Evaluation of Prediction Models in Medicine

Ameen Abu-Hanna\textsuperscript{a} and Niels Peek\textsuperscript{b}

\textsuperscript{a}Department of Medical Informatics, University of Amsterdam, The Netherlands
\textsuperscript{b}Health eResearch Centre, University of Manchester, UK

Abstract

The reliable prediction of outcomes from disease and treatment is becoming increasingly important in the delivery and organisation of health care. Advances in personalised medicine are catalysing the search for new prognostic markers. The uses of outcome predictions range from the level of individual patients, where they help doctors and patients to make treatment choices, to patient populations, where they support health-care managers in planning and allocating resources.

This tutorial focuses on methodologies for quantitative assessment of the performance of prediction models. The key to quantitative evaluation is the use of reliable methods for obtaining valid performance measures of unseen data with well-defined characteristics. In the medical scientific literature, some methods (e.g. ROC analysis) are frequently applied, whereas other methods (e.g. information criteria, and decision analysis) are largely neglected.

The tutorial will clarify the relevant methods and the relationship between them using conceptual and mathematical frameworks. It is explained under which circumstances specific methods are applicable and when they are not. In addition, attention will be paid to the various choices in the design of model evaluation procedures, and the relationship between model evaluation and the purposes for which a model has been built. All methods are illustrated with real-world examples from the domains such as cardiac surgery and intensive care medicine.

We will also address a framework for developing, evaluating, and reporting on prediction models. The participants should be acquainted with the concept of a prediction model; with the concept of probability; and with the notion that prediction models are fitted on a training dataset and tested on a test set. Upon following the tutorial the participants should be able to assess the performance of prediction models; know how to report on them; and be able to critique performance assessment reported in the literature.

Tutorial Speakers

Ameen Abu-Hanna, PhD, is Professor and Chair of the Department of Medical Informatics at the Academic Medical Center at the University of Amsterdam. He is Principal Investigator in the research area Methodology in Medical Informatics with interest in artificial intelligence, machine learning and decision support systems. He is associate editor of Journal of Biomedical Informatics and a former president of the European Society of AI in Medicine.

Niels Peek, PhD, is Reader in Health Informatics at the University of Manchester in the UK. He has a background in Computer Science and Artificial Intelligence. His research focuses on data-driven informatics methods for healthcare quality improvement, clinical decision support, and public health. He co-chaired workshops on Intelligent Data Analysis in bioMedicine and Pharmacology (IDAMAP) in 2005 (Aberdeen), 2006 (Verona) and 2011 (Bled), and chaired the 14th Conference on Artificial Intelligence in Medicine (AIME 2013). He chairs the IMIA working group on Data Mining and Big Data Analytics and is the current president of the European Society of AI in Medicine.

Both presenters published widely on prognostic models and have recently published a virtual issue on clinical prediction models in the Journal of Biomedical Informatics. They jointly presented similar tutorials at MEDINFO 2004 in San Francisco, USA; at MEDINFO 2010 in Cape Town, South Africa; at AIME 2009 in Verona during the Doctoral Consortium; and at AIME 2013 in Murcia, Spain.

General Organization of the Tutorial Proposal

The tutorial is provided as a 180-minute session and consists of five parts having the following structure:

Part One:

1. Introduction
1.1 Prediction models
1.2 Reasons to evaluate
1.3 Model building and evaluation

Part Two:
2. Performance measures
2.1 Accuracy Measures
   ROC analysis
   Generalized measures
2.2 Precision
   Calibration
   Indirect measures of precision
2.3 Decision Analytic approaches

Part Three
3. Model selection and validation
3.1 Model selection
   The bias-variance trade-off
   Cross-validation, Information Criteria
   Bootstrapping
3.2 Internal, temporal and external validation
3.3. Comparing different models

Part Four
4. Model validation with Big Data
4.1 Cohort selection / phenotyping
4.2 Dealing with missing data
   Missing predictors
   Missing outcomes

Part Five
5. A Framework for developing, evaluating and reporting on prediction models.

Handouts and interaction

In essence the tutorial will have the form of a presentation that is supported by projected slides. The presenters will alternate twice. Participants will be provided with handouts of the slides used by the tutorial presenters. Although such tutorials do not provide much opportunity for interaction, the participants will be challenged by the presenters during the tutorial in two ways: 1) by plenary raising questions about the covered material and 2) by allowing the participants to think about how the material applies to prediction models that the participants are acquainted with.