Big data opens a window onto wellness

Atul J. Butte

Institute for Computational Health Sciences, University of California, San Francisco Mission Hall, 550 16th Street, 4th Floor, Box 0110, San Francisco, CA 94158-2549

"It is difficult to make predictions, especially about the future," is a quote with varied attributions, but is an especially valid aphorism when making medical predictions. The idea of determining what future diseases might affect an individual is decades old, but our modern medical oracles commonly include the use of genetic associations -- such as *BRCA1* – to predict future breast and ovarian cancers, and the use of the Framingham Risk Scores for 10-year future coronary heart disease [1]. While the US Preventive Services Task Force (USPTF) currently lists 98 guidance documents for screening, counseling, and preventive medications [https://www.uspreventiveservicestaskforce.org/BrowseRec/Index/browse-recommendations], in general, these are written to be applicable to the entire population, and not specific subsets of the population. In addition, the USPTF does not represent a research organization that can innovate or discover new preventative strategies. There could seemingly be many prognostics and preventative measures waiting to be discovered.

With that in mind, many new organizations and groups have started to target increasingly larger populations of essentially healthy individuals for research studies using high-density measurements at faster time frequencies [2]. Price, et al., represent Arivale, one of these companies, with a stated goal of "provid[ing] individuals a scientific path to wellness." In this issue, Price, et al., show how 108 individuals were studied over 9 months with a greater level of resolution of measurements than had been previously attempted on a population of this size. Samples were drawn or collected at three time points during the 9 months. The number of measurements run on these samples was immense, including whole genome sequencing, a battery of clinical tests, metabolomes, quantitation of plasma proteins and peptides, and identification of the gut microbiome. In addition, home measurements and daily activity tracking were instituted, but the compliance rates for these appear to have been very low.

This was not just an observational study. Ten coaching sessions were offered over the 9 months to attempt to improve individual health, but exactly how the specificity of these sessions were driven by the genomic risk prediction and the ongoing clinical and molecular measurements was not clear.

While this seems like an immense effort, to put this in perspective, this study ran the length of a usual human pregnancy. With standard practice in the United States, a pregnant woman might undergo a dozen prenatal visits with counseling and advice giving, would get several clinical blood tests at various points, would typically get some ultrasound imaging, and might today get a non-invasive prenatal test involving some targeted genome sequencing.

This study compares with several recent studies looking at cohorts in newer detailed ways, such as 43 individuals studied with fitness, health, and environmental trackers yielding 250,000 daily measurements [3], a cohort of 10,000 with genome sequencing [4], and a cohort of more than 50,000 patients cared for in a conventional medical context but studied with exome sequencing [5]. Price, et al, could also be compared with the UK Biobank collecting baseline measurements (including imaging) on 500,000 individuals [6], or the Precision Medicine Initiative AllofUs Research Program, which will gather molecular, lifestyle, and environmental measurements on a million or more people [http://allofus.nih.gov/]. Of course, classic long-running efforts are also still useful, including the National Health and Nutritional Examination Survey (NHANES), which is cross-sectional but can include a few subsequent outcomes including diagnoses and death, as well as the Framingham Heart Study, now expanding to include microbiome studies and a third generation of offspring, as well as offspring spouses.

With more of these types of studies being done, it might be useful to consider six interesting lessons from this work, which lead to potential suggestions for future studies.

First, it was helpful that Price, et al., analyzed the consistency of several measurements, especially the clinical laboratory tests. Correlations were run on duplicate measurements across vendors, and it is still eye-opening that many standard clinical blood tests do not correlate well across technologies.

Second, genomes are indeed providing more useful medical information to healthy individuals. One of the individuals in this study was noted to have hereditary hemochromatosis, confirmed by genetics. In another recent study, many asymptomatic individuals were unexpectedly discovered to have genetic risk factors for treatable conditions, including familial hypercholesterolemia and breast cancer [5]. We are nearing the point where the sequencing of at least a few key genes could be justified as a screening tool in healthy individuals, perhaps at some point in early adulthood. But there are still unanswered questions. Are whole genomes needed, or are exomes enough? There have been many alternative approaches considered for combining risk alleles [7], but a comprehensive comparison across these many methods still needs to be performed. An ideal risk score calculator would also need to work with genotypes and exomes, not just whole genomes, would be adjustable given environment and behavioral risk and clinical testing, and would need to deal with missing SNPs and work across ethnic subpopulations. But even more pressing is understanding exactly what do we mean by predicting an individual is at risk for a disease, especially when that individual comes to the test with very different pre-test probabilities of disease (see Figure) [8].

Third, patients typically improve with more frequent medical services, but the question remains whether these individuals have been able to improve their health without all the molecular measurements? It is important to know how well these patients would have done with conventional preventative service. It will be more ideal in the future to have studies like this compare these individuals with a control group of conventional medical care. In fact, since many institutions are offering "concierge" style care to their wealthiest clients, it could be a valid question to ask whether these patients in these types of detailed molecular studies do better in terms out outcomes than conventional concierge care (if there is such a thing as conventional concierge care)!

Not only should future studies compare against conventional care, participants in these "interventional arms" should also receive conventional medical care. For example, it will be more important in the future to apply currently used risk scores to these types of cohorts. How much of the medical guidance here would have been covered anyway, if a Framingham Risk Score had been calculated as high?

The reader is also left wondering whether frequent visits with a doctor with just the clinical testing would yield the same outcomes? Price, et al., noted that 47 individuals in their cohort had an out-of-range fasting glucose at baseline, which suggests they have pre-diabetes. The USPTF recommends adults 40 to 70 years of age who are overweight to be screened for abnormal blood glucose [https://www.uspreventiveservicestaskforce.org/Page/Document/ RecommendationStatementFinal/screening-for-abnormal-blood-glucose-and-type-2-diabetes]. Did these participants not get that routine care from their current health providers? The same could be asked about smoking cessation strategies. The results of studies like Price, et al., are going to be more relevant and generalizable to larger audiences, if they can be properly evaluated against current standard medical care.

Fourth, it is natural to ask how long is "longitudinal"? With less than one year covered in this study, it is hard to know whether some of these changes were just seasonal. This comes up in the most significantly improved clinical variable in this study, the vitamin D level. Price, et al, started their study in April, which is near the known nadir of vitamin D levels in Americans [9]. Over the course of 9 months, the vitamin D levels rose, as might be expected given increasing sunlight exposure. Although an analysis was provided, the reader is still left wondering whether the effects seen in Price, et al., could just have been the effect of summer in the United States (or perhaps those few great days of summer in Seattle)? Ideally, studies such as these should run for more than one year, to compensate for seasonal effects.

Fifth, how much customization will we want or need for preventative care? For instance, we probably want all patients to receive their vaccines, for the promotion of herd immunity and public health. Yet at the same time, it is probably the case that not every man or woman will benefit equally from the exact same cancer screening regimen. It is relatively easy to find discover and invent screening tests with which to subject to patients. It is going to be much harder to have enough confidence to remove screening tests from medical care, such as periodic mammograms or prostate cancer studies.

Sixth, it is clear this field needs more statistical and informatics methodologies to analyze these types of data. Modern methods must be developed that can (1) perform deeper analyses of across different-modality measurements beyond correlation, (2) take better advantage of time and temporal information, and (3) can know what relationships are expected and which are interestingly different than expected. And these new methods will need to properly control for multiple hypothesis testing.

From the research perspective, we are certainly going to learn more from the longitudinal study of cohorts with detailed measurements. However, these measurements come with a price. Arivale currently offers their style of wellness testing for $3,499 [https://www.arivale.com/pricing/]. A competitor, Human Longevity, offers their services reportedly for $25,000 to $50,000 (depending on the panel of tests) [http://www.fiercebiotech.com/genomics/venter-s-human-longevity-starts-50-000-health-testing-service]. And neither of these replaces conventional medical care. This puts the price tag for participation in these cohorts beyond the affordable range of most people. It remains to be seen whether these efforts are going to be able to attract a diverse enough community to make the research findings worthwhile and broadly relevant to an entire population, and not meaningful to just a few.

In the end, with so much information being returned to the participants in this and similar ongoing studies, and with the expectation in a study like this that the patients are changing their behavior and lifestyle, the question will arise as to when do we stop calling these efforts "studies"? Does this mean we will then want to promote making all of these measurements on everyone as a matter of clinical care, or just the few measurements that we discover actually "matter"? Is the medical system ready to process and interpret all of these measurements, or at least ready with computational tools for decision support? Will we ever have enough evidence to justify the costs of comprehensive sophisticated molecular testing? In the end, it is important that decisions on whether to expand preventative services are based importance in public health and cost justification, and are then rigorously tested on whether proposed interventions improve outcomes. With those high bars, we have to see that we are just at the very beginning of understanding precision medicine for public health.
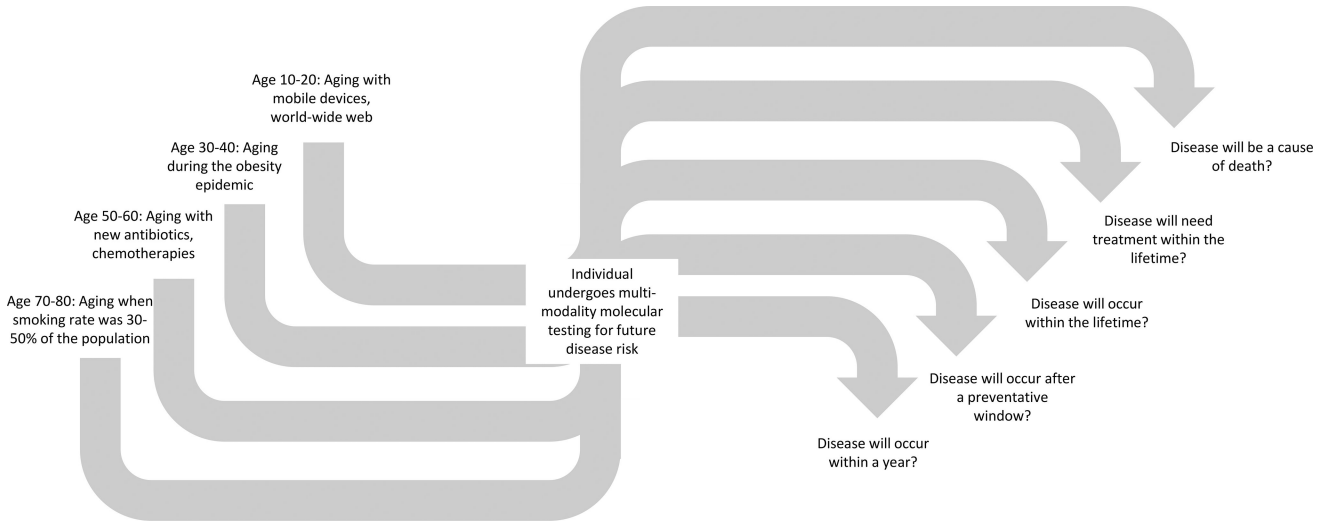
Acknowledgments

References

1. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998 May 12; 97(18):1837–47. [PubMed: 9603539]

2. Sheridan C. Omics-driven startups challenge healthcare model. Nat Biotechnol. 2015 Sep; 33(9): 887–9. DOI: 10.1038/nbt0915-887 [PubMed: 26348939]

3. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, Perelman D, Colbert E, Runge R, Rego S, Sonecha R, Datta S, McLaughlin T, Snyder MP. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information.

PLoS Biol. 2017 Jan 12.15(1):e2001402. eCollection 2017 Jan. doi: 10.1371/journal.pbio.2001402 [PubMed: 28081144]

4. Telenti A, Pierce LC, Biggs WH, di Iulio J, Wong EH, Fabani MM, Kirkness EF, Moustafa A, Shah N, Xie C, Brewerton SC, Bulsara N, Garner C, Metzker G, Sandoval E, Perkins BA, Och FJ, Turpaz Y, Venter JC. Deep sequencing of 10,000 human genomes. Proc Natl Acad Sci U S A. 2016 Oct 18; 113(42):11901–11906. Epub 2016 Oct 4. [PubMed: 27702888]

5. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, O'Dushlaine C, Van Hout CV, Staples J, Gonzaga-Jauregui C, Metpally R, Pendergrass SA, Giovanni MA, Kirchner HL, Balasubramanian S, Abul-Husn NS, Hartzel DN, Lavage DR, Kost KA, Packer JS, Lopez AE, Penn J, Mukherjee S, Gosalia N, Kanagaraj M, Li AH, Mitnaul LJ, Adams LJ, Person TN, Praveen K, Marcketta A, Lebo MS, Austin-Tse CA, Mason-Suares HM, Bruse S, Mellis S, Phillips R, Stahl N, Murphy A, Economides A, Skelding KA, Still CD, Elmore JR, Borecki IB, Yancopoulos GD, Davis FD, Faucett WA, Gottesman O, Ritchie MD, Shuldiner AR, Reid JG, Ledbetter DH, Baras A, Carey DJ. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016 Dec 23.354(6319) pii: aaf6814. doi: 10.1126/science.aaf6814

6. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015 Mar 31.12(3):e1001779. eCollection 2015 Mar. doi: 10.1371/journal.pmed.1001779 [PubMed: 25826379]

7. Morgan AA, Chen R, Butte AJ. Likelihood ratios for genome medicine. Genome Med. 2010 May 17.2(5):30.doi: 10.1186/gm151 [PubMed: 20497613]

8. Butte AJ. It takes a genome to understand a village: Population scale precision medicine. Proc Natl Acad Sci U S A. 2016 Nov 1; 113(44):12344–12346. Epub 2016 Oct 19. [PubMed: 27791179]

9. Kasahara AK, Singh RJ, Noymer A. Vitamin D (25OHD) Serum Seasonality in the United States. PLoS One. 2013 Jun 21.8(6):e65785. Print 2013. doi: 10.1371/journal.pone.0065785 [PubMed: 23805188]

Age 10-20: Aging with mobile devices, world-wide web

Age 30-40: Aging during the obesity epidemic

Age 50-60: Aging with new antibiotics, chemotherapies

Age 70-80: Aging when smoking rate was 30-50% of the population

Individual undergoes multi-modality molecular testing for future disease risk

Disease will be a cause of death?

Disease will need treatment within the lifetime?

Disease will occur within the lifetime?

Disease will occur after a preventative window?

Disease will occur within a year?

Figure.

An individual that undergoes multi-modality molecular testing (including germline DNA sequencing) already comes to the test with a wide range of pre-test probability of disease, given what he and she grew up with, including radical changes in smoking rates, environmental exposures, feeding patterns, and much more. Then, when a complex test then indicates a higher rate of a particular disease, what exactly does this mean? Does this mean a disease may appear in the next year, or within the lifetime? Is this a disease that the individual will need to be concerned about? Modeling these details in pre-test and post-test probabilities are crucially needed now, so individuals and their medical professionals can get more specific utility out of testing.