Precision medicine: discovering clinically relevant and mechanistically anchored disease subgroups at scale

Antony Rosen[1] and Scott L. Zeger[2]

[1]Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

American medicine is on the precipice of dramatic change, forced by disruptive technologies in measurement, computation, and communication. This change is inevitable, because society can no longer afford the nearly $1 trillion in annual waste, a major fraction of which is caused by poorly informed medical decisions and misaligned incentives. In America, if top-down solutions remain stalled, the needed change can only occur through better health decisions based on more valid measurements and analyses that improve medical decisions and the health of individuals and populations.

In our view, precision medicine, by this or any other name, is the science-based application of modern measurement and analysis to improve each health decision. Precision medicine must define clinically relevant and mechanistically anchored health and disease subgroups for which optimal strategies can be followed (when known) and discovered (when not). Precision medicine is the scientific framework of the learning health system that can bring informed innovations to clinical practice (1). As such, precision medicine has the potential to exploit the technology revolutions (2), as most other industries are doing, to improve the health of Americans at more affordable costs.

The scope of the challenge

America spends far too much on health care, given the health of its people. In 2018, US health care expenditures are estimated to total $3.79 trillion (18.1% of GDP) or $11,500 per person. The US is a world leader in access to advanced medical technologies and treatments for end-stage diseases, however, its health is below that of many other developed countries. For example, the US ranks 28th of 44 OECD countries in life expectancy at birth and 28th for people who have survived to 65 years of age (3).

If the US per-capita health care expenditure was equal to that of the second most expensive country, our health care would cost $1 trillion less each year. Similarly, an itemized analysis of the components of health care expenditures suggested $0.91 trillion in wasted expenses, approximately 40% of which was attributed to poorly informed or incented health decisions (4).

A major precision medicine opportunity is to provide evidence for health decisions so that better choices are made by patients, providers, and payers. Precision medicine must operate at the patient-provider interface and define clinically relevant subgroups for which optimal interventions can be devised and suboptimal choices avoided.

The precision medicine strategy

Three simultaneous technology revolutions are reshaping medicine: improved measurements, increased computational power, and better connectivity. The measurement revolution has improved the accuracy and sensitivity of measures by orders of magnitude, allowing the quantitation of thousands of parameters (genes, proteins, metabolites, physiological states, phenotypic features, imaging, and pathology) simultaneously across different biological, clinical, imaging, psychological, and social domains. Similarly, the speed and power of computation and data analysis have greatly accelerated the discovery of relevant subgroups and delineation of their mechanisms (5, 6). Last, the massive gains in connectivity enable efficient movement of health data, acquired in convenient settings for patients, rather than moving humans or hard copies of images and reports. We believe strongly that, if applied incisively, precision medicine produces novel measurement and analytical tools that can define disease subgroups and dictate optimal interventions. For example, in scleroderma, patients who present with concomitant cancer have immune responses to the large subunit of RNA polymerase-3 (encoded by POLR3A) and somatic mutations in the cancers, identifying a homogeneous subgroup of clinical relevance (7) and strongly suggestive of underlying mechanisms (8, 9).

Precision medicine is an opportunity to reimagine medicine. However, too narrow a focus on mainly omics data and deep learning, without anchors to careful phenotyping and disease trajectory over time, will fail to uncover disease mechanisms at scale and will not improve the effectiveness or costs of care. While genetic, social, economic, and environmental factors are the root causes of disease, and prevention is key to long-term population health, 97% of health expenditures in the US are currently for treatment. Secondary and tertiary care and population "metrics" are insufficient to guide those expenditures wisely. In our view, precision medicine's power is in the definition of more homogeneous disease subgroups and using these subgroups to inform important health decisions. Patients, providers, and payers must insist on it.

Assessing an individual's risk, trajectory, and optimal intervention

Precision medicine provides the scientific framework to address common questions that patients likely ask regarding their health state, trajectory, and the likelihood of benefit for intervention choices. Precision medicine contributes to the answers by developing better measurements of the current patient and of a reference group of "otherwise similar" patients with known outcomes to assess the probabilities of the patient's benefits and risks. This information can be used to estimate and balance the likely benefits and risks, given the patient's preferences and tolerances. Importantly, providing this knowledge at the point of care permits clinicians and

Conflict of interest: The authors have declared that no conflict of interest exists.
Reference information: *J Clin Invest.* 2019;129(3):944–945. https://doi.org/10.1172/JCI126120.

patients to make better-informed decisions. The alternative to precision medicine is to base decisions largely on qualitative considerations derived from the clinician's experience. The resulting ambiguities allow choices that do not benefit the patient's health and degrade system efficiency. Precision medicine informs clinician-patient risk assessments so that their belief is consistent with the actual frequencies in each patient's subgroup and avoids decisions biased by ignorance and/or perverse incentives.

Discovery of clinically relevant, mechanistically anchored subgroups

A central component of our thesis is that the ultimate value of defining disease subgroups is realized when they are tied to clinically relevant outcomes, which are usually strongly connected to the underlying mechanisms. Clinically relevant outcomes can be defined from many different perspectives, including (a) analysis of careful phenotypes and longitudinal trajectories; (b) differential outcomes and costs associated with specific interventions; (c) the experience and intuition of wise clinicians; and (d) patients' perspectives. Importantly, mechanistic anchors can be discovered by demonstrating greater homogeneity of measures from multiple domains within subgroups (9), as well as by perturbing the system and observing a coherent response within the subgroup.

We strongly believe that precision medicine must incorporate the broad scales of the data needed to inform a clinical decision. The most important measurements to inform a particular decision might be at the scale of biochemistry, physiology, structure, environment, or behavior of the individual or larger groups. In the early phases of precision medicine's development, it is likely that integrated measures closer to physiology and patient outcomes will better identify subgroups that matter in terms of burden, cost, and outcomes.

Learning health care systems of systems

Much has been written about learning health care systems (1, 10). The Johns Hopkins precision medicine program, called Hopkins *in*Health, is building a learning health system of systems. The core elements of each component system are: clear articulation of questions and decisions that address important unmet medical needs in domain-specific areas; a biomedical model (key variables and their known relationships) that is the foundation for answering the clinical questions in those areas; a science-quality clinical cohort database whose careful analysis provides the evidence to inform the answers; and a means for delivering that evidence, clearly communicated to the point of decision.

We need a system of systems, because the specifics vary widely from one clinical problem to another. To address the question of whether a prostate cancer is aggressive or indolent, key variables include serial measures of biomarkers, serial biopsy reports by core (11, 12), and, more recently, MRI images and a genomic risk score. But none of these assessments is relevant to managing autoimmune disease or diabetes or arrhythmia, which have their own state variables. The approach in each system is common, but the specifics are not. Of course, the domain-specific systems must also interact, as patients typically have multiple health issues. As common complexes of problems emerge, so must system clusters to address them.

To operate a system of systems, there must be a common infrastructure, including the template for generalizing the learning process; an information acquisition, management, analysis, and communication system that each subsystem can rapidly tailor to its needs; and a common business model that incents improved outcomes at more affordable costs, rather than earning more by doing more.

Concluding thoughts

The modern revolutions in measurement, computation, and communication afford an unprecedented opportunity to define clinically relevant and mechanistically anchored disease subgroups at scale. Subgroups will improve the prediction, prevention, and treatment of disease. Several challenges to realizing precision medicine's full potential exist, including the creation and wide adoption of connected technology platforms for the collection and analysis of research-grade clinical data; the construction of cascading incentives that encourage patients, providers, health systems, employers, and payers to participate in a learning health care system of systems; the building of clinical cohort databases comprising research-quality measurements from diverse domains; and research funding to pursue the iterative refinement of disease subgroups and their mechanisms. Precision medicine will only succeed if it couples the emerging tools of this era to real clinical problems and rebuilds medicine on a foundation of 21st century science to produce better health at more affordable costs.

Address correspondence to: Antony Rosen, 5200 Eastern Ave., Room 412 MFL Bldg Center Tower, Baltimore, Maryland 21224, USA. Phone: 443.287.0246; Email: arosen@jhmi.edu. Or to: Scott L. Zeger, 615 N. Wolfe St, Baltimore, Maryland 21209, USA. Phone: 410.502.9054; Email: sz@jhu.edu.

1. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010;2(57):57cm29.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–795.
3. OECD (2018). Life expectancy at birth (indicator). https://doi.org/10.1787/27e0fc9d-en. Accessed on December 20, 2018.
4. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA*. 2012;307(14):1513–1516.
5. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry*. 2012;169(2):141–151.
6. Chibon F, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med*. 2010;16(7):781–787.
7. Igusa T, et al. Autoantibodies and scleroderma phenotype define subgroups at high-risk and low-risk for cancer. *Ann Rheum Dis*. 2018;77(8):1179–1186.
8. Shah AA, Rosen A, Hummers L, Wigley F, Casciola-Rosen L. Close temporal relationship between onset of cancer and scleroderma in patients with RNA polymerase I/III antibodies. *Arthritis Rheum*. 2010;62(9):2787–2795.
9. Joseph CG, et al. Association of the autoimmune disease scleroderma with an immunologic response to cancer. *Science*. 2014;343(6167):152–157.
10. Schneeweiss S. Learning from big health care data. *N Engl J Med*. 2014;370(23):2161–2163.
11. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction of the pathologic Gleason score to inform a personalized management program for prostate cancer. *Eur Urol*. 2017;72(1):135–141.
12. Coley RY, Fisher AJ, Mamawala M, Carter HB, Pienta KJ, Zeger SL. A Bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer. *Biometrics*. 2017;73(2):625–634.