Biomarker research is at the forefront of the quest toward personalized medicine. It is hoped that the discovery of new biomarkers will aid in better understanding of disease pathways and in turn will lead to improved stratification of individuals at risk and to disease prevention. This avenue of research is actively pursued in all major domains of modern medicine, including, but not limited to, cardiovascular research (1) and cancer research (2). Risk stratification is usually based on the probability that a person has or will develop the event of interest within a prespecified time interval. These probabilities are usually derived from risk-prediction models that use regression-based approaches as well as nonparametric techniques (3). Published reporting guidelines have outlined what is expected of research reports that describe the impact of novel biomarkers (4) and genetic factors (5).

The ability to improve risk-prediction models is one of the key elements used to evaluate new biomarkers. Numerous metrics have been proposed to assess model performance and quantify its improvement. The area under the ROC curve (AUC), a frequently reported standard that originated in the diagnostic setting, equals the probability that given any 2 randomly selected individuals, one with an event (a future event in prognostic and a current event in diagnostic applications) and one without an event, the one with the event has a higher predicted risk. The AUC can be visualized as the area under the plot of sensitivity as a function of 1 minus specificity across all risk thresholds. Improvement in model performance can be quantified by using the increase in the AUC when the new biomarker is added to the risk model. An alternative measure of discrimination that focuses on absolute risks is called the discrimination slope. It is calculated as the difference in mean risks: that for people with events minus that for people without events. The difference in discrimination slopes, known as the integrated discrimination improvement, is used to measure the incremental value of new markers (6). For situations in which meaningful risk thresholds exist, one may focus on metrics that incorporate these thresholds into the assessment of improved performance. For example, according to the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) guidelines for primary prevention of coronary heart disease, individuals with a 10-year risk of coronary heart disease of >20% are considered at high risk, those with a 10-year risk of <6% are considered at low risk, and individuals with a risk between 6% and 20% are considered at “intermediate” risk (7). In this setting, improvement in model performance can be measured by increases in sensitivity for a fixed value of specificity or the net reclassification index, which quantifies the net proportions of people with and without events whose category assignment improved when the new biomarker or biomarkers were used (6).

The matched case control design is an approach commonly used to study the potential impact of new biomarkers and genetic factors. The choice of a case control design over a cohort design is based on the lower cost and increased feasibility of the former, especially when the incidence or prevalence of the event of interest is low.

The reasons for the matched case control design’s popularity are well summarized by Rose and van der Laan (8): “Matched sampling leads to a balanced number of cases and controls across the levels of the selected matching variables. This balance can reduce the variance in the parameters of interest, which improves statistical efficiency.” These authors note, however, that matching can lead to reduced efficiency in some circumstances, as was shown by Rothman and Greenland (9). Furthermore, matching creates a sample of controls that is not representative of the population (9) and can introduce bias. In general, it is recommended that the number of matching variables be reduced to an absolute minimum and that adjustments be made for the other potential confounders (10).

The above considerations and limitations of the matched case control design have been raised mainly in the context of the association of the exposure of interest with the outcome. In this issue of Clinical Chemistry, Pepe and coworkers (11) raise another important
problem with the use of matched case control studies, but this time with the focus on the assessment of the potential of new biomarkers to improve risk-prediction models. The authors assess the impact of the matched case control design on measures of model performance. Using data simulated under the assumption of normality, they were able to calculate the AUC and sensitivity at a fixed specificity, first theoretically and then separately for their matched case control sample. They observed that the impact of the new marker on model performance can be substantially underestimated when a matched case control design is used. On the other hand, the incremental value of the same marker added to a model with standard risk factors tends to be grossly overestimated (an AUC increment of 0.02 vs 0.17). These results are not specific to any given performance metric; the authors state that they were similar for all of the performance measures outlined above. Not surprisingly, the extent of bias depends on the correlation between the new marker and the matching variables. The practical example presented further illustrates the problem.

It is important to stress the large magnitude of the bias and its potential implications: The authors have demonstrated how reliance on matched case control studies can lead to the selection of biomarkers with less of a potential for improvement in model performance. Fortunately, it is possible to remedy the problem. The authors suggest a method to correct the bias due to the matched case control nature of the data. It is based on correcting the logistic regression score intercepts to calculate absolute risk, a technique well known in epidemiology (12). Given the results the authors present, the approach appears to work well and achieves its goal of removing the bias.

The study by Pepe et al. (11) addresses an important problem that applies to all biomarker or genetic studies that are based on matched case control data. The results are not surprising but are often overlooked by researchers who design and conduct them. The resulting bias is an effect of 2 inherent features of the matched case control design—(a) the decrease in the predictive ability of standard predictors, especially the ones used for matching, and (b) the nonrepresentative nature of the control sample. The first feature is related to the inverse relationship between the discriminatory capacity of the baseline model and the ability of new markers to improve it: The same marker can have a much more pronounced impact if the baseline model is weaker in terms of its discrimination ability (13). The second feature is a well-known epidemiologic phenomenon, which might have implications beyond those discussed here (9).

The implications of ignoring the problem are serious. The wrong choice of markers recommended for further testing and evaluation slows down the discovery process and may lead to the abandonment of more promising predictors, only because they “lost” in terms of their incremental value to weaker competitors because of the matched case control study design. From the opposite angle, new resources are likely to be wasted when markers with an apparently strong performance according to the case control data are deemed worthy of further studies when in fact their true ability to improve model performance is limited. The last problem might partially explain the limited yield of biomarkers and genetic factors when added to risk-prediction models (1, 14).

In conclusion, we strongly recommend that researchers involved in biomarker or genetic studies based on matched case control data be aware of the problems outlined by Pepe et al. (11) and at the very least apply the correction suggested.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References


