

The Danger of An Over-Personalized Insurance Industry

Krissy Gianforte
MIDS w231, Fall 2018

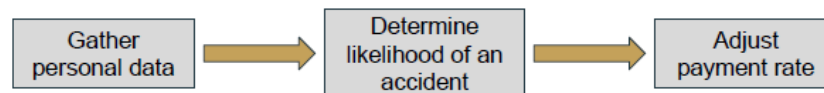
Insurance is a critical concept in the American economy. Insurance programs allow groups of people to pool together their futures; sharing both the chance of an incident and the associated expense among the larger community. Members of the group all agree to pay a small amount (a “premium”) into a pool, which can then be used to cover any one member’s large expenses, should an accident occur. Insurance is available for many areas of life, including health, property, automobile, pets, and more. Currently, both health and automobile accident insurance are mandatory in the United States. Traditionally, insurance rates have been set universally for large groups; for example, health plan premiums are set by employers and tend to be negotiated globally [1]. However, the automobile insurance industry provides an interesting example of more individualized pricing.

Automobile Insurance Incorporates Data

Since automobile insurance plans are typically purchased by individuals or households, rates can be set for fairly specific demographic groups. Data has already entered the pricing scheme, making insurance more expensive for groups that have statistically higher numbers of accidents. For example, young males may pay up to 20% more for auto insurance! [2] Insurers are pulling even *more* data into their pricing models by allowing customers to self-report personal information. StateFarm offers many discounts in this manner: customers who share their academic grades can receive a “Good Student Discount”, and customers who share their

student status and the distance between home and school can receive a “Student Away at School Discount” [3]. A new program called “Drive Safe & Save” even offers a discount based on driving habits. Customers volunteer to have their driving metrics (such as acceleration, “jerkiness”, left turn frequency, and mileage) measured by smartphone or OnStar and reported to StateFarm; that data then determines what sort of discount one may receive [4].

The model is simple: the insurance company uses customers’ personal information to better estimate the likelihood of an accident, then adjusts their premium costs accordingly.



This sort of model has been accepted easily in the automobile insurance industry, likely for a few reasons. First, these programs are structured as opt-in discounts, so no customers are forced into providing personal information if they do not wish to participate. Second, the information being collected is not deeply personal; much of it (such as student status and age) is already public information. Collecting driving metrics may border on intrusive, but the Lake Wobegon Effect [5] helps to explain why many customers have not objected: everyone believes they are a good driver, so they are willing to share their performance data!

What if this sort of personal-data-driven modeling was applied to other insurance industries? Would customers benefit from increased specificity and discounting, or would the concept of insurance collapse from over-personalization? This paper explores the first glimpses of a data-driven health insurance market and considers the economic and ethical consequences of such blending.

A Place for Data in Healthcare

Rapid changes or advancements in the medical industry are often stymied by strict regulations; the health insurance field is no exception, with many of the natural pathways to data incorporation prohibited by law. The HIPAA Privacy Rule requires that only the “minimum necessary” amount of health information be shared about patients [6], so insurers are not able to collect much personal health information about their customers. Even if the information is available, the Affordable Care Act (ACA) prohibits adjusting premiums based on factors such as gender, occupation, pre-existing conditions, and other predictors of health outcomes. In fact, only “five factors can legally affect a plan’s monthly premium: location, age, tobacco use, plan category, and whether the plan covers dependents” [7]. The Genetic Information Nondiscrimination Act of 2008 prevents discrimination based on DNA information, both in hiring and insurance decisions [8]. With so many restrictions, it may seem like the data-based model set by auto insurance providers could never find a place in the health industry.

However, healthcare is currently a contentious topic in the American government and federal programs are shifting quickly. In 2017, congress proposed an amendment the Genetic Information Nondiscrimination Act (H.R. 1313, [9]) that would allow employers to access employees’ genetic information. The supposed purpose of such access would be to tailor workplace wellness programs towards individuals with adverse health factors; however, the proposal sets a precedent for access that health insurers may soon be able to follow. In another segment of government, President Trump is working to extend the reach of short-term health insurance plans, making them available for up to 12 months and renewable for up to 36 months as opposed to the previous limit of three months [10]. Such plans typically cost up to 50% less

(making them a tempting option) but also cover significantly less than comprehensive, long-term coverage [11]. The terms of the ACA do not apply to these plans, so insurers are allowed to adjust rates or deny coverage based on any factors they deem relevant (including pre-existing conditions and other discriminatory factors).

Pulling From All Available Sources

These two federal movements are certainly changing the landscape of health insurance, making individually-set, data-based premiums seem possible. With their appetites whetted, some insurers are even beginning to incorporate non-medical data into their rate models. Though the correlations between these factors and actual health outcomes are less concrete, non-medical data is much less protected by law, so it makes for a tempting alternative.

Fitness tracking bands have offered an easy and seemingly harmless way to interweave health plans and personal data. Employers commonly negotiate discounts and incentives for their employees in the aggregate; if X% of employees wear trackers and Y% hit their step goals each week, a discount is applied for the entire program. Some insurers are beginning to explore options to interface directly with customers, further tailoring the rewards by collecting more detailed personal information [12]. The proposed program would be voluntary for individuals, akin to the opt-in safe driving programs currently offered by automobile insurers. It seems logical, then, customers should be allowed to offer up their own personal fitness data, particularly if it shows that they are at a lower risk of health issues and deserving of a discounted premium.

The logic and appropriateness of incorporating data is not so clear elsewhere, where insurers are turning to data brokers to purchase wide-reaching and detailed personal information about their customers without customers' knowledge or permission [13]. Ostensibly, the data is used for marketing purposes only, to make sure that targeted programs reach their intended audiences. (For example, a woman purchasing plus-sized clothing is at an increased risk of depression and should be notified of mental health care options.) It is hard to imagine, though, that these health prediction models will not slip into payment rate determinations as well.

Social media has also entered the health insurance picture, as it does with every issue in today's digital society. The European re-insurance firm SwissRe has started including Twitter-based sentiment analysis in their insurance rate determination algorithm, on the basis that more optimistic people are at a reduced risk of heart disease [14]. Facebook may be the next logical extension, where life events such as marriage, childbirth, and education milestones are publicly available and neatly collected for data scraping. User-posted photographs also offer an indication of personal hobbies, which may be tied to health status. Should a user whose profile picture depicts her finishing a marathon be eligible for a health insurance discount? Conversely, should a user with a passion for his football team and a habit of tailgating pay increased premiums based on his potential for alcoholism? The concept could be extended into all web activity, where Amazon purchases, magazine subscriptions, websites visited, and articles read may all offer personal information indicative of one's habits, hobbies, and mental state. Smart homes may be another extensive source of data, as recordings from Alexa in-home devices offer data on user's movement patterns and smart fridges reflect dietary habits.

Health predictions based on these factors may seem logical and even innovative for insurance actuaries, but they stand on questionable ethical grounds. Nissenbaum's Contextual Integrity concept suggests "locat[ing] contexts, explicat[ing] entrenched informational norms, evaluat[ing] flows, and evaluat[ing] these flows against norms based on general ethical and political principles as well as context-specific purposes and values" [15]. Social media posts are meant for friends and family, to keep them updated and involved in the same way that personal conversations and phone calls do. In that light, insurers are certainly not part of the intended flow of information, and scraping social media for health risk factors is surely an invasion of privacy. Solove's privacy taxonomy offers a concrete vocabulary for web activity and smart home data: "Surveillance – watching, listening to, or recording of an individual's activities" and "Secondary use – the use of information collected for one purpose for a different purpose without the data subject's consent" are certainly at play in this situation [16].

The Temptation of Genomic Data

On another side of the scientific spectrum, recent years have seen the emergence of direct-to-customer DNA testing services that can supposedly determine one's propensity for disease given just a small sample of saliva. One such company, 23andMe, offers an "Ancestry & Health Reports" package: a combination of 5 Genetic Health Risk reports, 5 Wellness reports, 25 Traits reports, and 40 Carrier Status reports [17]. The health outcome screenings are not clinically validated (resulting in an FDA ruling that 23andMe may only release "raw" DNA data [18]), but that caveat is underemphasized by the company and overwhelmed intense

temptation. If the weak link between web activities and health outcomes is enticing for insurance providers, the DNA data market is irresistible.

If the automobile insurance industry's "Drive Safe & Save" program serves as a model, health insurers may begin offering discounts to customers who volunteer their genetic information and show reduced risk of adverse health outcomes. At first glance, this may seem like an opt-in and therefore acceptable program, however, genetic information is at the complicated boundary of personal and yet impersonal data. DNA tells the story of an individual – and their family. As one article succinctly explains, "When you turn over access to your DNA, combine it with personally identifiable information (e.g., name), then provide survey data (e.g. questionnaires on sexuality, personality, drug use, lifestyle), you hand over the keys to not only your own identify but that of your relatives" [19]. Does one have the right to share his family's biological information in exchange for the personal benefit of an insurance discount? The story of Henrietta Lacks demonstrates just how complicated and contentious consent can be in this sort of situation [20].

Unfortunately, the question of consent may be ultimately irrelevant, as healthcare corporations are pursuing 23andMe's store of genetic data even without participants' permission. Medical companies such as Amgen, GlaxoSmithKline, Pfizer, and Genentech have recently entered partnerships with 23andMe to propel their pharmaceutical design work forward [21] [22]. In 2015, the company had already formed partnerships with 14 groups including both private companies and universities. Anne Wojcicki, the CEO of 23andMe, explained her ultimate vision:

Wojcicki says she envisions a time when 23andMe can allow virtually anybody to run queries on its huge store of genetic data using a web portal. "Need to find out if there's a possibility of asthma in this cohort? Just push a button," she said. [23]

For data scientists, such unrestricted access represents a true breakthrough in research potential. For customers in the healthcare industry, though, big brother feels closer than ever. When health insurers can run a simple web query to determine one's risk of asthma, diabetes, or even end-of-life conditions such as cancer, premium costs will surely become more individualized to reflect this knowledge.

Even in the absence of explicit corporate partnerships, 23andMe is making its genomic data available to a wide audience. Aggregated statements such as "30% of our female users share a particular genetic trait" are offered freely, under the privacy policy's assumption that "this information is different from 'individual-level' information and is not Personal Information because it does not identify any particular individual or disclose any particular individual's data" [24]. However, applying these statistics to an insurance pricing scheme by raising rates for subpopulations with an increased propensity for expensive medical conditions would surely qualify as unfair harm and even discrimination. Even worse, a misunderstood line between correlation and causation may cause these statistics to target already disadvantaged populations. As an example, consider a hypothetical statistic that *70% of Hispanic subjects suffer from diabetes*. In reality, this high rate of incidence may be due to lower wages among the Hispanic population, which force people to turn to less healthy but cheaper fast food. However, when a DNA research company such as 23andMe states that a particular race has an increased occurrence of a disease, the natural assumption is that it is driven by genomic factors.

The already impoverished Hispanic population in question would then begin to suffer increased insurance costs based on a supposedly genetic predisposition that did not really exist.

While 23andMe's privacy policy states that they "will not provide any person's data (genetic or non-genetic) to an insurance company or employer," and affirms their "long-time support of legislative efforts intended to prevent genetic discrimination and to safeguard individuals' genetic privacy," a detailed reading shows just how much harmful information can be released without customers' consent. The company is quite cavalier about sharing data within its network of corporate partners (which is expanding quickly, as mentioned above). Customers' DNA data – intensely personal and unchangeable information – is treated in the same manner as other companies' privacy policies treat phone numbers and email addresses. Even Niantic Inc, which deals in mobile gaming data collected from its PokemonGo app, achieves privacy standards equal to those of 23andMe. In fact, the spirit and wording of 23andMe's privacy policy closely matches Niantic's policy with regards to the transfer of assets between businesses:

We may share some or all of your Personal Information with other companies under common ownership or control of 23andMe, which may include our subsidiaries, our corporate parent, or any other subsidiaries owned by our corporate parent... In the event that 23andMe goes through a business transition such as a merger, acquisition by another company, or sale of all or a portion of its assets your Personal Information will likely be among the assets transferred. (23andMe Privacy Policy, [24])

If we are acquired by a third party as a result of a transaction such as a merger, acquisition, or asset sale or if our assets are acquired by a third party in the event we go out of business or enter bankruptcy, some or all of our assets, including your Personal Data, will be disclosed or transferred to a third party acquirer in connection with the transaction. (Niantic Privacy Policy, [25])

It is surprising and distressing to see that the care taken to see DNA data through a business transaction so closely matches the standard for comparatively inconsequential mobile gaming data.

Even if ethical implications did convince 23andMe to reduce the scope of its data sharing practices, customers may not be adequately protected. As the value of DNA information becomes more widely recognized, surely the databases of 23andMe will come under attack. The company is currently taking measures to protect its data; the privacy policy states that “23andMe employs software, hardware and physical security measures to protect the computers where customer data is stored. We use robust authentication methods to access our systems. Personal information and genetic data are stored in physically separate computing environments, which is in line with the industry standards for security” [24] (emphasis added). Once again, the policies used to handle intensely personal genomic data merely meet – and do not exceed – industry standards. Data breaches have become more common in recent years, and even large corporations such as Target and JPMorgan Chase have had their systems compromised. Extra care must be taken by 23andMe to protect their comparatively more valuable information stores.

Drawing Practical and Ethical Guidelines

The modern digital age is full of temptations for health insurance companies. An abundance of personal data is available, and prediction algorithms developed using big data analytics suggest that health outcomes will soon be anticipated well in advance of their

occurrence. At this horizon, established ethical frameworks can help keep innovation in check, ensuring that data subjects are treated fairly and ethically.

The Fair Information Practice Principles provide a convenient, well-rounded picture of what high-quality data should be: “relevant, accurate, timely, and complete” [26]. In actuality, any one of these qualities is difficult to guarantee for the data discussed in this paper. The accuracy of personal health data is difficult to ensure until technology advances a bit further; for example, fitness band trackers are easily fooled by shaking and arm movements, causing many wearers to record an unrealistic number of steps each day. (Not to mention participants who purposely “cheat” by shaking their trackers or fastening them to the family dog’s collar.) The relevance of personal data for health insurance decisions is also questionable at best, since the relationship between health outcomes and social media posts, habits, hobbies, or even DNA markers have not been clinically proven. Given this lack of clinical studies, it is also impossible to understand what qualifies as a “complete” data set. Can an individual’s health be modeled completely given his fitness tracking data but not his social media history? What about the inverse? Without a scientific study and proven models, it is impossible to know which interaction terms may be most significant, and the data picture may never be considered complete. Finally, attempting to ensure the timeliness of personal health data raises unexpected questions of coercion. Consider a young 23-year-old who provides his fitness tracker information to show that he is in good health and receive a discount on his insurance premium. Is this individual now committed to providing updated tracking information for his entire life, in order to keep his record up-to-date? If he suffers a health complication that would be reflected in his tracking data, is he obligated to share that information even if it would now

increase his premium payment? If he stops sharing, should his record be expunged completely, or should the last known data state be used for predictive modeling? Ultimately, all four of the Fair Information Practice Principles data-quality standards reveal limitations of data in the healthcare industry.

The Fair Information Practice Principles also demand that data be rationally used; that is, algorithms that make decisions based on personal information must make sense and be interpretable by human evaluators. Artificial intelligence and big data analytics skirt the boundary of this rule, with some models (such as neural networks) hiding their calculations behind complex interactions and layers. As health outcomes are predicted using more widespread data and the models become more obscure, customers may suffer increased distrust towards an already-suspect industry. Instead of setting rates for clear demographic groups such as “20-25 year-old males”, insurers will have to explain why one individual’s premium may differ from her sister or neighbor’s rate. It may even be possible for insurers to hide more sinister factors in their rate calculations as models grow in complexity; sexism and racism may enter into projections (consciously or unconsciously), and free-market factors such as the likelihood of a customer switching insurance providers may be rolled in. This sort of comprehensive tailoring is a far stretch from the vision of the Affordable Care Act, which simplified insurance availability and limited the impacting factors to just five simple categories!

At a high level, the very intent of health insurance may be eroded as data enters the industry. As explained above, the purpose of insurance is risk sharing. A large number of individuals, each with a limited chance of suffering a health incident themselves, pay money into a pool, which is then available to protect the small number of individuals who *do*

experience an event. As insurance is over-personalized, the very concept of risk- and expense-sharing is lost. If those with a higher risk of adverse events are forced to pay increased premiums, they are not benefiting from the group coverage model. In fact, taken to its most extreme form, a fully customized health insurance program would perfectly anticipate the health outcomes of each individual and set their premium to match the anticipated cost of their care. Certainly, the very core philosophy of insurance would be lost in such a system!

Protecting the Future of Health Insurance

This paper has shown that current legal protections are not sufficient to protect the future of health insurance; data is simply expanding faster than the federal government can act. Despite clear guidelines in the Affordable Care Act and the Genetic Information Nondiscrimination Act, insurance firms are circumventing legal fences and using personal data to anticipate health outcomes. Supposedly non-medical data sources enable even more tailoring based on private information such as social media activity, habits, and lifestyle choices. If action is not taken, vulnerable segments of the American population will begin to suffer dire consequences as their access to healthcare moves out of reach.

Unfortunately, an effective strategy is not obvious. In order to stay ahead of creative data scientists, federal regulations must protect the *intent* of insurance rather than trying to anticipate specific use models. However, laws based on intent are difficult to enforce, so some specificity is necessary. From this difficult and contradictory position, the government cannot take the legal actions necessary to effectively protect its citizens.

Perhaps instead, data scientists themselves must lead the way. As they develop new models to predict health outcomes based on obscure data points, they must constantly consider the ethical and social implications. Of course, scientific achievement is exciting and innovative development of predictive models should be celebrated. Applying these models to the world, though, requires somber and careful consideration. Are they best used by insurance firms, looking to tailor their programs and draw even more profit from customers? Or could these models be used to prevent adverse health outcomes and improve survival rates? The future is at the hands of the scientists.

References

- [1] D. Rook, "How We Got to Now: A Brief History of Employer-Sponsored Healthcare," JP Griffin Group, 17 Aug 2015. [Online]. Available: <https://www.griffinbenefits.com/employeebenefitsblog/history-of-employer-sponsored-healthcare>. [Accessed 13 Dec 2018].
- [2] P. Gusner, "13 things that affect your car insurance," 29 Jun 2018. [Online]. Available: <https://www.insure.com/car-insurance/car-insurance-factors.html>. [Accessed 13 Dec 2018].
- [3] StateFarm, "Auto Insurance Discounts and More Competitive Rates with State Farm," [Online]. Available: <https://www.statefarm.com/insurance/auto/discounts>. [Accessed 13 Dec 2018].
- [4] StateFarm, "Take Control of Your Discount With Drive Safe & Save," [Online]. Available: <https://www.statefarm.com/insurance/auto/discounts/drive-safe-save>. [Accessed 13 Dec 2018].
- [5] Wikipedia, "Lake Wobegon," 9 Dec 2018. [Online]. Available: https://en.wikipedia.org/wiki/Lake_Wobegon#The_Lake_Wobegon_effect. [Accessed 13 Dec 2018].
- [6] U.S. Department of Health & Human Services, "Summary of the HIPAA Privacy Rule," 26 Jul 2013. [Online]. Available: <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>. [Accessed 13 Dec 2018].
- [7] HealthCare.gov, "How insurance companies set health premiums," [Online]. Available: <https://www.healthcare.gov/how-plans-set-your-premiums/>. [Accessed 13 Dec 2018].
- [8] National Human Genome Research Institute, "Genetic Information Nondiscrimination Act of 2008," 6 Apr 2015. [Online]. Available: <https://www.genome.gov/10002328/>. [Accessed 13 Dec 2018].
- [9] U.S. Congress, "H.R. 1313 - Preserving Employee Wellness Programs Act," [Online]. Available: <https://www.congress.gov/bill/115th-congress/house-bill/1313>. [Accessed 13 Dec 2018].
- [10] B. Bryan, "Trump's plan to reshape the healthcare system without repealing Obamacare is coming together," 1 Aug 2018. [Online]. Available: <https://www.businessinsider.com/trump-obamacare-short-term-health-insurance-plans-2018-8>. [Accessed 13 Dec 2018].
- [11] A. Gorman, "Short-term insurance plans are enticing, low-cost alternatives for healthy people," 3 Dec 2018. [Online]. Available: https://www.washingtonpost.com/national/health-science/short-term-insurance-plans-are-enticing-low-cost-alternatives-for-healthy-people/2018/11/30/14e833e8-ece7-11e8-8679-934a2b33be52_story.html?noredirect=on&utm_term=.5b0c4f5768d4. [Accessed 13 Dec 2018].

- [12] A. Khan, "Can a Fitness Tracker Save You Money on Health Insurance?," 3 Oct 2014. [Online]. Available: <https://health.usnews.com/health-news/health-wellness/articles/2014/10/03/can-a-fitness-tracker-save-you-money-on-health-insurance>. [Accessed 13 Dec 2018].
- [13] M. Allen, "Health Insurers Are Vacuuming Up Details About You - And It Could Raise Your Rates," 17 Jul 2018. [Online]. Available: https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates?utm_source=newsletter&utm_medium=email&utm_campaign=sendto_newsletter&utm_term=.top. [Accessed 13 Dec 2018].
- [14] B. H. Neghaiwi, "In insurance Big Data could lower rates for optimistic tweeters," 23 Oct 2016. [Online]. Available: <https://www.reuters.com/article/us-insurers-bigdata-consumers/in-insurance-big-data-could-lower-rates-for-optimistic-tweeters-idUSKCN12N05R>. [Accessed 13 Dec 2018].
- [15] H. Nissenbaum, *A Contextual Approach to Privacy Online*, 2011.
- [16] D. J. Solove, "A Taxonomy of Privacy," *University of Pennsylvania Law Review*, vol. 154, no. 3, 2006.
- [17] 23andMe, [Online]. Available: <https://www.23andme.com/dna-reports-list/>. [Accessed 13 Dec 2018].
- [18] G. Neff, 13 Dec 2013. [Online]. Available: <https://slate.com/technology/2013/12/23andme-vs-the-fda-both-are-wrong.html> [Accessed 13 Dec 2018].
- [19] HealDove, "Risks and Benefits of 23andMe DNA Analysis: A Personal Experience," 12 Dec 2018. [Online]. Available: <https://healdove.com/health-care-industry/Should-I-Get-At-Home-Genetic-Testing>. [Accessed 13 Dec 2018].
- [20] R. Skloot, "The Immortal Life of Henrietta Lacks, the Sequel," 23 Mar 2013. [Online]. Available: <https://www.nytimes.com/2013/03/24/opinion/sunday/the-immortal-life-of-henrietta-lacks-the-sequel.html>. [Accessed 13 Dec 2018].
- [21] 23andMe, "Investors," [Online]. Available: <https://mediacenter.23andme.com/company/investors/>. [Accessed 13 Dec 2018].
- [22] 23andMe, "23andMe Announces Collaboration with Pfizer Inc. to Conduct Genetic Research Through 23andMe's Research Platform," 12 Jan 2015. [Online]. Available: <https://mediacenter.23andme.com/press-releases/23andme-pfizer-research-platform/>. [Accessed 13 Dec 2018].
- [23] M. Sullivan, "23andMe has signed 12 other genetic data partnerships beyond Pfizer and Genentech," 14 Jan 2015. [Online]. Available: <https://venturebeat.com/2015/01/14/23andme-has-signed-12-other-genetic-data-partnerships-beyond-pfizer-and-genentech/>. [Accessed 13 Dec 2018].

- [24] 23andMe, "Full Privacy Statement," 17 Jul 2018. [Online]. Available: <https://www.23andme.com/about/privacy/>. [Accessed 13 Dec 2018].
- [25] Niantic, Inc, "Niantic Privacy Policy," 30 Oct 2018. [Online]. Available: <https://nianticlabs.com/privacy/>. [Accessed 13 Dec 2018].
- [26] U.S. Department of Homeland Security, "Fair Information Practice Principles (FIPPS)," [Online]. Available: <https://www.dhs.gov/sites/default/files/publications/consolidated-powerpoint-final.pdf>. [Accessed 13 Dec 2018].