Predictive Analytics
Helping Guide the Implementation Research Agenda at the National Heart, Lung, and Blood Institute

Check for updates

Michael M. Engelgau*, Muin J. Khoury[†], Rebecca A. Roper*, Jennifer S. Curry*, George A. Mensah*
Bethesda, MD, USA; and Atlanta, GA, USA

Global biomedical research investments have resulted in important discoveries and many effective preventive and therapeutic interventions for heart, lung, blood, and sleep (HLBS) disorders—the focus areas for the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health (NIH). In most high-income countries, extensive efforts have synthesized this knowledge through systematic reviews and developed evidence-based guidelines for intervention delivery within health care systems and community-based settings [1—7]. In low- and middle-income countries, the World Economic Forum and World Health Organization have studied the economic toll of noncommunicable diseases and the cost of scaling up a set of proven-effective interventions, so-called best buys (e.g., providing drug therapy and counseling for eligible persons at high risk to prevent heart attacks and strokes) [8]. Thus, effective interventions are now available and recommended for implementation across the globe. However, adaptable and sustainable implementation strategies for HLBS interventions are lacking and return on investment for this vast knowledge base is diminished. This has led many to call for a more developed implementation research agenda—including leadership from institutions such as the World Health Organization [9], World Bank [10], academia [11], US Agency for International Development [12], and NIH [13—19].

Implementation research studies intervention delivery in real-world contexts of health care delivery systems and community public health systems, with the goal of delivering interventions optimally and sustainably across the entire socioecological context [20]. Several major challenges for implementation research include understanding key barriers and facilitators within the socioecological context, various health and community policies, delivery strategies within health systems (e.g., physical infrastructure, staffing and their skill levels, availability of diagnostics and therapeutics, geospatial issues), and community contexts (e.g., disease burden, community resources, social deprivation, and economic opportunities). Elements at each of these levels can impact intervention implementation and thus enhance or impair intervention delivery and outcomes.

Predictive analytics [21] offers novel approaches to use methodologies that could pinpoint key barriers and facilitators across the community context and lend insights as to what might be highly promising implementation strategies for research agendas. Implementing the NHLBI Strategic Vision [22]—where implementation research is prominent—will potentially benefit from predictive analytics informing the research agenda. In this perspective, we briefly review the opportunities and challenges of predictive analytics for refining the implementation research agenda targeted at reducing the population burden of HLBS disorders.

WHAT IS PREDICTIVE ANALYTICS?

Predictive analytics in health is a set of analytic procedures that take existing information and forecast future probabilities of disease patterns using population- and individual-level data along with biomedical and other types of data [21-24]. These annotated data can then drive analytic models that can make health decisions [24]—at both the individual and population levels. This is much different from the traditional hypothesis-driven biomedical research. Predictive analytics uses statistical algorithms, comprehensive data (e.g., geospatial, burden of disease, demography, variation in community and health care capacity and in local resources settings), and strives to understand complex interrelationships between determinants of health and the variability of health care and public health service delivery—and the likelihood of future health outcomes. Predictive analytics could help forecast potential solutions for populations, specifically vulnerable population subgroups, by simulating implementation strategies and then exploring the facilitators and challenges in specific contexts—while simultaneously considering available resources and capacity.

HOW PREDICTIVE ANALYTICS IS USED IN HEALTH

Currently, predictive analytics is used mostly as a business tool and now increasingly in health. Businesses may use current and past data to better understand their customers, products, and competitors while also identifying potential opportunities and risks [21]. This approach is not hypothesis driven per se. Rather, the data drive the analytical direction seeking novel relationships. For health, understanding how the community context can facilitate delivery of health interventions is key. Predictive analytics can simulate a wide variety of clinical-, systems-, and population-level delivery interventions and also predict outcome measures and forecast results from selected intervention delivery scenarios. These efforts are part of larger movements attempting to understand how best to

TABLE 1. Selected examples of predictive analytics use at the individual, population, and health system levels

Target Level	Objectives of Predictive Analytics Study	Scope of Data Included	Actions/Decisions
Individual	To construct a genomic risk score for CVD and to estimate its potential as a screening tool for primary prevention [30,33].	Large-scale genomic-wide data combined with targeted genetic association data.	Complements use of conventional risk factor predictions and predicts future CVD trajectory for clinical level prevention and treatment.
	To determine pharmacist interventions that maximize benefits and minimize risks from medications while improving outcomes [34].	Medical and medication records, health system claims data, physician notes.	Tailors medications use to individual benefits within the health system.
Population	To develop a predictive algorithm to estimate 5-yr risk of incident CVD in community settings [28].	Longitudinal population surveys.	Population risk levels for policy decisions.
	To use expressed Twitter language to characterize community-level psychological correlates of age-adjusted mortality from CVD [35,36].	Twitter language content, county-level mortality rates, census data.	Local CVD health intervention prioritization.
Health system	To determine population-level health experiences, population health, and health care costs [29,37].	Health system, social, environmental, and geographic data; community health status and vulnerabilities.	Targets resources to have largest impact.

CVD, cardiovascular disease.

align areas such as precision health, precision public health, and population health [25–27]. Predictive analytics can help determine efficient implementation research strategies that then can be tested in delivery systems.

Examples of predictive analytics vary broadly—from uses in individuals, populations, and health systems and can incorporate large amounts of data from a single or several sources with the goal of predicting disease risk or health events [24,28–37] (Table 1). The spectrum of data sources can range from genomics and epidemiology to social media and provide a range of results that include identifying disease burdens and risk factors, high health risk states, and strategies to target resources at high-risk or high-burden groups. A common theme is the use of large amounts of data, and in some cases, multiple types of data—beyond traditional health-related data—and to use them in innovative ways.

## SOME LIMITATIONS OF PREDICTIVE ANALYTICS

We need to better understand how prediction models actually impact health care decisions, patient outcomes, cost, and quality of care. While we can be thoughtful in identifying health information that is enlightening, whether it triggers a health decision action, can be challenging [38]. Other technical modeling challenges include data quality used for the analyses and its heterogeneity, model calibration (not able to account for the effects of unmeasured covariates that can lead to misleading results), and finally, gaining trust of the providers and consumers new to using this type of information and changing behaviors.

The challenges with the use of social media data need careful considerations. Limitations include sample bias due

to highly variable participation rates across age and population groups and the methods to extract useful content among large volumes of data that contain little relevant content [35,36]. Beyond sample bias challenges, interventions delivered by social media can also have short-term backlash when users realize they do not control the information they receive. Medium-term repercussions can include potentially driving unhealthy behaviors underground (e.g., participants will defer from tweeting unhealthy behaviors and/or develop or obfuscate it to avoid social media targeting). Finally, longer-term risks may include personal behavior pattern changes with what they share in the public social media domain.

HOW CAN PREDICTIVE ANALYTICS HELP ADVANCE IMPLEMENTATION RESEARCH

Implementation research is focused on the portion of the translational research spectrum that takes proven-effective interventions and studies their implementation in the actual context where they will be delivered [39] (Fig. 1). With T4 translation research, all the real-world factors that can facilitate implementation—or those that can be barriers to it—are critically important. For example, these can include the following domains: individual and patient factors; family dynamics; community, social, and physical environment; geographic location (e.g., urban/rural); social determinants of disease; educational and economic opportunities; social and physical mobility; and social and public health policy. These domains all can contribute to health risk, disease status, or health outcomes at the individual and at the population levels—and impact implementation efforts. They also are the drivers that will
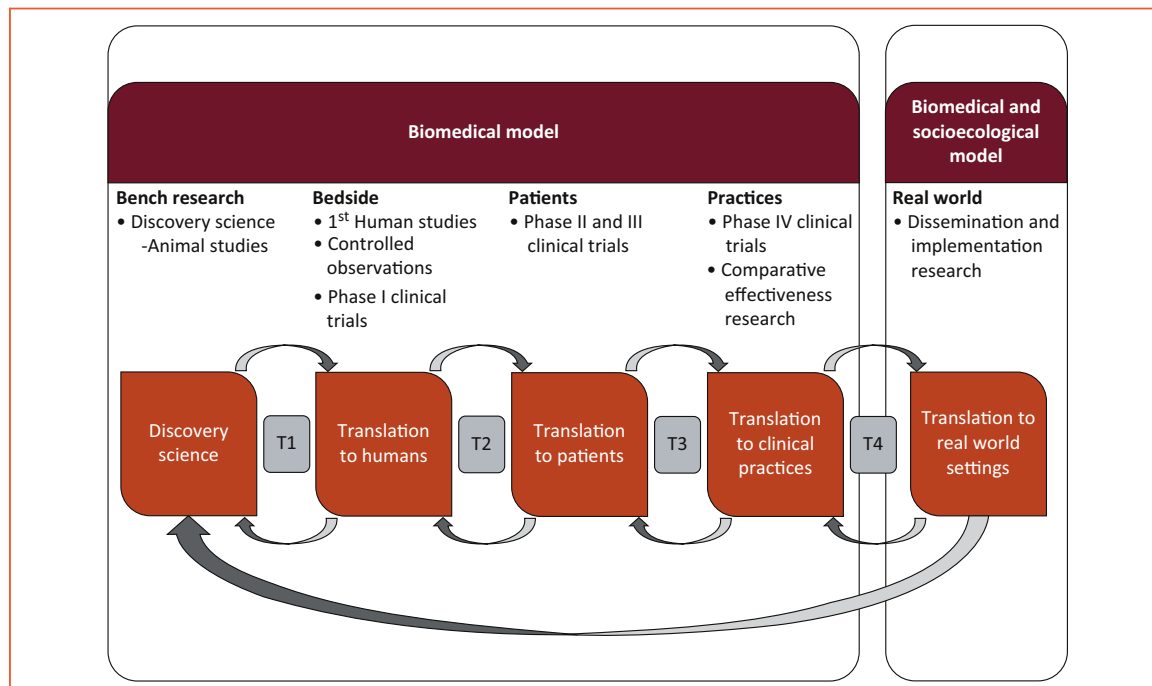
FIGURE 1. The translation spectrum [39] and T4 translation research. T1, bench to bedside or animal studies to humans; T2, translating science discovery to patients with specific diseases; T3, translating clinical insights to service delivery in clinical practices; T4, translating effective interventions to real-world settings.

enhance, and challenges that slow implementation of health interventions at both the individual and community levels.

As we found in the health examples using predictive analytics (Table 1), the ability to assimilate large and diverse data can help understand how these domains combine and interact across the socioecological spectrum from the individual to the community, and to the complex system-level environment. These should enlighten implementation strategies and focus in on identifying those at risk or with disease burden and uncover the key levers needed for successful intervention delivery. In characterizing complex systems in which people live and work, this will also provide better approaches to identify and address health inequities—that is, preventable health disadvantage due to poor socioeconomic and social determinants— driven by factors that go beyond those in the health sector [40]. Predictive analytics approaches can incorporate data that describe the burden of risk or disease in the context of the community environment while considering locale, assets, resources, and opportunity.

DISCUSSION

Predictive analytics provides a methodological approach that could be valuable in developing more refined implementation research agendas. It is currently becoming of great interest to NIH and NHLBI and was the topic of a recent NIH methodology seminar and is the topic for a NHLBI workshop planned for April 2019. It may help identify high-burden communities where health inequities are common, reveal important levers that can yield maximum return on benefit, and determine factors and drivers of future adverse health outcome, both clinical and public health, at the individual and population levels. It may also help tackle the complex systems where individuals and populations receive their clinical and public health care and also live, work, and play. All this information will help inform the implementation research agenda for NHLBI. However, the scope of predictive analytics is broader than described here. Precision medicine and precision public health are leveraging this tool along with other tools that employ big data approaches such as machine learning, artificial intelligence, neural networks, and data mining [21,25,26,41]. Use of these approaches can be tailored for the individual or population. The use of predictive analytics may enhance the emerging field of precision public health whose goal is to predict and understand population risks and tailor interventions to populations that are most likely to benefit from a specific intervention [25]. More broadly, it has been noted to have attributes that improve our ability to prevent disease, promote health, and reduce health disparities in populations by applying emerging methods and technologies for measuring disease, pathogens, exposures, behaviors, and susceptibility in populations; and developing policies

and targeted implementation programs to improve health [27]. Some learning from delivering individual-based precision medicine may actually be very beneficial and complementary to successfully delivering precision public health. For example, precision public health delivery may benefit from combining individual- and population-level prevention efforts because individual genetically targeted approaches can have population level impact (e.g., familial hypercholesterolemia). In addition, precision medicine methods can incorporate big data methods and be useful to help define new approaches for precision public health [27]. However, others have cautioned that precision public health should not focus solely on genetics—this may detract from the impact of considering broader social determinates [26]. Other challenges include ethical, legal, and privacy issues associated with accessing and analyzing big data from individuals and groups without consent. These same issues would challenge delivery of interventions through predictive analytics approaches. In sum, the benefits and the risks associated with a predictive analytic approach will both need to be weighted.

CONCLUSIONS

To date, research investments have yielded many interventions that we know are highly effective. Yet, they are not being delivered to have individual- or population-level impact. Implementation research can provide a means to determine optimal and sustainable strategies to deliver these interventions. Calls have been made for more efforts in this research area. Novel tools such as predictive analytics are currently tapping into big data and developing prediction models for health outcomes at the individual, population, and system levels. These models may incorporate big data from many domains and simulate the complex environments where interventions are delivered. These simulations may prove to be a valuable asset because they may be able to refine and make more efficient the implementation research agenda. This becomes even more important under current scenarios of limited resources. In conclusion, implementing the NHLBI Strategic Vision and its focus on implementation research will benefit from predictive analytics informing our implementation research agenda as we chart the future together.

REFERENCES

1. Bero LA, Grilli R, Grimshaw JM, et al. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. BMJ 1998;317:465–8.
2. United States Preventive Services Task Force. Recommendations for Primary Care Practice. Available at: https://www.uspreventiveservicestaskforce.org/Page/Name/recommendations. Accessed December 20, 2016.
3. The Community Guide. Your Online Guide of What Works to Promote Healthy Communities. Available at: https://www.thecommunityguide.org/. Accessed October 18, 2018.
4. Yawn BP, Buchanan GR, Afenyi-Annan AN, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. JAMA 2014;312:1033–48.
5. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018;71:1269–324.
6. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;63:2889–934.
7. US Department of Health and Human Services. Agency for Healthcare Research and Quality. Guidelines and Measures. Available at: https://www.ahrq.gov/gam/index.html. Accessed October 18, 2018.
8. World Health Organization. Tackling NCDs: "Best Buys" and Other Recommended Interventions for the Prevention and Control of Noncommunicable Diseases. Available at: http://apps.who.int/iris/bitstream/10665/259232/1/WHO-NMH-NVI-17.9-eng.pdf?ua=1. Accessed December 26, 2017.
9. World Health Organization. Implementation Research Toolkit. Available at: http://www.who.int/tdr/publications/year/2014/9789241506960_workbook_eng.pdf. Accessed December 26, 2017.
10. World Bank. Remarks as Prepared for Delivery: World Bank Group President Jim Yong Kim at the Annual Meeting Plenary Session. Available at: http://www.worldbank.org/en/news/speech/2012/10/12/remarks-world-bank-group-president-jim-yong-kim-annual-meeting-plenary-session. Accessed December 1, 2016.
11. Kruk ME. More health for the money-toward a more rigorous implementation science. Sci Transl Med 2014;6:245ed17.
12. US Agency for International Development. USAID's Implementation Science Investment: Improving HIV/AIDS Programming Through the Translation of Research to Practice: PEPFAR's Approach to Promoting Evidence-Based Programming. Available at: https://www.usaid.gov/sites/default/files/documents/1864/Implementation_Science_Brief_5_2.pdf. Accessed February 10, 2019.
13. National Institutes of Health, Fogarty International Center. Implementation Science Information and Resources. Available at: https://www.fic.nih.gov/researchtopics/pages/implementationscience.aspx. Accessed December 26, 2017.
14. Engelgau MM, Peprah E, Sampson UK, Mensah GA. A global health strategy to capitalize on proven-effective interventions for heart, lung, and blood diseases. Glob Heart 2015;10:87–91.
15. National Institutes of Health, National Cancer Institute, Division of Cancer Control and Population Sciences. Implementation Science. Available at: https://cancercontrol.cancer.gov/IS/. Accessed December 29, 2017.
16. National Institutes of Health. Department of Health and Human Services: Part 1. Overview Information: Dissemination and Implementation Research in Health. Available at: https://grants.nih.gov/grants/guide/pa-files/par-16-238.html. Accessed December 29, 2017.
17. National Institutes of Health. Department of Health and Human Services: Part 1. Overview Information: Implementation Science for the Prevention and Treatment of Mental and/or Substance Use Disorders in Low- and Middle-income Countries. Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-MH-17-650.html. Accessed December 27, 2017.
18. National Health Lung and Blood Institute. Center for Translation Research and Implementation Science. Available at: https://www.nhlbi.nih.gov/about/scienfic-divisions/center-translation-research-and-implementation-science. Accessed December 29, 2017.
19. Engelgau MM, Peprah E, Sampson UKA, et al. Perspectives from NHLBI Global Health Think Tank Meeting for late stage (T4) translation research. Glob Heart 2017;12:341–8.
20. Sampson UK, Chambers D, Riley W, Glass RI, Engelgau MM, Mensah GA. Implementation research: the fourth movement of the unfinished translation research symphony. Glob Heart 2016;11:153–8.

21. Miner l, Bolding P, Hilbe J, et al. Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research. New York, NY: Elsevier, Academic Press; 2015.

22. Mensah GA, Kiley J, Mockrin SC, et al. National Heart, Lung, and Blood Institute Strategic Visioning: setting an agenda together for the NHLBI of 2025. Am J Public Health 2015;105:e25–8.

23. Beal V. Predictive Analytics. Available at: https://www.webopedia.com/TERM/P/predictive_analytics.html. Accessed October 31, 2018.

24. Zhu L, Zheng W. Informatics, data science, and artificial intelligence. JAMA 2018;320:1103–4.

25. Dolley S. Big data's role in precision public health. Front Public Health 2018;6:68.

26. Chowkwanyun M, Bayer R, Galea S. "Precision" public health - between novelty and hype. N Engl J Med 2018;379:1398–400.

27. Khoury MJ, Galea S. Will precision medicine improve population health? JAMA 2016;316:1357–8.

28. Manuel DG, Tuna M, Bennett C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). CMAJ 2018;190:E871–82.

29. Califf RM, Harrell FE Jr. Individual risk prediction using data beyond the medical clinic. CMAJ 2018;190:E947–8.

30. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. J Am Coll Cardiol 2018;72:1883–93.

31. Janke AT, Overbeek DL, Kocher KE, Levy PD. Exploring the potential of predictive analytics and big data in emergency care. Ann Emerg Med 2016;67:227–36.

32. Russell J. Predictive analytics and child protection: constraints and opportunities. Child Abuse Negl 2015;46:182–9.

33. GWAS to the people. Nat Med 2018;24:1483.

34. Hernandez I, Zhang Y. Using predictive analytics and big data to optimize pharmaceutical outcomes. Am J Health Syst Pharm 2017;74:1494–500.

35. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018;1:47.

36. Eichstaedt JC, Schwartz HA, Kern ML, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015;26:159–69.

37. Miranda ML, Ferranti J, Strauss B, Neelon B, Califf RM. Geographic health information systems: a platform to support the "triple aim". Health Aff (Millwood) 2013;32:1608–15.

38. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: recalibrating expectations. JAMA 2018;320:27–8.

39. Mensah GA. A new global heart series—perspectives from the NHLBI. Glob Heart 2013;8:283–4.

40. Delpierre C, Kelly-Irving M. Big data and the study of social inequalities in health: expectations and issues. Front Public Health 2018;6:312.

41. Psaty BM, Dekkers OM, Cooper RS. Comparison of 2 treatment models: precision medicine and preventive medicine. JAMA 2018;320:751–2.