

## VIEWPOINT

# Potential Excessive Testing at Scale

## Biomarkers, Genomics, and Machine Learning

**Kenneth D. Mandl, MD, MPH**

Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts; and Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts.

**Arjun K. Manrai, PhD**

Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts; and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts.



Supplemental content

### Corresponding

**Author:** Kenneth D. Mandl, MD, MPH, Computational Health Informatics Program, Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115 ([kenneth\\_mandl@harvard.edu](mailto:kenneth_mandl@harvard.edu)).

[jama.com](http://jama.com)

**A culture** of advocacy and promotion for aggressive testing may arise when a biomarker or its sequelae yield financial benefit to drug and device manufacturers, procedure-based specialties, hospitals, or laboratory testing services or is increasingly requested by patients. Excessive testing can also lead to costly and harmful care, including false-positive results, overdiagnoses, and unnecessary treatments. Economic pressures, obfuscated intentionally or inadvertently, can drive increased use of biomarkers, a phenomenon that could be termed "biomarkup."

The volume of per-patient biomarker measurements for screening, monitoring, and diagnosing is poised to increase substantially. Furthermore, many of these tests will be directed at consumers.<sup>1</sup> Machine learning algorithms that will soon drive artificial intelligence in health care require large amounts of data and involve ever-expanding approaches to passively and actively capture patient- and clinician-generated data. The affordability of wearables and other connected devices is leading to continuous streams of "digital biomarkers" from individuals in their homes. Genomic measures in clinical care are expanding the number of biomarkers routinely measurable by a physician from a handful to potentially thousands.

In this Viewpoint, we discuss 3 mechanisms through which biomarker-based testing may be manipulated and recommend a systematic approach for recognizing, measuring, and counteracting the phenomenon in the genomic and artificial intelligence contexts.

### Modify the Threshold

Adjusting the threshold of a biomarker for disease definitions may significantly alter the population labeled with treatable conditions. For example, the 2013 change in the cholesterol practice guidelines with new cardiovascular risk calculators increased the number of adults eligible for statin therapy by an estimated 12.8 million<sup>2</sup> compared with previous guideline recommendations (Figure, A). To the public, cholesterol is probably the best known blood test biomarker. With the global statin market approaching \$23 billion, this is not a coincidence. Public awareness campaigns have made cholesterol a household word. Drug companies and practice guidelines educate physicians about the importance of testing and statin prescribing. Professional organizations focus guidelines on a readily available measurement. The Centers for Medicare & Medicaid Services levies financial penalties on health plans in which beneficiaries adhere poorly to filling their statin prescriptions.

Guidelines that lower fasting glucose thresholds signifying prediabetes medicalized a risk factor into a disease,<sup>3</sup> potentially increasing medical care, purchase of devices and supplies, and medication use.

Atrial fibrillation detected by a US Food and Drug Administration (FDA)-approved smartwatch or Parkinson

disease stage defined by a mobile device-based tapping test pose new opportunities to expand disease definitions and drive additional testing and treatment. The march toward value-based care is threatening to clash with a surge in ubiquitous technologies that may lead to additional costly testing. Many of these technologies are being marketed directly to consumers as well as to physicians.

### Increase the Complexity

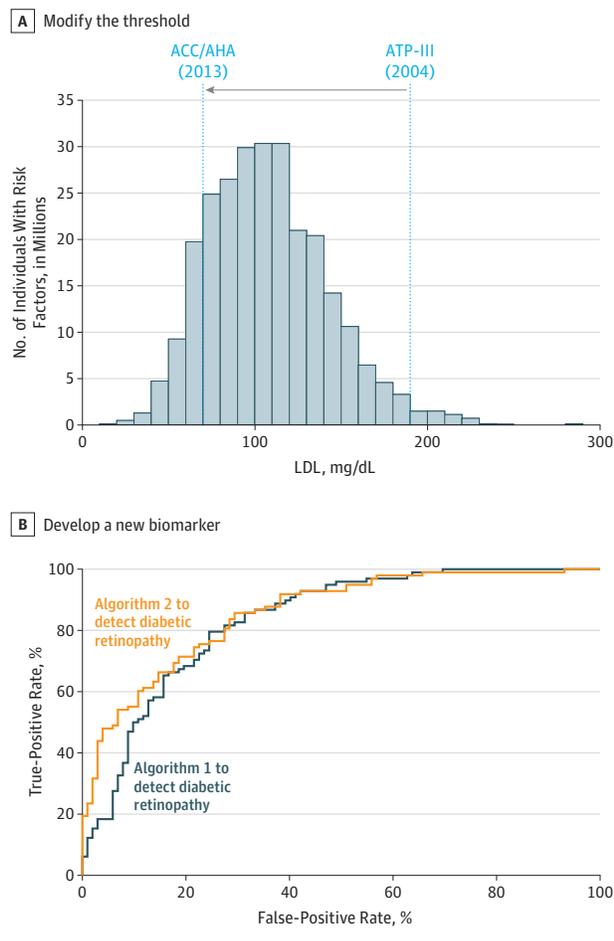
Genomic measurement exemplifies a multiple testing challenge in practice that is difficult to fully address and requires special care with regard to interpretation of biomarkers. The UK National Health Service announced a nationwide genomic medicine service, which at the outset is expected to sequence 30 000 patients per month. Geisenger announced the availability of genomic testing for its patient population. The FDA recently granted marketing authorization to 23andMe for a direct-to-consumer *BRCA1/BRCA2* test. It remains to be seen whether these tests are overused in low-risk populations, for whom positive results are likely to be erroneous and that may drive unnecessary follow-on testing.<sup>4</sup> Past experience predicts persistent strong advocacy and marketing that has already arisen around many genomic tests, most of which lack long-term studies demonstrating robust evidence of improved outcomes or survival (eFigure in the Supplement).

### Develop a New Biomarker

Interested parties, for example a medical subspecialty or a pharmaceutical company, may promote biomarkers that drive use or may even create new biomarkers for the condition of interest. Nonspecific biomarkers can be especially financially advantageous. For example, to promote its osteoporosis drug alendronate sodium (Fosamax), Merck helped develop bone densitometry and establish the Bone Measurement Institute, a not-for-profit company, that worked to increase the number of densitometers and achieve an optimal price point for the test. Infant formula manufacturers, through guidelines and sponsored patient and clinician education, promote specialty formula based on the nonspecific biomarker of feeding intolerance as a diagnostic for cow's milk protein allergy.<sup>5</sup> The American Pain Society, with funding from Purdue Pharmaceuticals, maker of OxyContin, established and trademarked the "Pain: the Fifth Vital Sign" slogan and successfully promoted more pain treatment.<sup>6</sup>

Big data, connected devices, and machine learning are yielding new digital biomarkers, many based on algorithms that are not readily interpretable. Even the simple choice of thresholds for 2 diabetic retinopathy algorithms with similar receiver operating characteristic curves could drive more positive tests and ensuing consultations and procedures (Figure, B). Individual investigators and companies

Figure. Pathways That Promote Increased Application of Biomarkers



A, Statin treatment recommendations changed substantially between the 2004 Adult Treatment Panel (ATP) and the 2013 American College of Cardiology–American Heart Association (ACC/AHA) guidelines. For primary prevention among patients without diabetes and with a 10-year cardiovascular disease risk of 7.5% or more, the guidelines switch amounts to lowering the low-density lipoprotein (LDL) cholesterol threshold from 190 to 70 mg/dL (to convert LDL from mg/dL to mmol/L, multiply by 0.0259). Survey-weighted estimates from the Centers for Disease Control and Prevention National Health and Nutrition Examination Survey (NHANES) 2013–2014 cohort show substantial increases in the population defined as having the risk factor.

B, New digital biomarkers are entering clinical care settings. Even with nearly identical testing characteristics, 2 machine learning algorithms to detect diabetic retinopathy using different thresholds along their respective receiver operating characteristic curves will drive substantially different numbers of ophthalmology referrals and possibly treatments.

are incentivized to make their algorithms appear to be both novel and state-of-the-art. Even if the performance of a new algorithm is superior in contextually appropriate patient cohorts, how thresholds are set in practice may be opaque and counteract the overall utility of the algorithm across diverse populations.

A health care system delivering value must use testing and services judiciously. Processes are needed to ensure that advocacy for a biomarker provides value and not just profit and ensure that educational programs for physicians in the interpretation of biomarker testing will protect patients and support transformation toward a safer, cost-effective health care system. Policy makers, guideline producers, payers, regulators, clinicians, and patients require knowledge to recognize paid advocacy, the coalescing of communities around lucrative tests, and tests that lead to expensive treatment, especially for large populations of asymptomatic individuals. However, influencers are not always free of conflicts of interest. In the context of advocacy for thyroid cancer screening (a demonstrably low-value program), government regulatory agencies and independent bodies such as the US Preventive Services Task Force are at possible risk of being “captured” by dominant, interested parties within the industry sector.<sup>7</sup> As the FDA grapples with emerging digital and genomic products, its regulatory framework should account for the public health consequences of overtesting at scale and should consider the interests of companies seeking approvals.

The parameters of an algorithm, thresholds for a positive test result, or an alert should be transparent and published for users of the device. Even the accuracy of a simple metric like step count can vary widely from device to device, user to user, and day to day. Performance of the tests in populations with differing risk factors, and therefore different prior probabilities of disease, should be documented and updated to promote continuous learning. Importantly, tests marketed directly to consumers will not be captured in medical records. A registry system should be developed to monitor use and results. In clinical genomics, variant annotation has been improved with publicly accessible resources documenting variant frequency data, along with centralized repositories sharing variant pathogenicity assertions, such as the ClinVar database, across previously siloed testing laboratories. Similar registry and data sharing efforts will help elucidate the analytical and clinical validity of the algorithms comprising the impending wave of health-related artificial intelligence and digital applications.

Patients have benefited from advances in biomarker development and selection. So too have drug and device manufacturers, procedure-based specialties, hospitals, laboratory testing services, and patent holders. In the 21st century, as the number of potential biomarkers expands exponentially, it will be important to ensure a system that benefits patients and improves their health.

#### ARTICLE INFORMATION

**Published Online:** February 8, 2019.  
doi:10.1001/jama.2019.0286

**Conflict of Interest Disclosures:** Dr Manrai reported receiving personal fees from doc.ai outside the submitted work. No other disclosures were reported.

#### REFERENCES

- Schwartz LM, Woloshin S. Medical marketing in the United States, 1997–2016. *JAMA*. 2019;321(1):80–96. doi:10.1001/jama.2018.19320
- Pencina MJ, Navar-Boggan AM, D’Agostino RB Sr, et al. Application of new cholesterol guidelines to a population-based sample. *N Engl J Med*. 2014;370(15):1422–1431. doi:10.1056/NEJMoA1315665
- Redberg RF. The medicalization of common conditions. *JAMA Intern Med*. 2016;176(12):1863. doi:10.1001/jamainternmed.2016.6210
- Gill J, Obley AJ, Prasad V. Direct-to-consumer genetic testing: the implications of the US FDA’s first marketing authorization for *BRCA* mutation testing. *JAMA*. 2018;319(23):2377–2378. doi:10.1001/jama.2018.5330
- van Tulleken C. Overdiagnosis and industry influence: how cow’s milk protein allergy is extending the reach of infant formula manufacturers. *BMJ*. 2018;363:k5056. doi:10.1136/bmj.k5056
- Kliff S. The opioid crisis changed how doctors think about pain. *Vox*. <https://www.vox.com/2017/6/5/15111936/opioid-crisis-pain-west-virginia>. Published June 5, 2017. Accessed January 30, 2019.
- Welch HG. Cancer screening, overdiagnosis, and regulatory capture. *JAMA Intern Med*. 2017;177(7):915–916. doi:10.1001/jamainternmed.2017.1198