Editorial

dmj
DIABETES & METABOLISM JOURNAL

Check for updates

Biomarker Score in Risk Prediction: Beyond Scientific Evidence and Statistical Performance

Heejung Bang[1,2]

[1]Division of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, CA,
[2]Clinical and Translational Science Center & Center for Healthcare Policy and Research, Davis School of Medicine, University of California, Sacramento, CA, USA

Common objectives in research are to (1) explain and (2) predict [1,2]. In practice, these are generally inter-dependent; for example, sound prediction relies upon clear explanation (e.g., causal mechanism), and comprehensible explanation is informed by predictors such as correlation and association.

From clinical and public health perspectives, type 2 diabetes mellitus has emerged as a critical concern in all regions of the world, including Asia. Current theory is that diabetes in Asian populations presents somewhat differently compared with Western populations, e.g., "Asian body mass index (BMI)" [3,4].

Notably, multiple studies, including randomized controlled trials, have demonstrated that it may be possible to prevent or delay diabetes onset. While prevention may be considered the highest goal, it could be costly and time-consuming at the population level; benefits are in the future and costs are immediate [5]. Diabetes is often accompanied by other conditions or adverse events that can be used for purposes of early prediction and treatment. Since diabetes or prediabetes may be present but undiagnosed, research methods to identify or screen 'current' at-risk individuals are as needed as those aimed at predicting and preventing 'future' cases. In fact, the question of how to identify those at high risk for diabetes has been studied for a few decades, resulting in publication of a number of prediction models and risk or screening scores. These tools may be tailored to accommodate specific populations or subgroups; for example, American Diabetes Association diabetes risk test, Centers for Disease Control and Prevention prediabetes screening test, Korean diabetes risk score, among others [6-8]. How-

ever, given the volume of models/scores and wide range of predictors, clinicians and patients may find it difficult to select the best tool or they ignore them altogether, citing "research for research sake," perceived lack of cooperation among investigators and with users, or frequently changing and often inconsistent results as reasons [9-12]. For example, more than 350 models for chronic obstructive pulmonary disease and for cardiovascular disease have been reported [11,12].

In this issue of the journal, Wang et al. [13] report on a proposed biomarker risk score to predict diabetes using the Singapore Chinese Health Study, a population-based cohort study, with a nested matched case-control design, using standard risk factors and (scientifically supported) biomarkers as predictors. The study employed conditional logistic regression for data analysis, area under the curve (AUC) P value and other information and re-classification statistics as evaluation measures, and internal validation and sensitivity analyses; these approaches are well accepted standards in research practice, despite some controversies in some measures [14]. No external validation, data limitations (e.g., no waist), uncertain cost implications, and future research direction (e.g., determination of cut-off values) were correctly discussed as potential limitations, indicating the authors' attempts at a balanced report. Here, I share several observations and viewpoints for readers pursuing similar lines of research.

First, the authors rightly mentioned cost and cost-effectiveness issues. Decisions involve trade-offs. While development of a biomarker test may create jobs, identify patients at risk and

---

Corresponding author: Heejung Bang https://orcid.org/0000-0002-5755-9951
Division of Biostatistics, Department of Public Health Sciences, School of Medicine,
University of California, One Shields Avenue, Med Sci 1C, Davis, CA 95616, USA
E-mail: hbang@ucdavis.edu

https://e-dmj.org

improve the public health, it will also incur costs (and possible other side effects). Despite the potential value of a statistically significant observation, cost might outweigh benefit; this will be particularly relevant in resource-limited settings. Despite expensive cutting-edge, high-tech or high-dimensional methods or best-fitting models, clinical or practical improvements may be disappointing [15-17] or even found to be no better than well known, easily obtained information at low or no cost, such as family history, BMI, or age. In that case, we may become "cost boosters" despite good intentions and the scientific knowledge gained.

Second, beyond prediction, what is the next step? After obtaining a set of predictors, odds/risk ratio (beta coefficient), risk estimate, or possibly risk status (say, high, or low), "so what?" A clinician (or patient) might use the information to recommend medication or additional tests, modified diet, increased exercise, or for emotional or financial preparedness. This point may best explain why so many prediction models are not used in the real world, in that there is no clear next step. Moreover, people tend not to worry much (or over-worry) about future events, and the importance of probability is perceived differently among individuals; for example, whether a 15% risk over 5 to 10 years is low or high [15,18].

Third, there are many biomarkers, and they are generally correlated, and levels of a variety of biomarkers would increase prior to the manifestation of glucose intolerance. It is entirely possible that different patient populations will be characterized by dissimilar sets of biomarkers, say, different "final four." Understanding and use of these biomarkers or anything novel requires expert clinical interpretation (vs. patient's self-assessment or learning). In implementation, the usual issues related to measurement and variable definitions come into play: different units (e.g., mg/dL vs. mg/L vs. mmol/L); continuous vs. categorical variables (e.g., yes/no or high/middle/low); not applicable or missing/censored/mismeasured (randomly or informatively) data. In addition, some biomarkers (such as ferritin) may be unknown to most non-experts, so that even a single variable may be a barrier or deal breaker in real world adoption of a model. Model developers and clinicians should ask if biomarker A is really better than biomarker B (e.g., glycosylated hemoglobin), considering the pros and cons of each. Models can compete, evolve, or be upgraded naturally as with scientific knowledge, but there may be unintended consequences related to confusion and mistrust. Nevertheless, these could be opportunities for patient-clinician communication

toward shared decision making and patient empowerment. Some clinicians view diabetes as a metabolic disease versus glucose disease and may treat biomarkers accordingly without need of a score. Clinicians may be called upon to explain biomarkers to patients—what they are, what their levels signify, why a given medication is needed, or even why disease definition can differ by sex/race/age/location. A patient's health education, including disease awareness and behavioral changes, may actually be stymied by highly technical and complex information. Also, which variables to use and how to use them in a risk model/score is an art, beyond science. While traditional regression may be well suited to education and simple and explainable models, black-box or push-the-button approaches (e.g., machine learning, artificial intelligence) can certainly have advantages in prediction or diagnostics. They should not compete [2,18-20].

Fourth, the terms "predictable," "preventable," "modifiable," and "actionable" are not synonymous. While most biomarkers are modifiable or may serve as a mediator or surrogate marker, medical history and socioeconomic status are essentially unmodifiable despite high predictive value. As Hume (1748) stated, "*The only immediate utility of all the sciences is to teach us how to control and regulate future events through their causes.*" Although journal editors and reviewers of prediction models tend to be too focused on AUC, too high AUC may imply "too late," such that predictor and outcome are virtually the same thing, or that a predictor must be just another (correlated) outcome or early marker of disease onset. Context and purpose always matter, beyond statistical performance (e.g., higher AUC, larger odds ratio, lower *P* value, largest sample size) in the prediction race [2,18,19].

Finally, the number #1 reason for any invention is need. Why need, for whom, when, etc. My personal motto for prediction modeling or method development as a statistician with heath economics/services and epidemiology training is: "*As long as some use or are willing to pay, a model can be a success.*" Researchers developing future models should emphasize this 'translational' aspect and practical use; perhaps a reasonable short-term goal could be intent to implement at the authors' own institution(s). The study of risk prediction is widespread in medicine and public health, with ever increasing availability of data and powerful computer and easy statistical modeling [11], but it should encompass fundamental principles of business/marketing and engineering. Without a consumer or user, a risk score is just another regression model, supplier of a bag

of odds/risk/hazard ratios, or one more paper in curriculum vitae (including my own!).

Regarding prediction modeling, Breiman stated: "*My attitude toward new and/or complicated methods is pragmatic. Prove that you've got a better mousetrap and I'll buy it. But the proof had better be concrete and convincing*" [2,20]. My hope is that we can maintain a disciplined, focused but pragmatic approach to conducting studies that will impart real knowledge, impact, use or value to diabetes research worldwide.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

REFERENCES

1. Shmueli G. To explain or to predict? Statist Sci 2010;25:289-310.
2. Breiman L. Statistical modeling: the two cultures. Statist Sci 2001;16:199-231.
3. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363:157-63.
4. Matsushita K, Tang O, Selvin E. Addressing challenges and implications of national surveillance for racial/ethnic disparities in diabetes. JAMA 2019;322:2387-8.
5. Haddix AC, Teutsch SM, Corso PS. Prevention effectiveness: a guide to decision analysis and economic evaluation. Oxford: Oxford University Press; 1996.
6. Poltavskiy E, Kim DJ, Bang H. Comparison of screening scores for diabetes and prediabetes. Diabetes Res Clin Pract 2016;118:146-53.
7. Ha KH, Lee YH, Song SO, Lee JW, Kim DW, Cho KH, Kim DJ. Development and validation of the Korean diabetes risk score: a 10-year national cohort study. Diabetes Metab J 2018;42:402-14.
8. Lee YH, Bang H, Kim HC, Kim HM, Park SW, Kim DJ. A simple screening score for diabetes for the Korean population: development, validation, and comparison with other scores. Diabetes Care 2012;35:1723-30.
9. Wyatt JC, Altman DG. Prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539.
10. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? J Am Coll Cardiol 2003;42:851-3.
11. Adibi A, Sadatsafavi M, Ioannidis JPA. Validation and utility testing of clinical prediction models: time to change the approach. JAMA 2020 Mar 5 [Epub]. https://doi.org/10.1001/jama.2020.1230.
12. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, Schlussel MM, Gerry S, Black JA, Heus P, van der Schouw YT, Peelen LM, Moons KG. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.
13. Wang Y, Koh WP, Sim X, Yuan JM, Pan A. Multiple biomarkers improved prediction for the risk of type 2 diabetes mellitus in Singapore Chinese men and women. Diabetes Metab 2020;44:295-306.
14. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology 2014;25:114-21.
15. Chow SC. Encyclopedia of biopharmaceutical statistics. 3rd ed. Boca Raton: CRC Press; 2010. Chapter 95, Cost-effectiveness analysis.
16. Fojo T, Grady C. How much is life worth: cetuximab, non-small cell lung cancer, and the $440 billion question. J Natl Cancer Inst 2009;101:1044-8.
17. Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, Garrett-Mayer E, Herbst RS, Lilenbaum RC, Sima C, Venook AP, Gonen M, Schilsky RL, Meropol NJ, Schnipper LE. American Society of Clinical Oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277-80.
18. Lee YH, Bang H, Kim DJ. How to establish clinical prediction models. Endocrinol Metab (Seoul) 2016;31:38-44.
19. Goldstein DG, Gigerenzer G. Fast and frugal forecasting. Int J Forecast 2009;25:760-72.
20. Raper S. Leo Breiman's "two cultures". Significance 2020;17:34-7.