



Department of Epidemiology, Biostatistics & Occupational Health

Biostatistics Seminars

Fall 2013

Andrew J. Vickers, PhD

*Department of Epidemiology and Biostatistics
Memorial Sloan-Kettering Cancer Center*

How do we know whether a predictive model is of clinical value? How do we know whether a molecular marker is worth measuring? A discussion of some simple decision analytic methods

Wednesday, September 18th, 2013

3:00 pm – 4:00 pm

Education Building, 3700 rue McTavish, room 129

ALL ARE WELCOME

Abstract:

There is increasing interest in and use of multivariable prediction models to aid clinical management. In oncology, it has been shown that such models are more accurate than the use of crude risk categories, such as those based on cancer stage. Accordingly, it has been suggested that multivariable models should be used to make decisions about patient care, such as whether a patient should undergo biopsy in the light of a raised PSA level. Research on molecular markers has mirrored the growth of prediction models: currently an enormous volume of papers are published examining whether a tissue or blood marker can predict the occurrence or course of disease.

Markers and models are currently evaluated in terms of accuracy using metrics such as the area-under-the-curve (AUC), sensitivity and specificity or the concordance index. A model is thought to be a good one if it is accurate; a marker is claimed to be of value if it increases the accuracy of a model. But how accurate is accurate enough? For instance, should we use a model with an AUC of 0.65, or only those with AUC's above 0.75? Similarly, if a marker improves AUC from, say, 0.65 to 0.68, is it worth using in the clinic? Or even taken the simple case of two binary diagnostic tests, with sensitivities of 91% and 51% and specificities of 40% and 78%: which is better?

Markers and models can also be evaluated in terms of calibration. But how much miscalibration would be "too much" to prevent clinical use of a model? What about the case where one model has better calibration and the other better discrimination, which model should be used?

The answers depends, of course, on what the model, test or marker will be used for. Evaluating models and markers in terms of clinical consequences is the remit of a field known as "decision analysis". The problem with traditional decision analysis is that it requires additional information, for example, on the benefits, harms and costs of treatment, or on patient preferences for different health states. Perhaps as a result, the number of papers in the literature using decision analytic methods is dwarfed by those that report accuracy.

In this presentation, I will describe some simple decision analytic methods that can be directly applied to the data set of a model or marker, without the need for external information. These methods can therefore be used to tell us whether or not to use a model in the clinic, or whether a marker is a good one. I will illustrate the use of the methods with some straightforward real-life examples.