Precision Medicine and its Imprecise History

Christopher J. Phillips[1]

[1]**Associate Professor of History, Carnegie Mellon University**

Column Editors' note: For this Mining the Past *article, column co-editor Christopher J. Phillips sketches a history of precision medicine. Though sometimes portrayed as a field with roots primarily in genetics, he describes the decisive historical role played by statisticians and data scientists, particularly a group of midcentury biostatisticians located at the National Institutes of Health.*

Keywords: precision medicine, personalized medicine, biostatistics, genetics, history

The origins of precision medicine are not precisely known. That's due in no small part to ongoing confusion about what precision medicine *is.* Confusion over the boundaries of a new scientific paradigm shouldn't surprise anyone, but even the basic terminology isn't clear in this case. What's the relationship of precision medicine to personalized medicine? What distinction, if any, is being made with evidence-based medicine? Haven't clinicians always striven to provide precise recommendations? As a systematic survey recently concluded, whether called precision medicine or personalized medicine, the phrase has come to refer to the way personal data and biomarkers—particularly genetic biomarkers—might be used to tailor treatments for individual patients (Schleidgen, Klingler, Bertram, Rogowski, & Marckmann, 2013). Nothing in this definition signals what's new about precision medicine, however—genetic information and other patient data have long been used to advance medical research and improve treatments. Only by delving deeper into what precision medicine has meant over time might we understand what's actually new about the age-old attempt to move from individual and seemingly idiosyncratic patient outcomes to generalizable knowledge about health and disease, and the crucial role statisticians have historically played in that process.

Despite the apparent breadth of the term, precision medicine's contemporary proponents effectively have two visions in mind. The first is essentially an advancement of pharmacogenetics—the development of pharmaceuticals on the basis of genetic information. Pharmacogenetics itself is not new, and the broader desire to use genetic data to improve health outcomes has its own long history. Nineteenth-century pioneers in biometry and statistics—including Karl Pearson and Francis Galton— were deeply interested in the relationship of genetics and disease and in particular in promoting eugenical reforms to avoid the manifestation of 'degeneration' in diseases ranging from mental illness to cataracts.[1] Others skeptical of orthodox eugenics' emphasis on individual variation, like the biometrician Raymond Pearl at Johns Hopkins, still attempted to reveal and measure the interaction of "constitutional" *and* environmental factors in the distributions of disease (Comfort, 2012). Though surprisingly little-remembered today, Werner Kalow's 1962 textbook *Pharmacogenetics* had already set out the program of linking therapeutic response to both the biochemistry of drug agents and to the role of genetics and evolution in shaping individual differences (Jones, 2013).

Precision medicine's proponents essentially coopted pharmacogenetics after the successful conclusion of the Human Genome Project around the turn of the century. Subsequent investments of the National

Institutes of Health (NIH) under Francis S. Collins attempted to capitalize on this new knowledge to transform genetic medicine far beyond the study of well-known mutations and chromosomal anomalies (Collins & McKusick, 2001). Indeed, some of the new discoveries have been profound; a handful of successful high-profile drugs based on the genetics of cancer cells—for example, Herceptin (trastuzumab), Erbitux (cetuximab), and Gleevec (imatinib), among others—have given hope that over time our understanding of more diseases will be transformed (Hamburg & Collins, 2010; National Research Council, 2011). Just as the 19th-century bacteriologist Robert Koch's postulation of a one-disease–one-organic-cause paradigm fit diseases like tuberculosis perfectly and others not at all, however, some diseases will likely be amenable to genetic approaches and others not so much.

A second vision proponents of precision medicine espouse is an increased ability to harness and aggregate new data sources concerning the manifestation and treatment of disease. The idea is that by identifying specific genes, biomarkers, or other factors that alter the probability of acquiring or alleviating disease, researchers will be able to design more precise interventions. This conception of precision medicine also draws on a long history of using biomedical data to tailor therapies to individual patients, to compare treatment outcomes numerically, and to develop statistical tools for moving back and forth between individual and aggregate data.

Physicians have, of course, long portrayed their job as tailoring therapeutic recommendations to patients' specific characteristics and particular manifestations of disease. This was true for premodern medical knowledge across most of the world, from traditional Chinese and Islamicate medicine to European humoral theory, which asserted that each person has a natural balance of humors or cardinal substances—with disease occurring as a result of imbalance. Though ideas about etiology and treatment may have been grounded in theoretical understandings (pneumonia caused by an excess of cold and moist phlegm should be treated by exposure to hot and dry substances), premodern physicians had to tailor that knowledge to the specific temperature, blood pressure, diet, and excretions of the patient in front of them. This was undoubtedly a form of personalized medicine.

The contrast that contemporary precision medicine advocates often make is instead with empirical studies of therapeutics, namely, the determination of which treatments result in measurably better outcomes. The idea of testing (trying) therapies on groups of patients and comparing outcomes also has a long history. If the Scottish surgeon James Lind's 18th-century study of treatments for scurvy is sometimes taken as the first formal medical trial, other informal examples can be found going back centuries, from Daniel's biblical trial of the effectiveness of Nebuchadnezzar's diet to the 10th-century Persian physician Abu Bakr Muhammad ibn Zakariyya al-Razi's test of bloodletting.[2] By the early 19th century, the strategy of testing interventions on different groups of people in order to gather evidence about therapeutic effectiveness was well established. During this period, the approach was reframed as the 'comparative method' and empirical reformers like Pierre-Charles-Alexandre Louis in France

and Elisha Bartlett in the United States were encouraging research programs that involved dividing patients into similar groups, treating each group differently, and then carefully comparing outcomes, keeping in mind concepts like probable error to help distinguish chance differences from real ones (Cassedy, 1984; Jorland, Opinel, & Weisz, 2005; Matthews, 1995). Though the comparative method was not widely practiced in medicine in the 19th century, historians have convincingly shown how it was adopted by physician reformers and government regulators in the 20th century in order to combat corrosive special interests and bias by reducing the emphasis placed on individual case reports (Greene, 2008; Greene & Podolsky, 2009; Marks, 1997; Podolsky, 2010). What was eventually termed evidence-based medicine might just as well be thought of as a scientific version of impersonal medicine, in the sense that its mechanisms were focused on identifying the best treatment for any given disease rather than the best treatment for any given patient. Or, as a recent survey of statistical measures in precision medicine concluded, patient heterogeneity was a "nuisance" for evidence-based medicine but a "blessing" for precision medicine (Kosorok & Laber, 2019).

The distinction, though, is perhaps not so clear-cut. Therapeutic outcomes, disease states, and individual biomarkers are almost never stable or invariable. Underlying all approaches to therapeutic testing—including the creation of 'similar' comparison groups—is a basic assumption about which differences matter: in what ways are diseases and individuals alike or different, and how might clinicians use that information to tailor treatments appropriately? A better distinguishing factor for contemporary precision medicine is the way its development was predicated on the availability of ever-larger amounts of data about relevant differences. As the chair of Stanford's Department of Genetics Michel Snyder explains in his introduction to personalized medicine, clinicians have always considered personalization as part of their work, but now the practice of medicine might become "more personalized" because medicine is "entering the era of big data" (Snyder, 2016, p. 1). That is, what's new is the degree of precision afforded by the volume of personal data being collected. The NIH, for example, launched its All of Us research program in May 2018 (following similar efforts such as the UK Biobank and the China Kadoorie Biobank) with an aim to understand "the relationships between circulating biomarkers or genetic variation as they relate to disease prevention." The crucial component was developing a data set large and diverse enough that "patterns will emerge that wouldn't be visible at a smaller scale," enabling researchers to have the "statistical power to make fine-grained predictions about how a given treatment will affect a given individual."[3] (The emphasis on diversity is not just for statistical reasons; it is also explicitly an effort to correct for decades of bias in medical research [Manolio, 2019] even though it is not at all obvious that more data will reduce existing racial inequities [Benjamin, 2019].) The novelty of precision medicine in the context of therapeutic tailoring is a matter of degree, not of kind, at least in its incarnation within multibillion-dollar efforts to enroll millions of patients' health information in the service of harnessing "big data."

The ability to predict how a "given treatment will affect a given individual" consequently relies—somewhat ironically—upon a history of *aggregating* patient data alongside measurements of therapeutic outcomes. By collecting more data linking particular factors to the risk of disease, population-level data could be used to make more precise clinical recommendations for individuals. For example, the landmark Framingham Heart Study was originally initiated in 1948 as a large prospective (cohort) study to establish incidence rates of heart disease in a classic public health or epidemiological sense. By the 1960s, however, statistical measures had been developed that enabled claims about specific biomarkers, behaviors, and other individual patient data to be linked to the probability of developing heart disease. Techniques of standardization and stratification that had once been used to control for confounders in aggregate studies were gradually replaced by prediction models for individual risk factors, even as researchers warned that the causal effect of a treatment could not be directly observed at the individual level (Keiding & Clayton, 2014; Susser, 1985). Statisticians continued to insist on the distinction between average effects in a population and an individual's probability of disease, but for many clinicians, such distinctions were less and less important as populations were sliced into ever-finer groups and associations of biomarkers with risks were made ever more precise. (Easy-to-use tools like the Framingham Risk Score only further blurred this distinction by appearing to calculate an individual's personal risk of disease on the basis of his or her specific medical data.)

However misunderstood or misapplied, new statistical methods were crucial to enabling patient data to be aggregated, generalized, and then applied back again to individual therapeutic recommendations. These methods were not a natural outgrowth of existing academic statistical practices, however, and the statistics behind the Framingham study were developed by a relatively small group of statisticians located at the National Heart Institute and led by Jerome Cornfield. This cluster had split off from the first group of statisticians at the NIH, hired in 1947 under sociologist Harold Dorn at the National Cancer Institute with backgrounds largely in industrial medicine, economics, sociology, and population surveys. Alongside their research into the statistics of observational studies, these NIH statisticians were also developing statistical methods for experimental settings, particularly dose-response studies and randomized clinical trials. Amid the NIH's expanding role in medical research, their methods enabled more precise claims to be made about which individual attributes or experimental conditions were likely to be causally associated with specific diseases. Though modern causal inference was still in the future, statisticians located at the NIH helped lay the foundations by showing how to carefully interrogate evidence of association in settings in which patients and government officials demanded answers to pressing practical questions —how, for example, might individuals reduce their personal risk of lung cancer or heart disease? Their work made a field like clinical epidemiology possible by showing how to combine clinical methods with seemingly distinct epidemiological methods. New statistical methods to model

responses to therapeutic interventions, to make clinical trials more adaptive, to reduce false positives in screening procedures, and to estimate the probability of disease on the basis of aggregated risk factors were all critical to this transformation—and all developed in part by NIH statisticians between 1950 and 1980.[4]

Researchers at the NIH made up only a small fraction of the international efforts of biostatisticians. Given the long history of precision medicine in its various guises, however, the postwar developments at the NIH are as sensible a marker as any for the origins of contemporary precision medicine. The NIH in this period became the largest global funder of medical research, including in biostatistical methods, and its largesse reorganized research activities of universities and independent institutes alike. The NIH was able to shape many subsequent developments through funding mechanisms, from promoting statistically sound study designs to encouraging 'comparative effectiveness research' and 'patient-centered outcomes research'. The NIH also took the lead at midcentury in using electronic computers to process the growing collection of medical data, an essential precursor for using 'big data' methods in medicine (November, 2012). If the idea of making medicine more precise and more personal goes back centuries, then many of the specific concepts, tools, and institutions at the heart of contemporary precision medicine have origins in the postwar NIH.

* * *

The ongoing work of data scientists and statisticians will be crucial to precision medicine, even as the meaning of 'precise' changes over time. The dream of turning medicine into a deterministic science has been around for many years, but the central question remains how researchers should move from individual patient outcomes to generalized knowledge that can be reliably applied in the future to a new patient who is similar in some respects and different in others. From this *longue durée* perspective, the human genome is simply the latest data used to subdivide patient populations and infer appropriate therapeutic recommendations. Moreover, even as an increasing amount of personal medical data is collected, there remains a big difference between identifying genes associated with a disease (or associated with elevated risk for a disease) and developing a successful new drug.[5] Genetic discoveries have provided powerful and novel insights into the pathology of disease, but the complexity of genetic expression and causal pathways suggests pharmacogenetic approaches are unlikely to render probabilistic models or statistical tools obsolete.

The histories we tell matter. If we portray precision medicine as emerging only after the development of big data machinery or the conclusion of the Human Genome Project, we risk obscuring the ways that statisticians and data scientists have long been trying to make medical practice more precise. And this longer history suggests we should be skeptical that the current paradigm will ultimately solve all the problems of medicine: as shown by the experience of Denmark and other countries far ahead of the United States and United Kingdom in their medical data collection efforts, the most far-reaching

benefits remain more promissory than proven (Hoeyer, 2019). Situating precision medicine historically reminds us not only of past breakthroughs, but also of the continued need for careful statistical analyses to make sense of the uncertainties that will inevitably remain.

References

Benjamin, R. (2019). Assessing risk, automating racism. *Science, 366,* 421–422. https://doi.org/10.1126/science.aaz3873

Bothwell, L. (2014). *The emergence of the randomized clinical trial: Origins to 1980* (Unpublished doctoral dissertation). Columbia University, New York.

Cassedy, J. H. (1984). *American medicine and statistical thinking, 1800–1860.* Cambridge, MA: Harvard University Press.

Collins F.S., & McKusick V. A. (2001). Implications of the Human Genome Project for medical science. *JAMA,* 285, 540–544. https://doi.org/10.1001/jama.285.5.540

Comfort, N. (2012). *The science of human perfection: How genes became the heart of American medicine.* New Haven, CT: Yale University Press.

Ellenberg, J. H., Gail, M. H., & Geller, N. L. (1997). Conversations with NIH statisticians: Interviews with the pioneers of biostatistics at the United States National Institutes of Health. *Statistical Science, 12,* 77–81.

Ellenberg, J. H., Gail, M. H., & Simon, R. M. (1994). Preface: Current topics in biostatistics. *Statistics in Medicine, 13,* 399.

Gillham, N. W. (2015). The battle between the biometricians and the Mendelians: How Sir Francis Galton's work caused his disciples to reach conflicting conclusions about the hereditary mechanism. *Science & Education, 24*(1–2), 61–75. https://doi.org/10.1007/s11191-013-9642-1

Greene, J. A. (2008). *Prescribing by numbers: Drugs and the definition of disease.* Baltimore, MD: Johns Hopkins University Press.

Greene J. A., & Podolsky, S. H. (2009). Keeping modern in medicine: Pharmaceutical promotion and physician education in postwar America. *Bulletin of the History of Medicine, 83,* 340. https://doi.org/10.1353/bhm.0.0218

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363*, 301–304. http://doi.org/10.1056/NEJMp1006304

Heijerman, H. G. M, McKone, E. F., Downey, D. G., Van Braeckel, E., Rowe, S. M., Tullis, E., … McCoy, S. (2019). Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: A double-blind, randomised, phase 3 trial. *The Lancet, 394,* 1940–1948. https://doi.org/10.1016/S0140-6736(19)32597-8

Hoeyer, K. (2019). Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science, 49*, 531–555. https://doi.org/10.1177/0306312719858697

Jones, D. S. (2013). How personalized medicine became genetic, and racial: Werner Kalow and the formations of pharmacogenetics. *Journal of the History of Medicine and Allied Sciences, 68*(1), 1–48.

Jorland, G., Opinel, A., & Weisz, G. (Eds.). (2005). *Body counts: Medical quantification in historical and sociological perspective.* Montreal, CN: McGill-Queen's University Press.

Keiding, N., & Clayton, D. (2014). Standardization and control for confounding in observational studies: A historical perspective. *Statistical Science, 29*, 529–558. https://doi.org/10.1214/13-STS453

Kosorok, M. R., & Laber, E. B. (2019). Precision medicine. *Annual Review of Statistics and Its Application, 6*, 263–286.

Magnello, M. E. (1999a). The non-correlation of biometrics and eugenics: Rival forms of laboratory work in Karl Pearson's career at University College London, Part 1. *History of Science, 37*, 79–106.

Magnello, M. E. (1999b). The non-correlation of biometrics and eugenics: Rival forms of laboratory work in Karl Pearson's Career at University College London, Part 2. *History of Science, 37*, 123–150.

Mackenzie, D. (1981). *Statistics in Britain, 1865–1930: The social construction of scientific knowledge.* Edinburgh, Scotland: Edinburgh University Press.

Manolio, T. A. (2019). Using the data we have: Improving diversity in genomic research. *The American Journal of Human Genetics, 105*, 233–236. https://doi.org/10.1016/j.ajhg.2019.07.008

Marks, H. M. (1997). *The progress of experiment: Science and therapeutic reform in the United States, 1900–1990*. Cambridge, UK: Cambridge University Press.

Matthews, J. R. (1995). *Quantification and the quest for medical certainty*. Princeton, NJ: Princeton University Press.

Middleton, P.G., Mall, M. A., Dřevínek, P., Lands, L. C., McKone, E. F., Polineni, D., … Jain, R. (2019). Elexacaftor–tezacaftor–ivacaftor for cystic fibrosis with a single Phe508del allele. *New England Journal of Medicine, 381,* 1809–1819. https://doi.org/10.1056/NEJMoa1908639

National Research Council. (2011). *Towards precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease.* Washington, DC: National Academies Press. Retrieved from https://www.nap.edu/catalog/13284/toward-precision-medicine-building-a-knowledge-network-for-biomedical-research

Olby, R. (1989). The dimensions of scientific controversy: The biometric Mendelian debate. *The British Journal for the History of Science, 2,* 399–320.

November, J. (2012). *Biomedical computing: Digitizing life in the United States.* Baltimore, MD: Johns Hopkins University Press.

Podolsky, S. H. (2010). Antibiotics and the social history of the controlled clinical trial, 1950–1970. *Journal of the History of Medicine and Allied Sciences, 65,* 327–367. https://doi.org/10.1093/jhmas/jrq003

Schleidgen, S. Klingler, C., Bertram, T., Rogowski, W. H., & Marckmann, G. (2013). What is personalized medicine: Sharpening a vague term based on a systematic literature review. *BMC Medical Ethics, 14,* 55. https://doi.org/10.1186/1472-6939-14-55

Snyder, M. (2016). *Genomics and personalized medicine: What everyone needs to know.* New York, NY: Oxford University Press.

Stigler S. M. (2000). The problematic unity of biometrics. *Biometrics, 56,* 653–658.

Susser, M. (1985). Epidemiology in the United States after World War II: The evolution of technique. *Epidemiologic Reviews, 7,* 147–177.

Footnotes

1. Historians have noted that the label of 'eugenics' elides disagreements among practitioners about the source of genetic variations and the extent to which positive or negative eugenic impulses

framed their overall statistical research between 1880 and 1930. See, for example, Gillham (2015), Mackenzie (1981), Magnello (1999a, 1999b), and Olby (1989). ↵

2. Surveys of the literature on the history of trials in medicine can be found in Bothwell (2014) as well as online at the James Lind Library, www.jameslindlibrary.org. Stigler (2000) discusses other trials, including ones by Galen, Asclepiades, and Avicenna. ↵

3. Information on the May 2018 launch of All of Us is at https://allofus.nih.gov/about/all-us-research-program-protocol and https://allofus.nih.gov/news-events-and-media/announcements. ↵

4. For an overview of the contributions of NIH biostatisticians in particular, see the special issues of *Statistical Science* introduced by Ellenberg, Gail, & Geller (1997) and *Statistics in Medicine* introduced by Ellenberg, Gail, & Simon (1994). ↵

5. To give just one high-profile example, NIH director Francis Collins was part of one of the teams that identified the gene responsible for the most common form of cystic fibrosis 30 years ago, and yet only in 2019 was a drug targeting that specific mechanism shown to be efficacious in a phase 3 trial (Heijerman et al., 2019; Middleton et al., 2019). ↵