

---

# Data Circulation in Health Landscapes

**Niccolò Tempini**

*University of Exeter*

**Antonio Maturo**

*University of Bologna*

**Elisabetta Tola**

*Formicablu*

**Abstract:** The crossing boundaries intends to open a dialogue between Science and Technology Studies, Social studies of Health and the emerging Data Journalism perspective. It explores major issues at stake in contemporary practices of producing and sharing data, with a focus on the COVID-19 pandemic.

**Keywords:** health data; platforms; risks; pandemic; data journalism.

**Submitted:** April 4, 2022 – **Accepted:** June 10, 2022

**Corresponding author:** Niccolò Tempini, Department of Sociology, Philosophy and Anthropology, University of Exeter & Alan Turing Institute, Byrne House FF8, Exeter (UK), [n.tempini@exeter.ac.uk](mailto:n.tempini@exeter.ac.uk)

\* \* \*

## **Polysocial Risk Scores and Behavior-Based Health Insurance: Promises and Perils**

*Antonio Maturo, University of Bologna*

## **Cotton Balls, Zinc Supplements and Predictive Analytics**

Once upon a time, a long time ago, around 2010, an irate father walked into a Target store on the outskirts of Minneapolis. He asked to speak with the manager, and upon their arrival, he waved coupons and vouchers in their face:

“My daughter got this in the mail!” he said. “She’s still in high school, and you’re sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?”<sup>1</sup>

The manager apologized profusely and stammered that he had no idea how this could have happened.

A few days later, the same manager called the father to apologize again, but something happened:

On the phone, though, the father was somewhat abashed. “I had a talk with my daughter,” he said. “It turns out there’s been some activities in my house I haven’t been completely aware of. She’s due in August. I owe you an apology.” (ibid.)

What led to this bewildering encounter was a new office in that retail location, where a mysterious new practice had been implemented: *Predictive Analytics*. A sudden change in the young woman’s shopping patterns had been noticed, signaled through her loyalty card, sparking an unanticipated chain reaction. Back in 2010, retailers had just started to collect intimate details about consumption habits. They had noted that:

Women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date.<sup>2</sup>

Because of this shift in purchasing habits, the young woman’s pregnancy had been made apparent in her data-double, even before her social identity.

This incident occurred over ten years ago, while paper mail was still the main form of promotion. In the meantime, self-tracking has exploded, generating enormous amounts of data, especially physiological and behavioral data. In addition, sophisticated algorithms can monitor the time we spend on a site, the physical places we visit, and the likes we place. By monitoring our credit cards, it is possible to know what we eat and how many calories we ingest. Especially in the context of COVID-19, unseen sensors can recognize who is running a temperature in a train station. In the field of health, therefore, there is not only big data but thick data: data that can tell us about our health from a clinical, physical and social point of view.

In this datafication of health, perhaps the two most disruptive and cutting-edge developments are “Polysocial Risk Scores” and “Behavior-Based Health Insurance”. These two areas, in some ways, overlap since the risk score is the basis of health insurance, of which Behavior-Based is the most advanced kind. Surrounding both are big players with keen interests and high expectations. Both Polysocial Risk Scores and Behavior-Based Health Insurance share close attention to social aspects of health, and both are

driven by the need to predict possible (individual) futures on the basis, of course, of quantification (e.g., datafication). Surrounding these developments, beyond innovative possibilities, are clear doubts and concerns about their implications and consequences in terms of social justice.

## The Polysocial Risk Score

In order to understand what Polysocial Risk Scores consist of, it is helpful to underline the main features of the Polygenic Risk Score, which in some ways acts as its prototype.

The Polygenic Risk Score estimates the risk that a person has of developing a disease from his or her genes. More precisely, the Polygenic Risk Score represents the total number of genetic variants that an individual has to assess their heritable risk of developing a particular disease since multiple genetic mutations and their interactions cause most diseases.

At first glance, Polysocial Risk Scores can be seen as the sociological version of the Polygenic Risk Scores, with the idea of the Polysocial Risk Score being developed in the context of the social theory of social determinants of health.

Social determinants of health are the factors that affect a person's health, namely education, income, type of work, type of housing, neighborhood, social cohesion, and others. These determinants affect health through lifestyles, health literacy, and access to care. Epidemiologists and health sociologists have repeatedly confirmed the influence of social context and social determinants on physiology. The determinants of health are strongly intertwined, e.g., how income influences health and how it, in turn, is affected by education; how the weight of income and how the weight of education affects a person's health, and how much, in turn, the weight of education on the possibility of acquiring higher income.

Therefore, the challenge is to weigh and estimate the conditioning of social determinants and their interactions with individual health. However, to date:

Most efforts to precisely quantify the influence of individual social determinants of health have failed, largely because the causal pathways are numerous, interconnected, and complex. (Figuerola et al. 2020, 1553).

The enormous amount of data that can now be acquired on people's health could mark a turning point for developing precise estimates of individual risk of becoming ill. Notwithstanding, one would have to arrive at a Polysocial Risk Score for each disease or health outcome, even in this case. One person would then have several Polysocial Risk Scores. Nevertheless, compared to the Polygenic Risk Score, there is a considerably more turbulent level of complexity:

One key difference is that unlike polygenic risk scores, which are not dynamic because the scores are based on an individual's genes, polysocial risk scores may change if an individual's social circumstances change. (ibid.)

Where the Polygenic Risk Score is static, the Polysocial Risk Score is (would be) dynamic. Moreover, the same social determinants have different weights in different social contexts. Indeed, and methodologically it is even worse with some social determinants being part of the context itself (e.g., social capital and social cohesion).

As Figueroa and colleagues (2020) illustrate, it is necessary to constantly collect, aggregate, and mobilize data from different domains regarding people's quality of life and sociodemographic data Polysocial Risk Scores need to be periodically updated. Above all, it is necessary to relate these "external" data to people's state of health, to their "internal" health data, and to their physiology.

Moreover, as scores are elaborated and processed by algorithms, in some cases, health data may result in biases and, in worst cases, social discrimination. As summarized by Leslie et al.:

AI systems can introduce or reflect bias and discrimination in three ways: in patterns of health discrimination that become entrenched in datasets, in data representativeness, and in human choices made during the design, development, and deployment of these systems (2021, 1).

Thus was the case of genetic data, as in the U.S, most genome-wide association study-based polygenic risk scores have been based on populations of European descent, neglecting the health of other ethnic minorities.

## **Pricing Risk: Behavior-Based Health Insurance**

Creating the Polysocial Risk Score would be something between miracle and mirage, yet this does not mean that attempts have not been made. On the contrary, the health analytics industry is a rapidly developing sector in the digital firms of Silicon Valley and the biotech industry of the Boston Area, with the American health insurance agencies leading the charge towards the construction of health risk scores, with the latter being interested in knowing the health status of their members. Moreover, actors that has most influenced this orientation of health insurance, at least according to some scholars, has been a legal provision contained in the Affordable Care Act (ACA), approved in 2010. As Liz McFall points out:

The ACA alternative introduced a "behavioural" approach (...) including new responsibilities to pay a "fair share" of the costs of the entire pool and be "as healthy as you can." The responsibility to be healthy is promoted by the provision of access to preventative care and treatments for chronic,

preventable disease. (...) This emphasis on behavioral responsibility is a great fit with data-driven healthcare innovations including wearable self-tracking devices and apps. (Mc Fall 2019, 60).

This provision has operated in “association” with other factors, primarily technology. As McFall (2019) and Schüll (2016) point out, digital technology and the ACA have been presented as a “dynamic duo” working together, and

compelling insurers, health care providers and consumers to cut costs (...) shifting the management of chronic conditions like diabetes and heart disease away from hospitals and doctors and into the hands of patients themselves (Schüll 2016, 318).

If over a decade ago the office of a chain store was able to learn of a customer’s pregnancy through her purchases of hygiene products, what can health insurers know about us today? What could insurance “providers” learn when they are given access to sociodemographic data, clinical data, genetic predisposition, and, more importantly, lifestyle data (not simply “lifestyle data” as in whether individuals are smokers or vegetarians, but all digital activities and data-doubles)? Moreover, some digital platforms have already identified rich sets of data points for proxies of social determinants of health:

individual purchasing behavior, consumer engagement with advertising, insurance claims, sentiment, and expression in online forums, credit histories, and online social networks (Rowe 2021, 4).

This data, in turn, is coupled with the mundane data generated by personal FitBits, generously gifted by health insurance agencies (Maturo and Moretti, 2018).

Before the spread of digital social networks, Christakis and Fowler (2010) wrote that social friend networks greatly influence personal decisions. Christakis and Fowler showed through animated sociograms based on accurate longitudinal research how certain behaviors may be “contagious”. Not only does a person have a high probability of gaining weight if their friend does, but also if their friend’s friend does, this can be further applied to divorce and smoking cessation. Today these analyses are immensely easier given the ease with which big data can be collected and processed. The predictive potential delivered to insurance agencies is enormous, leading to correlation taking the place of causation, with the latter becoming an obsolete 20<sup>th</sup>-century category (Anderson 2008).

Raschel Rowe (2021) has done thorough research on the platform “Opioid360”, a platform that combines browser histories, credit, insurance, social media, and traditional survey data to sell the service of risk

calculation in population health. Created as a tool that would support over-worked clinicians to see invisible signs of potential addiction in their patients, Opioid360 paved the way for broader applications to prevent chronic diseases. Most importantly:

By extending digital phenotyping imaginaries, Opioid360's presentation appealed to the notion that comprehensive personal data can offer behavioral science the precision that genomics has offered to identify rare diseases (Rowe, 2021, 4).

In their analysis of Vitality health insurance, McFall et al. (2020) make clear that:

Behaviour is Vitality's core brand value and its policies provide incentives to customers to meet behavioral targets, share their data with the company and share their progress on social media (McFall et al. 2020, 7).

The big switch that many health insurers have made is to link insurance premiums and access to specific policies to the constant digital monitoring of physical activity (InsurTech). In theory, through self-tracking, the premium costs could fluctuate every day, in connection with our physical states, instead of once a year. The extension of insurance surveillance to other aspects of our lives through the datafication of health raises big questions about social justice.

The encouragement of certain behaviors opens an extended reflection on the empowerment of the individual. In social studies of health, it is well known how social context affects a person's health and that certain social factors such as income make adherence to healthy lifestyles relatively easy for some people, while for others practically impossible.

When I arrive at around 8 o'clock outside my department, I often meet one of the ladies who clean the offices – being female, visibly overweight, doing an extremely physical job (maybe she has a disease or seeks satisfaction in food?). She gets up at 4.45 a.m. to start work before 6 a.m. When she greets me at 8 a.m., she lights a cigarette with her South Italian accent before getting into the car. She inhales in big puffs as if it were a prize, a seal, or as we say today in the field of gamification, an award for the work done. However, it is not her avatar who is smoking, unfortunately. Her face is tired, and she is in a hurry – maybe she will light another one soon: she has to go to the other side of the town to do some more cleaning, and there is a lot of traffic by then. Just before entering the department, out of the corner of my eye, I see a colleague of mine jogging through the beautiful palm trees on our campus.

## Algorithmic Forecasting and Insurance Customization

According to Barry and Carpentier (2020), insurance can be defined as

the transformation of unknown individual uncertainty, or chance, into a measurable aggregate risk. Technically, it consists of pooling uncertainty and applying the law of large numbers (Barry and Carpentier 2020, 3).

In this way, the occurrence of catastrophic events for one person was remedied by adding small amounts set aside by all. Through statistical predictions, it is relatively easy to predict that a certain number of insured people will fall ill without knowing who exactly. At least until now, insurance has been based on the concept of socialized actuarialism. However, as early as 1996, O'Malley glimpsed the advance of privatized actuarialism, a more refined approach based on:

a technology of governance that removes the key concept of regulating individuals through collectivistic risk management and places the responsibility for risk management back on the individual (O'Malley 1996, 197).

Thus, whereas traditional insurance was based on prediction (i.e., aggregate predictions at the macro level), the new behavior-based insurance is based on forecasting (i.e., attention to the individual's future at the micro-level). This mode of insurance makes policyholders more responsible for their daily actions and health. However, many scholars question whether, technically, behavior-based insurance can still be considered insurance. Based on the distinction between individual fairness and social fairness, Cevolini and Esposito, effectively summarize how the ancient principle of solidarity can be undermined by new insurance policies:

Algorithmic prediction could radicalize the principle of segmentation, culminating in the extreme case of "segments of one." This would almost automatically mean the end of the risk-pooling on which the principle of risk-sharing is based (Cevolini and Esposito 2020, 4).

The end of risk-pooling carries significant implications as to whether Polysocial Risk Scores have the potential to become a central tool in healthcare. In this regard, a crucial issue here concerns what would happen if Polysocial Risk Scores are calculated and accredited by institutions.

Considering that constructed indicators tend to become objective entities, Polysocial Risk Scores can be employed in different contexts and by different actors; from public health departments, government officials,



technology companies, investors, and private insurance companies (Neresini 2015). In a world that is increasingly computerized, quantified, and managed by algorithms, health scores could be mobilized for a variety of purposes. Some of these uses could be noble and others less so:

Health risk scores are not only useful for immediate patient classification or public health program planning, they are also useful to investors seeking to leverage or hedge their risk exposure. (Rowe 2021, 9).

Although indirectly, a strong impetus for developing health scores has undoubtedly come from COVID-19 pandemic. The pandemic has bolstered the trend of health quantification through the robust joint growth of medicalisation and digitalisation. Most importantly, COVID-19 pandemic has spurred surveillance. To put a long story short: *9/11 increased police surveillance, big data stimulated capitalist surveillance, and COVID-19 hyperbolically accelerated molecular surveillance*. Molecular surveillance can be seen as the scrupulous and precise monitoring of our physiological motions and their instantaneous transformation into data. A panopticon of our internal states, or more precisely: the *endopticon* (Maturó 2015). However, this surveilling is not perpetrated by shadowy officials of mysterious agencies wearing thick-lensed glasses in smoke-filled rooms of some governmental molecular surveillance departments but by algorithms themselves. Programs that react to numbers that exceed certain thresholds, to parameters that measure, compare, and discriminate our physiological motions, collect our behavioral habits and read our molecules' silent but vivacious lives.

Yuval Noal Harari, the author of the successful *Homo Deus*, in an article published in the *Financial Times* on April 19, 2020 entitled *The world after the Coronavirus*, fears a dystopian scenario:

Hitherto, when your finger touched the screen of your smartphone and clicked on a link, the government wanted to know what exactly your finger was clicking on. But with Coronavirus, the focus of interest shifts. Now the government wants to know the temperature of your finger and the blood-pressure under its skin. One of the problems we face in working out where we stand on surveillance is that none of us know exactly how we are being surveilled, and what the coming years might bring. Surveillance technology is developing at breakneck speed, and what seemed science-fiction 10 years ago is today old news.<sup>3</sup>

Harari's concerns reaffirm that health scores will soon be the subject of a Black Mirror episode. Behavior-Based Insurance and Polysocial Risk

Score have disturbing implications, starting with the de-politicization of health, which is no longer understood as a public and social issue but as a business and private concern. The challenge, however, is not to assume ipso facto Luddite or apocalyptic attitudes. It is necessary to find a catalyst that brings health back to the center of public discourse. In a society dominated by chronicity, the masses (of patients and caregivers) should become aware of their strength.

## References

- Anderson, C. (2008) *The end of theory: The data deluge makes the scientific method obsolete*, in <https://www.wired.com/2008/06/pb-theory/> (retrieved June 3, 2022).
- Barry, L. and Charpentier, A. (2020) *Personalization as a promise: Can Big Data change the practice of insurance?*, in “Big Data & Society”, 7 (1), pp. 1-12.
- Cevolini, A. and Esposito, E. (2020) *From pool to profile: Social consequences of algorithmic prediction in insurance*, in “Big Data & Society”, 7 (2), pp. 1-11.
- Christakis, N. and Fowler, J. (2011) *Connected: The Surprising Power of Our Social Networks and How They Shape Our Life*, Boston, Little Brown and Co.
- Figueroa, J.F., Frakt, A.B. and Jha, A. (2020) *Addressing Social Determinants of Health: Time for a Polysocial Risk Score*, in “Journal of American Medical Association”, 323 (16), pp. 1553-1554.
- Fourcade, M. and Healy, K. (2013) *Classification Situations: Life-chances in the Neoliberal Era*, in “Accounting, Organizations, and Society”, 38 (8), pp. 559-572.
- Leslie, D., Mazumder, A., Peppin, A., Wolters, M.K. and Hagerty, A. (2021) *Does “AI” stand for augmenting inequality in the era of covid-19 healthcare?*, in “British Medical Journal”, 372, pp. 1-5.
- Maturo, A. (2015) *Doing Things with Numbers: The Quantified Self and the Gamification of Health*, in “Eä – Journal of Medical Humanities & Social Studies of Science and Technology”, 7 (1), pp. 87-105.
- Maturo, A. and Moretti, V. (2018) *Digital Health and the Gamification of Life: How Apps Can Promote a Positive Medicalization*, London/New York, Emerald.
- McFall, L. (2019) *Personalizing solidarity? The role of self-tracking in health insurance pricing*, in “Economy and Society”, 48 (1), pp. 52-76.
- McFall, L., Meyers, G. and Hoyweghen, I.V. (2020) *Editorial: The personalisation of insurance: Data, behaviour and innovation*, in “Big Data & Society”, 7 (2), pp. 1-11.
- Moretti, V. and Maturo, A. (2021) “Unhome” Sweet Home: *The Construction of New Normalities in Italy during Covid-19*, in D. Lupton and K. Willis (eds.), *The COVID-19 Crisis: Social Perspectives*, New York, Routledge, pp. 90-102.
- Neresini, F. (2015) *Quando i numeri diventano grandi: che cosa possiamo imparare dalla scienza*, in “Rassegna Italiana di Sociologia”, 56 (3-4), pp. 405-431.

- 
- O'Malley, P. (1996) *Risk and responsibility*, in A. Barry, T. Osborne and N. Rose (eds.), *Foucault and Political Reason*, Chicago, University of Chicago Press, pp. 189-207.
- Rowe, R. (2021) *Social determinants of health in the Big Data mode of population health risk calculation*, in "Big Data & Society", 8 (2), pp. 1-12.
- Schull, N.D. (2016) *Data for life: Wearable technology and the design of self-care*, in "BioSocieties", 11 (3), pp. 317-333.