OPEN FORUM

Will Big Data and personalized medicine do the gender dimension justice?

Antonio Carnevale[1] · Emanuela A. Tangari[1] · Andrea Iannone[1] · Elena Sartini[2]

Abstract
Over the last decade, humans have produced each year as much data as were produced throughout the entire history of humankind. These data, in quantities that exceed current analytical capabilities, have been described as "the new oil," an incomparable source of value. This is true for healthcare, as well. Conducting analyses of large, diverse, medical datasets promises the detection of previously unnoticed clinical correlations and new diagnostic or even therapeutic possibilities. However, using Big Data poses several problems, especially in terms of representing the uniqueness of each patient and expressing the differences between individuals, primarily gender and sex differences. The first two sections of the paper provide a definition of "Big Data" and illustrate the uses of Big Data in medicine. Subsequently, the paper explores the struggle to represent exhaustively the uniqueness of the patient through Big Data is highlighted prior to a deeper investigation of the digital representation of gender in personalized medicine. The final part of the paper put forward a series of recommendations for better approaching the complexity of gender in medical and clinical research involving Big Data for the creation or enhancement of personalized medicine services.

1 Big Data: a definition still in progress

The use of Big Data in medicine has been defined as one of the top ten revolutions in the coming decade (Shaikh et al. 2014). Indeed, the analysis of Big Data (also known as "mining") opens entirely new possibilities for the optimization of health care process, the identification of clinical interdependencies between diseases, and decision support (Carnevale and Tangari 2021). However, experts and companies may not always mean the same thing when they use the expression "Big Data". Some medical researchers argue that, for the most part, we tend to define Big Data as just "a lot of data" (Willems et al. 2019). Furthermore, confusion exists as to whether Big Data is a technology in it of itself (Hilbert 2016). Let us make clear: it is not. Indeed, Big Data is an expression that is used to refer not only to a gargantuan quantity but also to a mining process through technological means that have developed greatly over the last decade. Indeed, computer algorithms' ability to sort through and automatically identify in data recurring patterns, is the technological development that has made possible refining "the new oil." This controversial label, provided by British mathematician Clive Humby, is meant to highlight data's inherent worth in today's digital economy: many stand to gain financially and/or politically from the control of data itself.

The recurring patterns spotted in data, so elusive to the untrained human eye given the vastness in which they are immersed, lead to new intelligence and therefore a competitive advantage over—in the case of health—disease. However, the "spotting" of the patterns can occur transparently, i.e. in a way that allows human researchers to reconstruct the processes automatically performed by machines, or in

✉ Antonio Carnevale
a.carnevale@cyberethicslab.com

Emanuela A. Tangari
e.tangari@cyberethicslab.com

Andrea Iannone
a.iannone@cyberethicslab.com

Elena Sartini
elena.sartini2@gmail.com

[1] CyberEthics Lab., Via Giuseppe Vasi, 18/A, 00162 Rome, Italy

[2] Bologna, Italy

a so-called "black box," i.e. with a logic that even highly trained technicians struggle to grasp. Yet, before delving into the specificities of data in health care, it might be helpful to briefly sum up the main characteristics of Big Data.

In 2001, Doug Laney, the then-Vice President and Service Director of the company Meta Group, ideated the "3V" model for Big Data management: *Volume*, *Velocity*, and *Variety*.[1]

- Volume: every day, over the course of many activities, we generate data. With "Volume" we refer to the sheer mass of them, something so large that traditional technologies could not wrangle. The volume of data is growing continuously and international analysts estimate that the amount of data worldwide in 2025 will increase by 530% compared to 2018.[2] This is why it has been difficult to identify a threshold above which data can be defined as "big." For now, let us consider the limit of 50 Terabytes or data volumes that grow more than 50% on an annual basis.
- Velocity: data are created and acquired ever more rapidly thanks to the vast proliferation of devices with embedded sensors that collect data in real-time. The challenge enterprises face is the need not only to wrangle data but also to analyse them in real-time, so that they can make decisions with the greatest timeliness possible.
- Variety: the data that exist today are incredibly diverse in terms of their sources. While in the past data tended to be generated and kept within a single organization, today companies and public entities tend to acquire more data from external sources. Data is created not from only enterprise transaction and resource planning systems, but also sensors, social networks, and data made available by public entities such as the Bureau of Labor Statistics (commonly known as "open data"). One continuum on which data can be placed is that of their degree of structure: organized elements with standardized relationships to one another are known as "structured data," while elements whose relationships are not organized in a predefined way are known as "unstructured." Data cover the entire range. For example, all the credits and debits on one's bank accounts are structured data (they are organized in a table that lists their relationship to merchants,

time, place, etc.) while the data contained in a YouTube video are unstructured and require processing to answer questions such as: "Is that an animal or a person's face? If it is an animal, what kind of animal is it?"

Today, the 3V paradigm is no longer robust enough and has been updated with a fourth characteristic.

- Veracity: analysts love to harp that "Bad data are worse than no data." Data must be reliable if they are to be used to reveal truths. Guaranteeing data quality and integrity is necessary if analyses are to be considered useless, equitable, and sustainable.

Yet, even with the addition of veracity, the paradigm is still not sufficiently strong. Experts are realizing that referencing the volume of data and the velocity with which it must be analysed addresses aspects of a relationship, and thus of knowledge, that point to the value-based structure of society. For this reason, a fifth "V" has been introduced[3]:

- Value: by definition, a datum is a base point, a coded representation of an entity, of a phenomenon, of a transaction, or of an event. Per se, it is of little to no value, since value—to paraphrase Hume—exists in the eye of the beholder. Indeed, data is mined, as mentioned, through manual and/or automatic analytical processes. The information generated often has meaning—if found to be at all significant—only for those involved in the process. By following a well-known hierarchy made popular by Richard Askoff, a computer scientist, we can say that information may lead to knowledge, i.e. what is obtained when information is used to make decisions and take action; in other words, when information is "put into practice." In Askoff's hierarchy, as with the passage between data and information, the action of transitioning from knowledge to the next and final level, wisdom, implies the mediation of human sensibilities. Wisdom, according to Askoff, is the empirical awareness necessary to undertake the proper course of action in a given context. But how are contexts examined? How are certain courses of action deemed proper? It depends on the priorities of those involved in the process. Human beings lade data and information with ideological and socioeconomic values, creating knowledge and wisdom

---

[1] See more details here: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[2] International Data Corporation, *The Digitization of the World* (sponsored by Seagate), Nov. 2018. https://www.seagate.com/our-story/data-age-2025/.

[3] This 5V approach has been exposed for example by Anil Jain, Vice President and Chief Medical Officer at IBM Watson Health https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data.

whose relevance depends on the expectations and the contextual needs of the humans involved. Perhaps it is for this reason that efforts to (i) develop methodologies to evaluate the monetary worth of data within organisations on their balance sheets[4] and (ii) pass legislation creating a "data dividend" in the United States[5] have been woefully unsuccessful. In other words, the value of data is known precisely to only a select few and it is known—consciously or not—a priori with respect to the mining processes necessary to extract it. Outside a world made up of norms and values, it is impossible to "put into practice." (Habermas 1996). We leave in background the important question of the relationship between awareness and action: as Aristotle noted challenging Socrates, knowing what is right is not the same as putting it into practice, because it requires a set of human capacities that cannot be traced back to pure knowledge, starting from a specific virtue that is *phronesis*, i.e. practical wisdom. It is a special kind of knowledge that is action-oriented and requires experience; the ability to assess conditions and situations and the resources to put the best action into practice; knowledge of oneself and the world, awareness of the relationship between oneself and others, awareness of nature. The *phronesis* is therefore the form connected to consciousness and judgement that can be translated into the action of human existence, the *praxis*: in it, the moment of knowledge, awareness, and action are intrinsically connected. The *phronesis*, like practical philosophy, requires knowledge of the universal, but above all, it requires a certain experience of life, which makes the human being able to observe and distinguish cases because action always takes place in such individual cases.

---

[4] In the private sector, companies increase their economic value by placing intangible assets such as rights to certain intellectual properties in specific portions of their financial statements. However, the data an organization possesses are not recognized as intangible assets, despite the fact they can be exploited for monetary gain. Part of the reason this is not possible is that, while say a patent has a value on the market, data sets (often protected by privacy laws) may have no value outside of the company in which they are created.

[5] Under the 1st Session of the current Congress of the United States (116th), Senate Bill 1951 sponsored by Senator Mark Warner. [D-VA] was introduced in June 2019 in an attempt to require the Securities and Exchange Commission (SEC—the governmental agency that regulates the purchase and sale of stock, bonds, and other securities) "to disclose to users what types of user data are collected, and the usage and value of that data." A hearing on the bill, referred to the Committee on Banking, Housing, and Urban Affairs, was held in October 2019. https://www.congress.gov/bill/116th-congress/senate-bill/1951/all-actions.

2 Big Data in medicine: sources and potentiality

One of the most important uses of Big Data in medicine regards the creation or enhancement of personalized and precision medical practices. Personalized medicine can be defined as the consideration of the genotypical and phenotypical (environment, lifestyle, social relationships, etc.) characteristics of each individual receiving health care.[6] Analyses of Big Data in pathology serve to support this practice by providing more apt information about each patient's unique characteristics. This information allows symptoms to be treated or cured more efficiently and in harmony with deontological practices on one hand and socioeconomic realities on the other. Indeed, personalized medicine is distinguished for being "precise," "preventive," "predictive," and "participatory" (Hood and Friend 2011).

At a closer glance, however, it becomes clear that what makes this medical practice "precise" is not the aid of sophisticated technology such as robotic surgery, but the contribution made by each patient as a unique individual. The same goes for the "preventive" and "predictive" characteristics. *Theoretically*, by using insights from Big Data, everyone should be able to learn about their own clinical history and genetics, thus favouring their personal understanding of how to reduce their probability of suffering from certain illnesses and of which treatments would likely be most effective for them. *Practically*, however, these outcomes are possible only if everyone consents to formalizing their own clinical history so that their information can be consulted by health care specialists beyond their primary care doctor. In fact, one of the main sources of Big Data is patients themselves: patient-related outcome measurements (PROMs) and patient-related experience measurements (PREMs) record all sorts of measures through computer and mobile apps provided by their caregivers or sourced by the patients themselves.

Beyond those provided by patients, data from a large variety of sources are utilized in the practice of personalized medicine. These data are *derived* from patients. These include various data subjects and are usually recorded in electronic patient files for clinical purposes. As Willems et al. argue, these data contain the clinical data of patients, diseases, treatments, and outcomes, and may include "demographic details such as gender and age, presenting symptoms, family history, comorbidity, radiological data (such as CT, MRI, PET, US) as well as solid and liquid tissue-based analysis (such as histopathological diagnosis/features, immunohistochemistry, DNA/RNA sequencing experiments,

---

[6] https://ghr.nlm.nih.gov/primer/precisionmedicine/precisionvspersonalized.

blood analyses, and whole genome BAM files)" (2019, p. 9). But also, data from in vitro experiments can be an important source. Another source of Big Data includes the computational analysis of these data. These processed data comprise indirect and computed data, including radiomics and digital image analysis as well as genetic expression and mutation analyses.

3 The risks of a data-driven medicine in research and patient's representation

The use of Big Data in medicine opens the doors to a series of complex problems, above all in moving from the *descriptive analytics* software, which allows information represented by the data to be displayed with graphs, tables, and other visual supports that highlight the underlying medical phenomena, to *predictive analytics* software, which through more or less transparently developed IT algorithms identify correlations between the provided medical data with an "if … then" logic (Cato et al. 2016; Nicoletta et al. 2018). In this section, we will analyse the main challenges that experts in the field ascribe to Big Data when attempting to represent and describe a patient's uniqueness.

3.1 The risk of obsolescence for human theorization in medical diagnoses

The first risk is quite vast, somewhat epistemological. To create value from data, it is crucial to turn raw data into useful information and actionable medical and patient-centred knowledge. Some even aim at producing a conceptual framework of predictive clinical capabilities. This process involves formidable computer algorithms. Machine learning algorithms develop their own rational patterns of computation and data mining by themselves, making the scientific method obsolete. Against this backdrop, Chris Anderson has claimed the well-known "end of theory": "Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity […] the numbers speak for themselves" (Anderson 2008). Could it be that smart algorithmic searching through oceans of data can spare us the labour (and the joys) of learning and theorizing how the world works? (Succi and Coveney 2019).

One risky field could be, in this sense, if diagnostics were entirely centred on Big Data analyses. This would heavily impact the scientific method. In fact, clinical studies—unlike physiological ones—do not attempt to derive universal laws, but rather to explain an event or a series of events connected to specific illnesses or pathologies. It's one thing to know, it's another thing to recognize. And in clinical studies, what

matters is the latter. This human ability might be affected negatively by the widespread and uncontrolled usage of Big Data analytics.

3.2 The risk of losing a holistic vision of society's health problems (or technoscientific holism)

A second problematic aspect follows. Medical pathologies conserve in their aetiology the characteristics of a more general symptomatology that harkens back to the causal models of both medical and social sciences. The diagnosis that derives from the concrete interaction between doctor and patient establishes an assertion on a given set of symptoms (the pathology), but it is nonetheless an assertion on mediated truths. In other words, a diagnosis—correct or not—describes a state of being that is factual or erroneous with respect not to the actual reality, but to a set of conceptual and causal factors that are mediated in a variety of ways. This holistic ability would be challenged by a widespread usage of Big Data analytics in medical diagnoses. Some scholars have called this new way of making patients in the digital age a 'technoscientific holism' (Tretter 2018; Vogt et al. 2016) which assumes that, "by means of more granular mechanisms and technologies to capture ever more aspects of patients' bodies and lives, it is possible to represent them objectively as *wholes*. At the same time, however, this *holism* neglects the organismic aspect of human health and disease" (Pot et al. 2019, p. 172).

Moreover, such a loss would concern not only doctors and health care workers. The loss would reflect vastly on what many call "wisdom of crowds." The holistic vision, in fact, needs sufficiently well-educated people who gather and judge information separately and make their decisions independently. As James Surowiecki stated in his book, influencing people's decisions will increase the likelihood of mistakes, which might be costly. Moreover, the information basis may get so biased over time that no one, including government institutions and intelligent machines, might be able to make reliable judgments (Surowiecki 2004).

The cases of COVID-19 deaths in the United States and the United Kingdom are emblematic of the impact of non-health factors on health care outcomes.

In the US, the Center for Disease Control and Prevention (CDC) has analysed the distribution of deaths from the disease by race at the country, state, and county levels.[7] At the country level, as of July 29, 2020, Non-Hispanic Black

---

[7] The CDC receives data only as far as the county level, and not all counties gather and report data in the same way. Nevertheless, through specific methodological corrections and by weighing the number of deaths with the proportion of population of a certain race, the analyses were possible. https://www.cdc.gov/nchs/nvss/vsrr/covid19/health_disparities.htm.

or African Americans accounted for 16.4% of deaths due to COVID-19 though making up 13.4% of the total population, while Non-Hispanic Whites accounted for 41.6% of deaths due to COVID-19 though making up 60.1% of the total population.[8] The reasons behind the disproportion are essentially two. First, there is a higher concentration of Black Americans in lower-wage front-line jobs[9] and in the geographical areas hit hardest by the virus. Second, among Black Americans there is a higher incidence per 1,000 people of risk factors for COVID-19 such as diabetes.[10]

In the UK, similar findings were mirrored in reports by the Office of National Statistics (ONS) and the Institute for Fiscal Studies (IFS). Examining data from March to May 2020, the former reported that: "the mortality rate for deaths involving COVID-19 was highest among males of Black ethnic background at 255.7 deaths per 100,000 population and lowest among males of White ethnic background at 87.0 deaths per 100,000."[11] The latter, instead, found that: "Occupational exposure may partially explain disproportionate deaths for some groups. Key workers are at higher risk of infection through the jobs they do. More than two in ten black African women of working age are employed in health and social care roles. Indian men are 150% more likely to work in health or social care roles than their white British counterparts."[12]

While these analyses evidently rely on public, open data, to the extent of our knowledge they do not use Big Data, e.g. those data with more or less medical relevance generated by the usage of technological equipment. In medicine, however, it could be conceivable to increase the complexity of the algorithms used to analyse Big Data to include socioeconomic data or those from other domains of life. However, it follows that the output of such an analysis would no longer be strictly medical and that, as highlighted by Russo (2021),

implications and recommendations would no longer be confined solely to the health field, but would spill over into areas of policy, such as public health or labour law. Therefore, if social factors are not accounted for and a solely clinical view of the patient is favoured, then health care professionals establish aetiologies of the disease that by their very nature restrict the set recommendations being put forward, at the expense of more definitive potential solutions.

3.3 The risk of data determinism based on provisory and incomplete inference

One of the most well-known risks among experts is that Big Data elicit their predictions by computing interoperable but often incomplete and unstructured datasets (Anderson 2008). From a meta-perspective, Big Data might be understood as a maker for a new data-driven determinism that can re-shape the future by transforming statistical possibilities into real-life predictionary probabilities. This ontological deficit is not just a mathematical question of data granularity and the need of new computational models. There is a related societal and ethical questioning. Imagine a Big Data engine that hands out electronic health record (EHR) data for helping doctors to identify which approaches will be effective for which patients based on genetic, environmental, and lifestyle factors. EHR data are generated and collected in heterogeneous everyday situations and settings, which by definition are entrenched in social structures and difficult to disentangle (Prosperi et al. 2018). Are we completely sure to have in place the best countermeasures to detect and balance the endemic indeterminacy of data against human complexity? Are we capable of managing this risk of a data determinism that celebrates quantity and probability and curtails quality and social innovation?

When Big Data are employed to support human decisions, one crucial aspect should be seriously taken into account: Big Data, however statistically fast and voluminous, can never extract "realistic" predictions since a data-driven prediction is a math operation that, therefore, risks losing sight of the meanings that human beings entrust to their predictions. Predictions are value-laden assessments. This means they are constructed by a calculatory logic that can compute but lacks a substratum of personal experiences. Therefore, the real determinism in using large amounts of data does not lie in *their* forecasting ability, rather in *our* caution in using them to avoid social sorting and digital discrimination. Outside the context of health care, in the report published by the *European Network Against Racism* (ENAR), the authors detail a "hardwiring" effect of racialised police responses based on certain Big Data collection strategies (Williams and Kind 2019). Data-driven approaches will always deliver some output, but this might be just an "opinion" of an intelligent machine rather than a

---

[8] The percentages of deaths from COVID-19 are weighted distributions of the population. Total population percentages are taken from the United States Census Bureau. https://www.census.gov/quickfacts/fact/table/US/RHI725219.

[9] Rho, H.J., Brown, H., Fremstad, S., *A Basic Demographic Profile of Workers in Frontline Industries*, Center for Economic and Policy Research, April 07, 2020. https://cepr.net/a-basic-demographic-profile-of-workers-in-frontline-industries/.

[10] Appendix Table IV in Center for Disease Control and Prevention, *National Diabetes Statistics Report*, 2020. http://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf.

[11] Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 15 May 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronaviruscovid19relateddeathsbyethnicgroupenglandandwales/2march2020to15may2020.

[12] Platt, L., Warwick, R., (2020), Are Some Ethnic Groups More Vulnerable to COVID-19 than Others?, in The Institute for Fiscal Studies Deaton Review. https://www.ifs.org.uk/inequality/chapter/are-some-ethnic-groups-more-vulnerable-to-covid-19-than-others/.

fact. "This becomes clear if we assume to run two identical superintelligent machines in different places. As they are not fed with exactly the same information, they would have different learning histories, and would sometimes come to different conclusions. So, super-intelligence is no guarantee to find a solution that corresponds to the truth" (Helbing 2019, p. 57).

3.4 The risk of a totalitarian digital society based on biomedical data control

The creation, development, and implementation of Big Data driven systems in the field of medicine are not immune from criticisms also from a legal perspective. While Big Data in medicine can concretely contribute to the well-being of humankind, their use in biomedical research may be subject to misuse or exploitation, which might put in peril fundamental rights such as privacy, personal data protection, the right of non-discrimination, and the right to self-determination. As a matter of fact, data can be knowledge, and nowadays, knowledge is power.

In other words, the risk is creating a society where each data regarding a specific individual is digitally transformed, analysed, stored, combined with other data, and used for different purposes with several means, by different actors. Considering elements such as market share, market dimension, and market power as well (widening, therefore, the perspective to include not only a pure data protection point of view but also a competition law perspective) in this society, few private and/or public entities would know (almost) everything about a huge number of individuals. Besides profiting in monetary terms from this knowledge, they could also profit in terms of policy making, public decisions, allocation of funds, or even elections.

The aforementioned considerations are extremely current when it comes to considering personal data in the field of medicine, which, by using the words of the EU legislator, are sensitive. This trait can be easily understood by making reference to some of the definitions provided within the EU General Data Protection Regulation 2016/679 ("GDPR"). In article 4 of GDPR, a distinction among genetic data, biometric data, and data concerning health has been made, which for the purposes of ease of reading in the present paper, can be jointly referred to as "biomedical" data, considering that it is possible to distinguish a common trait in all three definitions. Biomedical data are indeed the representation of those features intrinsically tied to the very nature of an individual. By analysing, or processing to use the language of GDPR, many, potentially indefinite sub-data and sub-information can be inferred.

Without prejudice to the above, however, neither data nor its analyses are a problem per se. Indeed, questions start to arise with the creation of the correlations between the sources of the data, the databases, the findings derived by the analysis of the data, and with the identification of the subjects that will profit (in scientific as well as in monetary terms) from the entire set of processing operations.[13] By connecting the (millions of) dots, it is possible not only to re-create the complete health history of a person but also to know his or her habits, to predict potential genetic diseases or diseases consequential to the place of living. Without the intention to undermine any legitimate clinical study, it is undeniable that the knowledge of the past, present, and potential future medical conditions of individuals results extremely useful and profitable for entities who are essentially functioning just to make profits (insurance providers and pharmaceutical companies to name two), as well as to policymakers in terms of definitions of new health strategies or policies for the allocation of funds.

The possibility to collect and then analyse this huge flow of biomedical data in the field of medicine is now facilitated by the implementation of wearable devices during (even if lawful and legitimate) clinical protocols, where, we, as human beings, are the direct source of our biomedical data. This practice should be then added to all the devices through which we, on a daily basis, individually register personal information (e.g. mobile phone health app which are counting our steps by default). Therefore, leaving aside for a moment clinical protocols, assuming that many of the devices that we use daily collect those data and that many of those devices are interconnected with other devices, a domino effect that has as object individual personal information becomes evident. Things get even more complicated when the Internet of Things is taken into consideration, with its extrinsic potentiality of having all of our devices as sources of our own biomedical data.

After having briefly considered the potential effects of the use and correlation of biomedical Big Data, it is now interesting to acknowledge also that the entities that might benefit or profit from collecting "raw" biomedical Big Data or from their analyses, might be either private or public. If the benefits for the private sector might be easily spotted considering pharmaceutical and insurance companies, the view only becomes murkier once the public sphere, such as the State, is considered, as well. In the paper "The Dark Side of Numbers", Seltzer and Andersen (2001) present five examples[14] on how the collection of personal data referred

---

[13] The dictum whose origins can be traced back to Hume, and which any student of statistics learns immediately, must be taken into account in this matter: correlation is not causation. This poses subtle problems regarding the interpretation of data, its methods, models, and boundaries as highlighted by Russo (2015).

[14] In particular the examples cited by the authors concerned the Holocaust, the recording/listing of Native Americans in the United States, the internment of Japanese Americans during World War II,

to religious affiliation, ethnicity, or other sensitive traits, has led to some of the darkest moments in human history. To recall just one, the "comprehensive system of population registration" (i.e. a census) developed in the 1938 and perfectioned in 1941 in The Netherlands was then used by Nazis to identify with absolute certainty Jewish citizens and where they lived prior to deporting them to concentration camps. With hope that such tremendous acts never happen again, there are governments or policymakers that might benefit in other concrete ways from the creation of large biomedical databases. Indeed, it would be possible to imagine that in a future not that far away, by aggregating certain sets of data, even for statistical purposes, public entities might decide to change the dietary system of certain schools in certain areas, due to the genetic pre-conditions of the parents of the pupils. In addition, considering that nowadays the limits between the private and the public sector is becoming more and more blurred, the risk that decisions affecting public life are taken by private entities might occur with a higher frequency rather than in the past.

4 Gender and medicine

After analysing the problems in the representation of the patient, in this part, we address the first crucial point of the article: what are the difficulties of representing the gender dimension in medicine?

4.1 A gender-averse diagnostic medical model

The gender dimension, while opening up contingent and specific questions, poses a preliminary discriminant that is the consequence of the diagnostic paradigms that have arisen over the course of the history of medicine. In fact, gender poses basic epistemic questions, such as those of "difference" and "intersectionality" (Crenshaw 1989; Gilligan 1993).[15] Difference and intersection, therefore, imply dialectics and dialogue which seem to be the characteristics neglected by the present medical diagnostic model. Nowadays there is a shift from medical data gathered through analogue, narrative, and human-mediated exchanges towards digital ones. In the Western world, the biomedical model that became prevalent in the nineteenth century meant that

information about the body was no longer created in an active dialogue between the healer and the patient, but was supposed to come from the healer interrogating the passive patient's body directly. "This process largely coincided with the rise of laboratory medicine and the spread of hospitals, whose paradigmatic form of knowledge generation was systematic and quantitatively oriented clinical observation and scientific lab experiments, respectively (Jewson 2009). Both assumed that 'facts' ought to be separated from values and that medicine should rest upon the former. It was against this background, and in particular with the emergence of the evidence-based medicine movement in the 1990s, that randomised controlled trials and the data and information generated through them became the 'gold standard' (Timmermans and Berg 2010) of evidence in medicine" (Pot et al. 2019, p. 171).

4.2 Pre-comprehensive gender discrimination in medical research

There is an underlying gender discrimination that operates before medical research even produces its results. Already in the screening and definition phase of trials, some choices are consistent for example, when accounts from women about their discomfort or pain are taken less seriously than men's (e.g. Hirsh et al. 2014) or when women continue to be underrepresented in medical research (e.g. Duma et al. 2018; Geller et al. 2011; Phillips and Hamberg 2016). Such discriminations often act on unconscious, pre-understanding schemes, which do not always allow researchers to question the relationship between science and society. This implies, therefore, the risk of being blind to those forms of injustice that make certain social groups vulnerable (as described in Paragraph 3.3), the same social groups to which the patients involved in the trials belong. It is not a case that such thoughtlessness affects, in addition to gender, also other aspects of human vulnerability, like race/ethnicity or age (e.g. Denson and Mahipal 2014; Hamel et al. 2016). Furthermore, even when the gender dimension is taken into consideration, it is treated as a naturalistic male–female contrast. In medical practice and research, gender is understood as sexual dimorphism. As Anne Fausto-Sterling (2000) has shown, the binary categories of sex and gender are a 'social decision' and not the representation of inevitable biological realities.

4.3 Poor gender representation in clinical trials due to socio-economic factors

Beyond pre-comprehensive discrimination, there is also contingent discrimination in medical research: the stringent requirements of clinical trials often mean that socioeconomic factors limit the representation of women therein. In

---

Footnote 14 (continued)

practices undertaken under the Soviet Union and the genocide in Rwanda.

[15] While the sex and gender differentiation insists on converging dimension of gender inequalities, the concept of "intersectionality" emphasises the "double jeopardy" and how gender interacts with other social identities to shape bias.
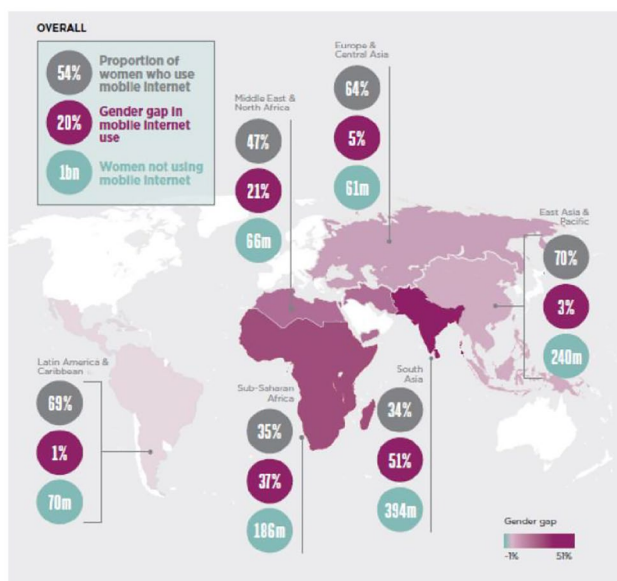
Fig. 1 The mobile Internet gender gap

fact, trials often entail lengthy periods of time under clinical supervision, periods that can conflict with caring and other responsibilities most often held by women (Seeman 2010). Furthermore, women (and more specifically those from a lower socioeconomic status and/or those belonging to ethnic minority groups) show high levels of ethical concern regarding the participation in medical research due to the history of harmful medical experiments, often targeting their reproductive health (Killien et al. 2000).

4.4 The internet mobile gender gap

An additional source of undesired biases in digital biomarkers is the unbalanced access and use of digital devices among people with different sexes and genders as well as education and income (Cirillo et al. 2020). As demonstrated by *The Mobile Gender Gap Report* of 2020,[16] despite progress, the gender gap in mobile internet use overall in low- and middle-income countries remains substantial, with over 300 million fewer women than men accessing the internet on a mobile. The gender gap is widest in South Asia at 51% and remains fairly consistent in other regions such as Sub-Saharan Africa, which has the second-largest gender gap at 37% (Rowntree et al. 2020). Figure 1 shows this gender gap, highlighting how much less likely a woman is to use mobile internet than a man.

5 Gender, Big Data, and personalized medicine

On the basis of the previous argumentation on the relationship between medicine and gender, in this part of the article, we address the second crucial point of the article: what are the difficulties of representing gender in a type of medicine that admits precision and personalization through the use of Big Data?

Gender medicine, which has been gaining ground since the 1990s, develops from and around differences, from the process of prevention to diagnosis and follow-up (Ballard et al. 2016). The objective of a tailor-made therapeutic approach, consequent to these differences, requires countless interventions on several levels: individual awareness, training of the personnel involved, research and study in the technical, social, psychological, and humanistic fields (Koren et al. 2019). The EU-funded project PERSIST (*Patients-centered SurvivorShIp care plan after Cancer treatments based on Big Data and Artificial Intelligence technologies*) is aimed at creating a system capable of improving treatments for cancer survivors and improving their health and well-being, through various phases and objectives that make use of Big Data models to facilitate decisions on the best treatments.

Far beyond sexual and biological difference, gender medicine focuses on social, cultural, and territorial distinctions to carry out an articulated and complex analysis; from this analysis it is possible to achieve equality and equity-based on diversified and personalized medical, political, and social choices, adapted to the specific needs of people and their constitution. Complex analysis and general awareness for those involved at all levels of decision-making are fundamental to orient methodologies, political and medical decisions, and non-discriminatory and increasingly appropriate technological innovations.

One could argue that AI-health technologies act as a double-edged sword (Cirillo et al. 2020). They represent a shift away from the male 'standard patient' by trying to comprehensively and objectively represent many different aspects of patients' lives and bodies. Indeed, in contrast to traditional datasets, Big Data can involve a change in scales—speed, capacity, continuous generation—as well as a change in the relationality, flexibility, repurposing, and de-contextualization of data (Metcalf et al. 2016). As Pot et al. argue: "When people are characterised individually with their anatomical, genomic, endocrinological, metabolomic, and other relevant dimensions, we may no longer need generic labels such as gender. The newly founded groups could correspond with clusters along existing gender categories in some respects—such as around pregnancy and childbirth—but not in others, such as in endocrinological terms where 'similar' patients

would be grouped together in ways that cut across traditional categories of sex and gender. Personalised and precision medicine could thus also be viewed as a way to overcome existing tensions between the desire for the abolishment of gender bias, on the one hand, and the important critique of the invisibility of women and gender minorities in biomedicine, on the other" (2019, p. 178).

5.1 Systemic gender biases in clinical trial

Against this backdrop, the barriers already mentioned are not the best conditions for enhancing artificial intelligence in favour of gender-sensitive precision medicine. Indeed, despite the explicit break with the past that Big Data epistemologies typically claim, many of the systemic biases around gender that characterised clinical trial data continue to be found in Big-Data-centred practices and technologies as well (Cirillo et al. 2020).

5.2 Electronic phenotyping in electronic health records

But the problem does not concern only the data collected in clinical studies, which de facto are more governable towards bioethical requirements of respect for human rights and the dignity of the person (as happens in research projects funded by the European Commission). The aspect that generates the greatest concern is instead educating artificial intelligence by feeding it with data from EHRs. Yet, biases become harder to detect due to their entanglements with the everyday social practices through which such data are generated. "An example of this is electronic phenotyping. Electronic phenotyping seeks to discover patient characteristics that the patients themselves did not disclose or that they were not even aware of, using advanced statistical analysis of large sets of EHR data" (Pot et al. 2019, p. 175). It can be used, for example, to detect yet undiagnosed patients with diabetes (Holt et al. 2014) or those with a high risk of suicide (Tran et al. 2014). Ultimately, since EHR data are generated and collected in heterogeneous everyday situations, their integrity depends on a whole series of factors both endogenous and exogenous with respect to medical practice (Prosperi et al. 2018, p. 10). Women for example are more likely to face other barriers that lie outside the health care system, such as care responsibilities, limited access to transportation, or a lack of workplace flexibility. Furthermore, women receive lower-quality care, which is associated with their more frequent use of health care services (Dehkordy et al. 2016).

5.3 Big Data versus data feminism?

Feminist's criticisms on the use of Big Data in health care are becoming stronger, understandably more elaborate and refined. If the first thing that comes into the mind of the reader is that gender discrimination in medicine and clinical experimentation occurs because researchers discriminate against women during the recruitment phase, the reality is instead much more complex. As a matter of fact, gender discrimination in data processing in the field of medicine occurs with a frequency, with a "normality", and on so many levels, that it goes undetected.

If this statement seems extremist, a mention of Caroline Criado Perez's work is worthwhile. In her book *Invisible Women. Exposing Data Bias in a World Designed for Men* (2020), the author intelligently notes that even the human body is pictured as a "male" body in medicine textbooks when it comes to describing the image of "neutral body parts". Moreover, she pointed out that, for the same disease or pathology, the female body might show different symptoms other than the male body (just one example concerns heart attacks), without then even considering that there are some diseases or health conditions that do not affect the male body (e.g., pre-menstrual syndrome).

One of the problems is that, in the context of gender issues, "a widely used approach to ensure fairness in data processing is to remove some sensitive information, such as sex or gender, and all other possible correlated features. However, if inherent differences exist in the underlying population, such as sex differences in disease prevalence, this procedure is undesirable as the outcome would be less fair towards specific minorities. Indeed, the learned patterns that apply to the majority group might be invalid for a minority one" (Cirillo et al. 2020, p. 7). The adoption and perpetration of this approach can lead to additional (intentionally or unintendedly) forms of discrimination, which however might put in danger the life of the human being concerned. By not involving women in clinical trials for the development of a new drug, its negative effects will not correspond completely to reality. And the worst part is not that women are not voluntary involved in the clinical trials, but that women simply cannot participate in many clinical trials, due to work and family constraints.

Instead of promoting the neutrality of the medicine and of its research, but more in general of the data, medicine and research, in general, might find better and more suitable solutions by adopting a differential approach, where female and male presences are differentially present in research studies (depending on what kind of studies are conducted).

5.4 Male-dominated culture in computer science

Although there is now literature that demonstrates that there is a hidden and untold story about women pioneers in computing (Abbate 2017), gender-sensitive approaches to computing continue to be rare; there is a lack of gender balance in computer science. Basically, one gender-related hypothesis

to explain the underrepresented presence of women in computer sciences can be found in pre-comprehensive factors of the reviewers of their scientific and technical findings. Very often this pre-comprehensive realm—permeated of stereotyped opinions, filtered by beliefs nourished more by common sense than by the analysis of facts—is unconscious at the individual level. Such a make-believe background leads, on the one hand, to an unconscious underestimation of the results of the techno-scientific work of women and, on the other, to an intrinsic disregard of women that contaminates one's own willingness and self-determination to carry out and evaluate their findings, correspondingly helping to enhance the image of women as subjects with intellectual skills that aren't apt to the computer sciences (Cohoon and Aspray 2006).

This same gap is the harbinger of an ideologizing, male-dominated culture that continues to produce a pervasive belief that societal issues can be solved by better technologies. In our view, the inability to see gender equality as a possible driver of computer science research inevitably produces a dystopian image of data-driven technology as keen predictive analytics for approaching and solving all the risks and diseases of contemporary society. Consequently, this assumption tends to increase the risk that built-in gender biases continue to go unnoticed, alongside exaggerated expectations and reliance on technologies (Broussard 2018).

6 Will Big Data and personalized medicine really do the gender dimension justice?

Now is the time to try to answer the question: will Big Data and personalized medicine really do the gender dimension justice? According to the professional experience we have in various research projects and initiatives, it is hard to provide a definitive and exhaustive answer. This is not a scientific or epistemic conclusion and it must not lead to the assumption of an empirical relativism of "case-by-case". In considering only the significance of detailed cases, indeed, the social and economic causes that often determine the sex and gender inequalities could be neglected. The uniqueness of the patient, therefore, should never stricto sensu depend on her/his clinical data (e.g.: woman, mother, post-breast cancer follow-up operation). Rather, the latter should be read and crossed with other analyses—to give examples: access to technological means that favour personalized medicine; social mobility that allows participation in clinical trials; equal representation in science. Therefore, if on the one hand case-by-case is good for clinically contextualizing precision medicine, on the other hand, the universality of the fight against certain forms of injustice should never cease, but rather be reinforced and become an important evaluation element in the patient's medical treatment.

The road towards gender-sensitive sustainability of the use of Big Data in precision health care is still long. A policy-based recommendation could be to approach Big Data in medicine while being aware of the intrinsic risks of enacting technoscientific holism. In their aetiologies, medical pathologies preserve the characteristics of a more general symptomatology that harkens back to the co-causal models of both medical and social sciences. We are all bodies, but we demand recognition from others. Our being included within a biological identity is only one-half of the process of subjectivization. The other half lies within the social coordinates of our identification (Butler 1997) in the complex and dialectical game between sexuality and gender. This is evident precisely in the ways of medical diagnosis. A diagnosis—correct or not—describes a state of being that is factual or erroneous with respect not to the actual reality, but to a set of conceptual and causal factors that are mediated in a variety of ways (Russo 2015).

In our view, data-driven diagnosis could also be informed by predictive analytics models on the condition that the use of *black box* artificial intelligence systems (Nicoletta et al. 2018) is highly limited. We need to rethink the Big Data architecture from the "automatic decision" models to "augmented decisions" ones (Lamanna and Byrne 2018). In the latter case, the software does not decide autonomously, but allows a human being with strong expertise in a certain field, for example, a doctor, to decide based on the evidence discovered by the model and to integrate this additional information with his/her previous knowledge. Against this backdrop, integrating technological decision-making processes with increased human presence can augment gender sensitivity in computing for health care. Indeed, greater attention paid to identity aspects, many of which are currently subordinated or excluded from the socio-technical order, can augment medical sciences' inclusivity and active concern for social justice.

On the basis of these preliminary considerations, at the end of the article, we provide some recommendations on how to approach a gender-sensitive use of Big Data in personalized and precision medicine.

7 Some gender-sensitive recommendations for Big Data in personalized medicine

7.1 No super-intelligences in health contexts

It would be preferable to engage the ICT medical research community to distribute collective intelligence rather than creating a few extremely powerful super-intelligences we may not be able to control.

7.2 Explainable artificial intelligence to avoid technoscientific holism

Getting an explicable justification of how and why the Big Data models reach their conclusions is now becoming more and more crucial since there is an increasing need to understand the specific parameters used to draw clinical conclusions with relevant impact on patients' lives. Indeed, the EU directive 2016/680 General Data Protection Regulation (GDPR) states the "right to an explanation" about the output of an algorithm (see also the EU White Paper on *Artificial Intelligence: a European approach to excellence and trust*[17]). The term "explainable artificial intelligence (XAI)" is used to refer to algorithms that are able to meet those requirements implementing privacy protection (Wachter et al. 2017) and responsible research and innovation (Stahl and Wright 2018). XAI is a relatively young field of research and its applications so far has not been particularly involved with sex and gender differences.

7.3 Data empathy

An understanding of the contextuality of data (Ferryman and Pitcan 2018) is not only necessary to interpret them in meaningful ways, but also to increase sensibility for potential gender and other biases. Context sensitivity also includes reflections about the reliability and accuracy of results when analyses are based on data that has been collected for a specific purpose and is subsequently analysed for other purposes.

7.4 Data bodies to avoid the invisibility of gender

In clinical trials, the underrepresentation of women was visible because the gender or sex of participants was recorded; missing women were, literally, missing bodies (Casper and Moore 2009). "The 'data bodies' that digital medicine operates with often have no explicit gender attribution anymore; when digital epidemiology analyses entries in online search engines or the movement of people throughout a city via their phone's geolocation, these data are often used without any information on gender. The non- or underrepresentation of women and gender minorities is thus invisible" (Pot et al. 2019, p. 180).

7.5 Interweaving individual variation with group classifications

The invisibility of gender occurs also in cases when gender is taken into consideration. Indeed, when medicine insists too much on personalisation, the individual uniqueness supersedes and resists group classifications, including gender. This, together with the integration of data from multiple sources and the hypothesis-free mining of datasets, makes biases harder to detect.

7.6 Do not "Tokenize" the keywords

Before evaluating the impact of any application of personalized medicine, a meta-question should always be asked: "Personalized medicine is fine, but for whom?" In other words, it is essential to never lose sight of the realistic relationship between the transformation that medicine promotes and the implications that this transformation has for concrete, embodied patients. Let us take the "gender-sensitive" formula for example. If this formula does not represent an effective opportunity for everyone—patients, family members, doctors—to transform a real practice, "gender-sensitive" could remain a merely symbolic, rhetorical form of political correctness, reduced to a linguistic code applied to reality to represent it but not to change it.

8 Conclusions

Gender biases challenge the quality of results yielded from Big Data analytics in precise and personalized medicine. The gender issues entail biomedical considerations, in an indirect way, since they help remind us that the definition of Big Data is still in progress and the specialist knowledge for analysing them is still in development. At the same time, gender biases persist in the context of biomedical Big Data in a direct way, creating new configurations of old problems. As Pot et al. (2019) state, digital practices in medicine reconfigure the gendering of societies in at least two ways: by making genderless visible, and by making gender bias harder to detect.

The aim of this article has been to dissect these double-faced gender biases. On one hand, we have analysed the problems that characterise the generic relationship between medicine and gender; on the other, we have outlined the conceptual and practical obstacles that still make applying artificial intelligence algorithms to large streams of biomedical data contradictory and not very gender friendly.

In the core of the paper, we summarised different viewpoints through which the application of data-driven technology can be understood as problematic from a scientific, ethical, or societal perspective. Solving these problems and

---

[17] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

reducing gender biases in those areas of medical research where gender-sensitive instances occur, starts not with the development of a new generation of predictive machines capable of not creating them in the first place, but with implementing the recommendations offered. Data intelligence can stimulate new situational awareness in health care professionals and encourage the participation of actors both within and outside of the health care system, thus making it more prone to satisfying patients' rising demand for more holistic care.

References

Abbate J (2017) Recoding gender: women's changing participation in computing. MIT Press, Cambridge

Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete. Wired 16(7). Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Ballard DW et al (2016) Optimizing clinical decision support in the electronic health record. Clinical characteristics associated with the use of a decision tool for disposition of ED patients with pulmonary embolism. Appl Clin Inform 7(3):883–898. https://doi.org/10.4338/ACI-2016-05-RA-0073.

Broussard M (2018) Artificial unintelligence: how computers misunderstand the world. The MIT Press, Cambridge, MA

Butler J (1997) The psychic life of power. Theories in subjection. Stanford University Press, California

Carnevale A, Tangari E (2021, forthcoming). Philosophical News 22, special issue on "Ethics, Health Data, and Bio-Citizenship"

Casper MJ, Moore LJ (2009) Missing bodies: the politics of visibility. NYU Press, New York

Cato KD, Bockting W, Larson E (2016) Did i tell you that? Ethical issues related to using computational methods to discover non-disclosed patient characteristics. J Empir Res Hum Res Ethics 11(3):214–219. https://doi.org/10.1177/1556264616661611

Cirillo D et al (2020) Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj Digit Med 3(81). https://doi.org/10.1038/s41746-020-0288-5

Cohoon JM, Aspray W (2006) Women and information technology. Research on underrepresentation. MIT Press, London

Crenshaw K (1989) Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chicago Legal Forum 140:139–167. Available at: https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

Dehkordy SF et al (2016) The link between everyday discrimination, healthcare utilization, and health status among a national sample of women. J Women Health 25(10):1044–1051. https://doi.org/10.1089/jwh.2015.5522

Denson AC, Mahipal A (2014) Participation of the elderly population in clinical trials: barriers and solutions. Cancer Control 21(3):209–214. https://doi.org/10.1177/107327481402100305

Duma N et al (2018) Representation of minorities and women in oncological clinical trials: review of the past 14 years. J Oncol Practice 14(1):e1–e10. https://doi.org/10.1200/JOP.2017.025288

Fausto-Sterling A (2000) Sexing the body: gender politics and the construction of sexuality. Basic Books, New York

Ferryman K, Pitcan M (2018) Fairness in precision medicine. Data & Society. Available at: https://datasociety.net/output/fairness-in-precision-medicine/

Geller S et al (2011) Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? J Women Health 20(3):31520. https://doi.org/10.1089/jwh.2010.2469

Gilligan C (1993) In a different voice: psychological theory and women's development. Harvard University Press, Cambridge, MA

Habermas J (1996) Between facts and norms: contributions to a discourse theory of law and democracy. MIT Press, Cambridge

Hamel LM et al (2016) Barriers to clinical trial enrolment in racial and ethnic minority patients with cancer. Cancer Control 23(4):327–337. https://doi.org/10.1177/107327481602300404

Helbing D (2019) Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies. In: Helbing D (ed) Towards digital enlightenment essays on the dark and light sides of the digital revolution. Springer Nature, Cham, Switzerland

Hilbert M (2016) Big data for development: a review of promises and challenges. Dev Policy Rev 34(1):135–174. https://doi.org/10.1111/dpr.12142

Hirsh A et al (2014) The influence of patient sex, provider sex, and sexist attitudes on pain treatment decisions. J Pain 15(5):551–559. https://doi.org/10.1016/j.jpain.2014.02.003

Holt TA et al (2014) Identification of undiagnosed diabetes and quality of diabetes care in the United States: cross-sectional study of 115 million primary care electronic records. CMAJ Open 2(4):E248–E255. https://doi.org/10.9778/cmajo.20130095

Hood L, Friend SH (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 8(3):184–187. https://doi.org/10.1038/nrclinonc.2010.227

Jewson N (2009) The disappearance of the sick-man from medical cosmology, 1770–1870. Int J Epidemiol 38(3):622–633. https://doi.org/10.1093/ije/dyp180

Killien M et al (2000) Involving minority and underrepresented women in clinical trials: the National Centers of Excellence in Women's Health. J Women Health Gender Based Med 9(10):1061–1070. https://doi.org/10.1089/152460900445974

Koren G et al (2019) "A patient like me." An algorithm-based program to inform patients on the likely conditions people with symptoms like theirs have. Medicine 98(42):e17596. https://doi.org/10.1097/MD.0000000000017596

Lamanna C, Byrne L (2018) Should artificial intelligence augment medical decision making? The case for an autonomy algorithm. AMA J Ethics 20(9):E902–910. https://doi.org/10.1001/amajethics.2018.902

Metcalf J, Keller EF, Boyd D (2016) Perspectives on big data, ethics, and society, council for big data, ethics, and society. Available at: http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/

Nicoletta M et al (2018) Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità. JAMD 21(3). Available at: https://www.jamd.it/wp-content/uploads/2018/11/2018_03_03.pdf

Perez CC (2020) Invisible women. Exposing data bias in world designed for men. Vintage Books, New York

Phillips SP, Hamberg K (2016) Doubly blind: a systematic review of gender in randomised controlled trials. Glob Health Action 9(1):29597. https://doi.org/10.3402/gha.v9.29597

Pot M, Spahl W, Prainsack B (2019) The gender of biomedical data: challenges for personalised and precision medicine. Somatechnics 9(2–3):170–187. https://doi.org/10.3366/soma.2019.0278

Prosperi M et al (2018) Big data hurdles in precision medicine and precision public health. BMC Med Inf Decision Making 18(139). https://doi.org/10.1186/s12911-018-0719-2

Rowntree O, Shanahan M, Bahia K, Butler C, Lindsey D, Sibthorpe C (2020) The Mobile Gender Gap Report 2020, GSM Association, London

Russo F (2015) Causation and correlation in medical science: theoretical problems. In: Schramme T, Edwards S (eds) Handbook of the philosophy of medicine. Springer Science+Business Media. https://doi.org/10.1007/978-94-017-8706-2_46-1

Russo F (2021) Value-promoting concepts in the health sciences and public health. Philosophical News 22, special issue on "Ethics, Health Data, and Bio-Citizenship", edited by Antonio Carnevale and Emanuela Tangari (forthcoming)

Seeman MV (2010) Women's issues in clinical trials. In: Hertzman M, Adler L (eds) Clinical trials in psychopharmacology. Wiley-Blackwell, Chichester

Seltzer W, Anderson M (2001) The dark side of numbers. Soc Res 68(2):481–513

Shaikh AR et al (2014) Collaborative biomedicine in the age of big data: the case of cancer. J Med Internet Res 16(4):e101. https://doi.org/10.2196/jmir.2496

Stahl BC, Wright D (2018) Ethics and privacy in AI and big data: implementing responsible research and innovation. IEEE Secur Priv 16(3):326–333. https://doi.org/10.1109/MSP.2018.2701164

Succi S, Coveney PV (2019) Big Data: the end of the scientific method. Philos Trans Roy Soc Math Phys Eng Sci 377:20180145. https://doi.org/10.1098/rsta.2018.0145

Surowiecki J (2004) The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Doubleday, New York

Timmermans S, Berg M (2010) The gold standard: the challenge of evidence-based medicine and standardization in health care. Temple University Press, Philadelphia

Tran T et al (2014) Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. BMC Psychiatry 14(76). https://doi.org/10.1186/1471-244X-14-76

Tretter F (2018) From mind to molecules and back to mind: metatheoretical limits and options for systems neuropsychiatry. Chaos 28(10):106325. https://doi.org/10.1063/1.5040174

Vogt H, Hofmann B, Getz L (2016) The New Holism: P4 systems medicine and the medicalization of health and life itself. Med Health Care Philos 19(2):307–323. https://doi.org/10.1007/s11019-016-9683-8

Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the general data protection regulation. Int Data Privacy Law 7(2):76–99. https://doi.org/10.2139/ssrn.2903469

Willems SM et al (2019) The potential use of big data in oncology. Oral Oncol 98:8–12. https://doi.org/10.1016/j.oraloncology.2019.09.003

Williams P, Kind E (2019) Data driven policing: the hardwiring of discriminatory policing practices across Europe. European Network Against Racism, Brussels. http://bit.ly/2PTYmSl