

Quantifying Biomarker Risks and Benefits

April 21, 2014

G Brennan

The 1000 Genomes Project and other “big data” studies have identified an unprecedented number of potential cancer biomarkers. However, as the number of potential biomarkers grows, sorting through them to identify the most useful predictors for both populations and individual patients becomes a substantial problem. In a recent paper published in *BMC Medical Informatics and Decision Making*, Drs. Xiaohong Li and Thomas L. Vaughan (Public Health and Human Biology Divisions), in collaboration with researchers at the University of Washington, developed a mathematical index called the naïve **R**atio of **P**opulation **B**enefit (RPB), which takes into account both the benefits of biomarkers as screening tools as well as the potential adverse effects from using a biomarker inappropriately, as a framework to quantitatively assess the utility of biomarkers.

Biomarkers are useful in several scenarios including the prediction of individuals in a population who will develop disease (risk prediction), identifying individuals with diseases that are currently asymptomatic (early detection), identifying pathways for therapeutic intervention (targeted prevention/therapy), and identifying risk factors that increase the possibility of disease (risk prevention). Conceptually, a given biomarker is likely to be more useful for some of these categories than others, so using published biomarkers clinically without a well-established framework for evaluation can lead to either underdiagnosis, missing true cases of cancer, or overdiagnosis, improperly diagnosing disease-free patients with cancer. Therefore, integrating both the potential risks of over- or underdiagnosis with the potential benefit is a critical step to translate these biomarkers from the bench to the bedside.

To illustrate this point, Li, *et al.* demonstrate that traditional measures for biomarker efficacy, such as odds ratios, sensitivity, and specificity, may have different impacts depending on the situation in which they are applied. For example, a biomarker with low prevalence in the population will have very low sensitivity, a measure of the true positive rate of detection, even if the biomarker is very strongly associated with disease. Therefore, even combining multiple low-frequency biomarkers into an assay may have limited utility for population level analyses such as risk prediction or early detection. Because false-positives and false-negatives are unavoidable, Li, *et al.* developed the naïve ratio of population benefit (RPB), a mathematical tool that considers the adverse effects of misclassification when evaluating the overall benefit of a given biomarker for population level

reduction of disease burden. In addition, the RPB index considers disease prevalence, a variable that is not generally taken into account by current methods for biomarker evaluation.

Translating newly identified biomarkers into informative tools for epidemiological and clinical use is the next challenge for large-scale studies. Li, *et al.* propose a risk benefit approach to analyze the utility of these biomarkers for both preventive approaches and risk prediction/early intervention strategies. The current study provides an excellent framework for analyzing biomarkers at a single time point. "It is well known that disease development is a stochastic, dynamic process. The future work needs to incorporate time as a factor in this quantitative evaluation process," said Dr. Li.

[Li X, Blount PL, Reid BJ, Vaughan TL.](#) 2014. Quantification of population benefit in evaluation of biomarkers: practical implications for disease detection and prevention. *BMC Med Inform Decis Mak* (1):15.



**What do we do?
Ratio Population Benefit (RPB)**

$$RPB = \frac{wf_1 \int_c^\infty f_d(x) dx - (1-w)f_2 \int_c^\infty f_{\bar{d}}(x) dx}{wf_1 \int_c^\infty f_d(x) dx + (1-w)f_2 \int_c^\infty f_{\bar{d}}(x) dx + wf_1 \int_{-\infty}^c f_d(x) dx + (1-w)f_2 \int_{-\infty}^c f_{\bar{d}}(x) dx}$$

Image courtesy Dr. Xiaohong Li

Logogram of potential risk factor and biomarker categories (top), and the equation for the Ratio of Population Benefit (bottom). See the paper for definitions of the variables.