Debunking a Widely Used Methodology for Assigning Importance to Biomarkers

July 21, 2014

ND Weber

An important component in making medical care decisions is risk prediction. Being able to accurately foretell the likelihood of an event, such as the appearance of a disease, based on predictive factors like biomarkers, age, sex, or family history of illness is becoming more and more common in health care. But how do we identify biomarkers that would improve upon current models used for risk prediction? Since 2008, the net risk reclassification index (NRI), a method for measuring the usefulness of adding a new biomarker or set of biomarkers to a current risk prediction model, has been widely used. However, a recent study carried out by Drs. Margaret Pepe, Holly Janes, and Christopher Li of the Public Health Sciences Division questions the validity of this method for assigning importance to new biomarkers. Their study was published in the *Journal of the National Cancer Institute*.

The NRI was devised as an alternative to the traditional way of determining the usefulness of a new biomarker for improving risk prediction. Previously, researchers considered the improvement in the area under the receiver operating characteristic curve (ΔAUC), a statistical measure of discrimination, when a new biomarker was added to the original risk model. However, promising new biomarkers often failed to produce large increases in the area under the curve. The alternative method, the NRI, is based on combining probabilities for increases in predicted risks for subjects who have events together with decreases in predicted risks for subjects who do not have events, both scenarios that are favorable for a biomarker. In the six years since its introduction, the NRI has gained traction in research, particularly in cardiovascular research, but also increasingly in cancer research publications. "In 2013 alone, the statistic appeared in almost 600 papers including papers published in the most prestigious medical journals," said Dr. Pepe, lead author of the study. Despite its growth in popularity, it was not fully known whether the NRI truly was a robust method for assigning importance to biomarkers correctly.

Dr. Pepe and co-workers chose to test the NRI method by performing simulated studies on an already existing dataset. They utilized a population dataset of 10,000 individuals with an event rate of 10.2%. They chose four biomarkers with no predictive ability and repeatedly calculated the NRI on a randomly chosen subset of patients to test for a positive statistically significant result. The rate of these false-positive results would be expected to be approximately 5.0%, based on the definition of statistical significance. Upon performing 5000 simulations, the rate at which the NRI generated a false-positive result was 63%, 23% or 34% depending on the dataset chosen for calculating the NRI.

This finding was quite alarming. Dr. Pepe explained, "Our results show that positive conclusions based on the NRI statistic are quite likely to be false even in very well designed biomarker studies."

It is worrisome that a methodology of such high prevalence in the research community would be found to be so error-prone. Hopefully, the NRI is only an outlier and not representative of other common statistical metrics. Dr. Pepe believes that, "For the most part statistical methodology is

evaluated rigorously before being applied routinely to real data," and that, "Somehow this invalid methodology based on the NRI statistic slipped in between the cracks."

There is a lesson to be learned here. According to Dr. Pepe, "this occurrence reminds us to always critically examine the fundamentals and to not take for granted that some procedure works just because we expect it to. That statement of course applies to all of science, not just to statistical procedures."

The important direct conclusion to be drawn from the study is the invalidation of the NRI. "The implication of our paper is that scientists should not use the NRI statistic for evaluating biomarkers. Moreover, readers and reviewers must be skeptical about the results of biomarker studies that indicate good biomarker performance based primarily on statistical significance of the NRI statistic," said Dr. Pepe.

Pepe MS, Janes H, Li CI. 2014. Net Risk Reclassification P Values: Valid or Misleading?*J Natl Cancer Inst*. 106: dju041.

See also: Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. 2014. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology*. 25: 114-21.
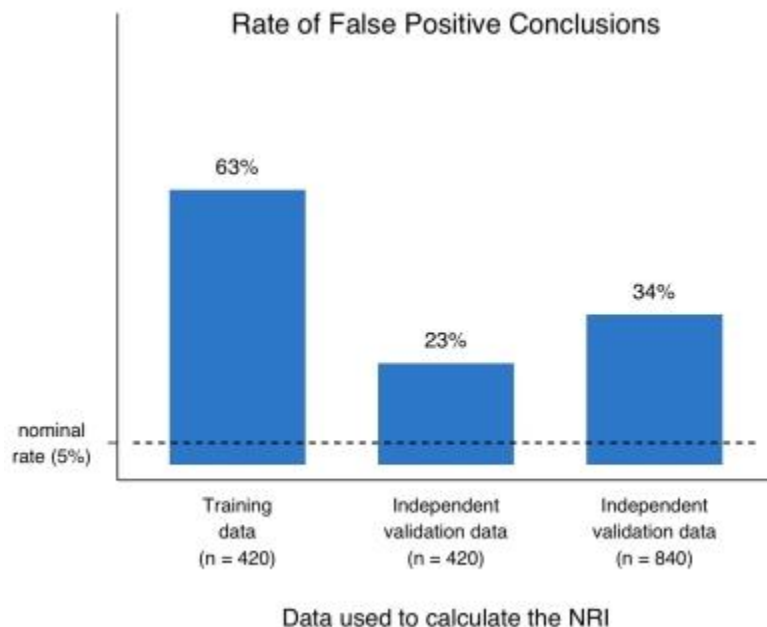

Image provided by Dr. Margaret Pepe.

Chance of a false-positive conclusion for four useless biomarkers when using the Net Reclassification Index (NRI) statistic p-value. A training set of 420 subjects of whom 10% had events were used to fit risk models with and without the biomarkers. The expected rate of false-positive conclusions due to random chance is 5%, but the actual rates are much higher.