Development, Validation, and Application of Risk Prediction Models - Course Syllabus 2011-2012

Follow this and additional works at: http://digitalcommons.wustl.edu/mphs_materials
Part of the Medicine and Health Sciences Commons

Recommended Citation
http://digitalcommons.wustl.edu/mphs_materials/6

This Syllabus is brought to you for free and open access by the Master of Population Health Sciences at Digital Commons@Becker. It has been accepted for inclusion in Teaching Materials by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
This course will present detailed discussion of the methods of predictive modeling, with applications to clinical and population health settings. When does adding another variable to prediction improve classification? At what trade-off to implementation in real world settings? The philosophy of the course is that learning should move away from the dictionary definition, “to receive instruction, to be informed, to commit to memory” to one that works for how computers learn, where committing to memory is a trivial task: things (students in a course like this?) learn when they change their behavior in a way that makes them perform better in the future. A performance based definition. We explore how prediction models fit into this concept of learning focusing on their development and implementation to improving health outcomes.

Building from traditional risk factor identification to model refinement and validation of prediction, a number of statistical approaches will be reviewed. Each method is motivated by clinical examples. Basic concepts and philosophy of supervised and unsupervised data mining as well as appropriate applications will be discussed. Topics covered will include model development and validation; regression approaches, model selection, inference, averaging; classification and regression trees (CART), multivariate adaptive splines (MARS), neural networks, random forests, and bagging and boosting. Approaches to validation will be discussed and strategies for estimation of added value with expanded variable lists will be a key focus of this applied quantitative methods course.

**Course note:** Biostatistics I and II (M21-560 and M21-570) are required prerequisites. Meets Mondays 2 p.m. to 4 p.m.

**Text:** (available on reserve at the library)

**Data mining Practical Machine Learning Tools and Techniques.** Witten IH, Frank E, Hall MA. MorganKaufman, Burlington Mass, 2011. (from Waikato University, New Zealand; has Weka web based software).

A more statistical theory based text for reference is


**Course Activities:**

Lectures, discuss assigned papers, critique manuscripts, and computer labs.
Evaluation:

Students will outline a formal strategy to develop, validate and evaluate performance of a clinically relevant model, including strategies to evaluate the benefit of adding new markers to a model and the clinical and public health utility of an extended model for risk prediction and classification.

Competencies:

Develop the knowledge and skills to design, implement, and evaluate epidemiology-related, health services or clinical research projects of clinical or public health significance including:

- Develop the knowledge and skills with biostatistical methods and computer software for performing appropriate analyses of public health services or clinical outcomes data.

- Develop the knowledge and skills with the definitions in basic issues involved in the clinical prediction rules including, design, development, validation, implementation, and interpretation of results for their application in clinical or public health settings.

- Apply the principles of dissemination and implementation science to the evaluation of evidence for use of risk prediction models in clinical and public health programs. Design and implement strategies with appropriate integration of evaluation to inform the refinement of clinical and public health programs that will lead to improved health and wellness of the population.

- To achieve this competency, students will: Understand the development, implementation, evaluation and refinement of guidelines as they relate to risk prediction models.

- Apply principles of study design and evaluation to T2 research and implementation projects.

Topics from list above

Clinical risk prediction: 1. Overview of clinical and public health risk prediction

Wald N. When can a risk factor be used as a worthwhile screening test? BMJ 1999; 319:1562-5


Pencina, D’Agostino et al...Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in Medicine 2008; 30;27(2):157-72 and discussion 207-12


2. Stats approaches and issues... Logistic regression


Huang, ... Olsen et al. Use of Medicare claims to rank hospitals by surgical site infection risk following coronary artery bypass graft surgery. Infect control Hosp Epidemiol 2011:32:775-83

3. Stats approaches and issues... ROC analysis

OBGYN example Esplin and Macones commentary


4. Stats approaches and issues... Sample size

5. *Stats approaches and issues*…

Data mining techniques in propensity score estimation (Schneeweiss examples)


Schneeweiss et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiol 2009;20:512-22

6. *Stats approaches and issues*… *Instrumental variables*


Rassen et al. Instrumental Variables II: Instrumental variable application in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance J Clin Epidemiol 2009;62:1233-41

7. *Applications Development and validation in urology practice*

Kattan M et al – prostate develop and validate…

Nam et al Prospective multi-institutional study evaluating the performance of prostate cancer risk calculators J Clin Oncol 2011; 29:2959-64

8. *Applications CVD*

Nancy Cook references


Cardiovascular risk prediction in diabetic men and women using HbA1c vs diabetes as a high-risk equivalent. Arch intern med 2011:

9. Applications Pharmaco-epidemiology

Schneeweiss references

10. Applications breast cancer models and comparisons


Colditz GA, Rosner BA. Cumulative risk of breast cancer to age 70 years according to risk factor status: Data from the Nurses' Health Study. Am J Epidemiol 2000;152(10):950-64.


11. Applications Stepwise regression:

12. Applications Multivariate pattern analysis (MVPS) using machine learning algorithms (Schlaggar lab)

Doesenbach et al Prediction of individual brain maturity using fMRI. Science 2010; 329:1358-61

13. Applications .. RPART


14. Applications with other methods

From development to implementation and improved outcomes

Dubberke E, et al. Development and validation of a Clostridium difficile infection risk prediction model. Infection control hospital epidemiol 2011; 32: 360-6

Additional examples ...

Examples drawn from, Random forests, MCMC modeling, Bayesian approaches, MARS, Bagging and boosting