UMC Utrecht
Julius Center

How to appraise Individual Participant Data meta-analysis in diagnostic and prognostic prediction research

Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM
on behalf of the IPD-MA methods group (Cochrane)

Risk prediction models

- Predict absolute probability of
  - presence certain outcome (diagnosis)
  - future occurrence certain outcome (prognosis)
- Often developed using multivariable regression analysis
  - Logistic regression, Cox regression, …
  - Examples: Framingham, Gail, APACHE, …
- Actual performance often disappointing
  - Development dataset too small or local
  - Varying baseline risk across study populations
  - Varying predictor effects across study populations
- Limited **transportability** or **generalizability**
  - External validation strongly recommended
  - Individual participant data meta-analysis

Potential aims of an IPD meta-analysis

1. Develop and directly validate a new prediction model
2. Evaluate performance of an existing prediction model
3. Compare the performance of competing models
4. Adjust and combine the most promising competing prediction model(s)
5. Examine the added value of a specific predictor

Important steps of an IPD meta-analysis

1. Pre-specifying a study protocol
2. Identifying relevant studies
3. Assessing the risk of bias
4. **Statistical analyses**
5. Reporting

Focus on statistical analyses in this presentation. Details on all steps will be submitted around the end of 2014.

Statistical methods: handling of missing data

- Distinguish between MCAR, MAR, MNAR
- Account for between-study heterogeneity
- Systematically missing predictors

Recommended approaches

- Multiple imputation models stratified per study
- Hierarchical multiple imputation models
  - Resche-Rigon *et al*, Stat Med 2013
  - Jolani, Debray, *et al*, submitted

Statistical methods: developing and directly validating a new prediction model

- Investigate **heterogeneity** in baseline risk (or hazard) and predictor effects
- Facilitate implementation in new study populations
  - Estimate stratified intercept term (or baseline hazard)
  - Avoid heterogeneity in predictor effects
  - Adopt parametric survival models
- Apply **internal-external cross-validation**
  - Iteratively discard one study for external validation and use the remaining studies for model development (Debray *et al*, Stat Med 2013)
  - In case of few studies: bootstrapping techniques (Cai *et al*, Biometrics 2011)

Statistical methods: evaluating the performance of one or more existing prediction models

- Summarize model performance across various study populations
  - Pooled performance & prediction intervals
- Identify modifiers of model performance (similar to subgroup analysis in intervention research)
- Investigate degree of relatedness between development and/or validation samples (Debray *et al*, JCE 2014)
  - Interpret achieved model performance in terms of case mix differences
  - Distinguish between model **reproducibility** and model **transportability**

Statistical methods: adjusting and combining the most promising models

- Combine literature models into a meta-model (Debray *et al*, Stat Med 2012 & Stat Med 2014)
- Facilitate implementation in new study populations
  - Stratified intercept term (or baseline hazard)
  - Avoid heterogeneous predictors
  - Adopt parametric (survival) models
- Further research needed

Statistical methods: examining the added value of a specific predictor

- Compare the performance of statistical models with and without the predictor of interest
  - Discrimination
  - Calibration
  - Re-classification
- Eventual models not required to yield absolute outcome probabilities in new participants
  - Random effects distributions can be used to account for between-study heterogeneity
  - No need to adopt parametric survival models
- Investigate between-study heterogeneity in the evaluated predictor effect

Reporting: important issues

PRISMA-IPD underway!!

- Choice of statistical methods
  - Missing data
  - Between-study heterogeneity
  - Predictor selection
  - …
- Estimates of baseline risk/hazard and predictor effects
  - Allow calculation of absolute risk in new participants
- Estimates of model performance
  - In overall & in individual studies
  - Prediction intervals