? *JAMA Intern. Med.* **178**, 1507–1508 (2018).
44. Alley, D. E., Asomugha, C. N., Conway, P. H. & Sanghavi, D. M. et al. Accountable health communities-addressing social needs through medicare and medicaid. *N. Engl. J. Med* **374**, 8–11 (2016).
45. Alaa, A. M. & van der Schaar, M. in *Advances in Neural Information Processing Systems* Vol. 30 (eds Guyon, I. et al.) (NeurIPS, 2017).
46. Chang, C.-H., Mai, M. & Goldenberg, A. Dynamic measurement scheduling for event forecasting using deep RL. In *International Conference on Machine Learning* 951–960 (PMLR, 2019).
47. Coughlin, L. N. et al. Developing an adaptive mobile intervention to address risky substance use among adolescents and emerging adults: usability study. *JMIR mHealth uHealth* **9**, e24424 (2021).
48. Snyder, J. J. et al. Organ distribution without geographic boundaries: a possible framework for organ allocation. *Am. J. Transplant.* **18**, 2635–2640 (2018).
49. Mantelero, A. in *Group Privacy* 139–158 (Springer, 2017).
50. Gasser, U., Ienca, M., Scheibner, J., Sleigh, J. & Vayena, E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. *Lancet Digit. Health* **2**, e425–e434 (2020).
51. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **1**, 389–399 (2019).
52. *Privacy and the COVID-19 Outbreak* (Office of the Privacy Commissioner of Canada, 2020); https://priv.gc.ca/en/privacy-topics/health-genetic-and-other-body-information/health-emergencies/gd_covid_202003/
53. Langarizadeh, M., Orooji, A., Sheikhtaheri, A. & Hayn, D. Effectiveness of anonymization methods in preserving patients' privacy: a systematic literature review. *eHealth* 80–87 (2018).
54. Smith, M., Szongott, C., Henne, B. & Von Voigt, G. Big data privacy issues in public social media. In *Proc. 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)* 1–6 (IEEE, 2012).
55. Yearby, R. Structural racism and health disparities: reconfiguring the social determinants of health framework to include the root cause. *J. Law Med. Ethics* **48**, 518–526 (2020).
56. Fiesler, C. & Proferes, N. 'Participant' perceptions of Twitter research ethics. *Soc. Media Soc.* **4**, 2056305118763366 (2018).
57. Sandhaus, S., Kaufmann, D. & Ramirez-Andreotta, M. Public participation, trust and data sharing: gardens as hubs for citizen science and environmental health literacy efforts. *Int. J. Sci. Educ. B* **9**, 54–71 (2019).
58. Chunara, R. & Cook, S. H. Using digital data to protect and promote the most vulnerable in the fight against COVID-19. *Front. Public Health* **8**, 296 (2020).
59. Liu, X., Zhang, B., Susarla, A. & Padman, R. Youtube for patient education: a deep learning approach for understanding medical knowledge from user-generated videos. Preprint at https://arxiv.org/abs/1807.03179 (2018).
60. Dawkins-Moultin, L., McDonald, A. & McKyer, L. Integrating the principles of socioecology and critical pedagogy for health promotion health literacy interventions. *J. Health Commun.* **21**, 30–35 (2016).
61. Hong, S. J., Drake, B., Goodman, M. & Kaphingst, K. A. Race, trust in doctors, privacy concerns, and consent preferences for biobanks. *Health Commun.* **35**, 1219–1228 (2020).
62. Tanner, M. A. & Wong, W. H. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* **82**, 528–540 (1987).
63. Daughton, A. R., Chunara, R. & Paul, M. J. Comparison of social media, syndromic surveillance, and microbiologic acute respiratory infection data: observational study. *JMIR Public Health Surveill.* **6**, e14986 (2020).
64. Sun, B., Feng, J. & Saenko, K. Return of frustratingly easy domain adaptation. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 30 (AAAI, 2016).
65. Pearl, J. & Bareinboim, E. Transportability of causal and statistical relations: a formal approach. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 25 (AAAI, 2011).
66. Scepanovic, S., Martin-Lopez, E., Quercia, D. & Baykaner, K. Extracting medical entities from social media. In *Proc. ACM Conference on Health, Inference, and Learning* 170–181 (ACM, 2020).
67. Abdur Rehman, N., Saif, U. & Chunara, R. Deep landscape features for improving vector-borne disease prediction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops* 44–51 (IEEE, 2019).
68. Relia, K., Akbari, M., Duncan, D. & Chunara, R. Socio-spatial self-organizing maps: using social media to assess relevant geographies for exposure to social processes. *Proc. ACM Hum.Comput. Interact.* **2**, 1–23 (2018).
69. Relia, K., Li, Z., Cook, S. H. & Chunara, R. Race, ethnicity and national origin-based discrimination in social media and hate crimes across 100 US cities. In *Proc. International AAAI Conference on Web and Social Media* Vol. 13, 417–427 (AAAI, 2019).
70. Harper, S., Lynch, J. & Smith, G. D. Social determinants and the decline of cardiovascular diseases: understanding the links. *Annu. Rev. Public Health* **32**, 39–69 (2011).
71. Marmot, M. Social justice, epidemiology and health inequalities. *Eur. J. Epidemiol.* **32**, 537–546 (2017).
72. Akbar, M. & Chunara, R. Using contextual information to improve blood glucose prediction. In *Proc. Machine Learning Research* Vol. 106, 91–108 (PMLR, 2019); http://proceedings.mlr.press/v106/akbar19a.html
73. Quisel, T., Kale, D. C. & Foschini, L. Intra-day activity better predicts chronic conditions. Preprint at https://arxiv.org/abs/1612.01200 (2016).
74. Glymour, C. & Glymour, M. R. Commentary: race and sex are causes. *Epidemiology* **25**, 488–490 (2014).
75. Bauman, A. E., Sallis, J. F., Dzewaltowski, D. A. & Owen, N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *Am. J. Prevent. Med.* **23**, 5–14 (2002).
76. Verma, S. & Rubin, J. Fairness definitions explained. In *2018 IEEE/ACM International Workshop on Software Fairness (Fairware)* 1–7 (IEEE, 2018).

77. McCradden, M. D., Joshi, S., Mazwi, M. & Anderson, J. A. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit. Health* **2**, e221–e223 (2020).
78. Chen, I. Y., Agrawal, M., Horng, S. & Sontag, D. Robustly extracting medical knowledge from EHRS: a case study of learning a health knowledge graph. In *Pacific Symposium on Biocomputing 2020* 19–30 (World Scientific, 2020).
79. Obermeyer, Z. & Mullainathan, S. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proc. Conference on Fairness, Accountability and Transparency* 89 (ACM, 2019).
80. Braveman, P. A., Egerter, S. A., Cubbin, C. & Marchi, K. S. An approach to studying social disparities in health and health care. *Am. J. Public Health* **94**, 2139–2148 (2004).
81. Penman-Aguilar, A. et al. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. *J. Public Health Manag. Pract.* **22**, S33 (2016).
82. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med.* **169**, 866–872 (2018).
83. Tichenor, M. & Sridhar, D. Metric partnerships: global burden of disease estimates within the World Bank, the World Health Organisation and the Institute for Health Metrics and Evaluation. *Wellcome Open Res.* **4**, 35 (2019).
84. Buolamwini, J. & Gebru, T. Gender shades: intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* 77–91 (PMLR, 2018).
85. Agarwal, C. & Hooker, S. Estimating example difficulty using variance of gradients. Preprint at https://arxiv.org/abs/2008.11600 (2020).
86. Hooker, S., Moorosi, N., Clark, G., Bengio, S. & Denton, E. Characterising bias in compressed models. Preprint at https://arxiv.org/abs/2010.03058 (2020).
87. Suresh, H. & Guttag, J. V. A framework for understanding unintended consequences of machine learning. Preprint at https://arxiv.org/abs/1901.10002 (2019).
88. Krieger, N. Refiguring 'race': epidemiology, racialized biology, and biological expressions of race relations. *Int. J. Health Serv.* **30**, 211–216 (2000).
89. Bonham, V. L., Green, E. D. & Pérez-Stable, E. J. Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA* **320**, 1533–1534 (2018).
90. Crenshaw, K. Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *Univ. Chicago Legal Forum* 139–167 (1989).
91. Morris, J. N. Uses of epidemiology. *Br. Med. J.* **2**, 395 (1955).
92. Evans, C. R., Williams, D. R., Onnela, J.-P. & Subramanian, S. A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Soc. Sci. Med.* **203**, 64–73 (2018).
93. Benjamin, R. *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity Press, 2019).
94. Mitchell, S., Potash, E., Barocas, S., D'Amour, A. & Lum, K. Algorithmic fairness: choices, assumptions, and definitions. *Annu. Rev. Stat. Appl.* **8**, 141–163 (2021).
95. VanderWeele, T. J. & Robinson, W. R. On causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology* **25**, 473 (2014).
96. Diez-Roux, A. V. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am. J. Public Health* **88**, 216–222 (1998).
97. Mhasawade, V. & Chunara, R. Causal multi-level fairness. Preprint at https://arxiv.org/abs/2010.07343 (2020).
98. Card, D. E. et al. *The Impact of Health Insurance Status on Treatment Intensity and Health Outcomes* (RAND, 2007).
99. Pearl, J. & Bareinboim, E. External validity: from do-calculus to transportability across populations. *Stat. Sci.* **29**, 579–595 (2014).
100. Mitchell, S., Potash, E., Barocas, S., D'Amour, A. & Lum, K. Prediction-based decisions and fairness: a catalogue of choices, assumptions, and definitions. Preprint at https://arxiv.org/abs/1811.07867 (2018).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to R.C.

Peer review information *Nature Machine Intelligence* thanks Melissa Mccradden, Marcello Ienca and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.