

# Validation of predictive models

- What to validate?
  - selection of covariables: *model specification*
  - estimation of coefficients: *model quantification*
  - predictive ability: *model performance*
- Types of validity
  - apparent (own sample)
  - internal (own population)
  - external (other population)

# Internal validity: techniques

- Split-sample
  - e.g. fixed 2/3 development, 1/3 validation
- Cross-validation
  - e.g. alternating  $\frac{1}{2}$  development,  $\frac{1}{2}$  validation (split-half)
  - 3/4 development, 1/4 validation (split-quarter)
  - $n-1$  development, 1 pt validation (jack-knife)
- Bootstrap
  - $n$  in bootstrap sample for development (drawn with replacement),  $n$  for validation

# Internal validity: characteristics

- Random splitting
- Efficiency
  - bootstrap  $n$ , jack-knife  $n-1$ , cross-validation part of  $n$
  - high stability for bootstrap; lowest for 1 analysis of an independent sample (split-sample)
- Feasibility
  - bootstrap can incorporate only automated procedures, independent sample full process
- Face validity
  - independent sample high
    - but cheating easy (repeat for better result, if disappointing validity)
  - bootstrapping low, “statistical trick”

# External validity (generalizability)

- Temporal validation
  - e.g. same investigators, cross-validate in recent years
- Spatial validation (other place)
  - e.g. same investigators, cross-validate in centers
- Fully external
  - e.g. other investigators, other center
- Characteristic: **non**-random splitting
- See Justice Ann Int Med 1999

# Statistical vs clinical validity

- Statistical validity
  - ‘goodness of fit’: linearity, additivity
  - reliable and discriminating predictions
- Clinical validity
  - ‘satisfactory performance’
  - related to intrinsic prognostic information
  
  - See Altman & Royston, Stat Med 2000