Summary of main challenges and future directions

Martin Schumacher

Institute of Medical Biometry and Medical Informatics, University Medical Center, Freiburg
Outline

• Some historical milestones

• The "optimal" cutpoint problem revisited

• Methodological achievements

• Predictive performance

• High-dimensional genomic information

• Current limitations of systematic reviews

• Summing up
Some historical milestones


1984: Buyse, Staquet & Silvester (eds.): Cancer Clinical Trials: Methods and Practice.

Requirements for confirmatory prognostic factor studies according to Simon & Altman

1. Documentation of intra- and inter-laboratory reproducibility of assays
2. Blinded conduct of laboratory assays
3. Definition and description of a clear inception cohort
4. Standardization or randomization of treatment
5. Detailed statement of hypotheses (in advance)
6. Justification of sample size based on power calculations
7. Analysis of additional prognostic value beyond standard prognostic factors
8. Adjustment of analyses for multiple testing
9. Avoidance of outcome-orientated cut-off values
10. Reporting of confidence intervals for effect estimates
11. Demonstration of subset-specific treatment effects by an appropriate statistical test
Some historical milestones


1984: Buyse, Staquet & Silvester (eds.): Cancer Clinical Trials: Methods and Practice.


The “optimal” cutpoint problem revisited

• Popular data-driven approach for evaluating the effect of a continuous marker

• Gives best separation into two groups (minimal P-value)
P-values of the logrank test as a function of all possible cutpoints for S-phase fraction (A) and Kaplan-Meier estimates of event-free survival probabilities by S-phase fraction using “optimal” cutpoint (B) in the Freiburg DNA study.
The “optimal” cutpoint problem revisited

- Popular data-driven approach for evaluating the effect of a continuous marker

- Gives best separation into two groups (minimal P-value)

- Leads to inflation of type I-error (~50% instead of 5%)

- Corrected P-values can be calculated (Miller & Siegmund, 1982; Lausen & Schumacher, 1992; and subsequent publications)

- Leads to overestimation of effects (shrinkage factors and correct confidence intervals can be obtained by bootstrap resampling)
Estimates of cutpoints and log-hazard ratio in 100 repetitions of randomly allocated observed SPF values to event-free survival times in the Freiburg DNA study before (a) and after (b) correction.

(a)

(b)
Some historical milestones


1984: Buyse, Staquet & Silvester (eds.): Cancer Clinical Trials: Methods and Practice.


2001: Crowley (ed.): Handbook of Statistics in Clinical Oncology.

Methodological achievements

• Design of prognostic studies; sample size considerations
• Regression modeling (time-to-event data) and related aspects
• Tree-based and machine learning approaches
• Development and validation of prognostic models
• Interactions between markers and treatment
• Performance measures
• Reporting guidelines
Nevertheless, …

• … the situation is still unsatisfactory

• … the medical literature is full of conflicting results,

• … many studies are of poor quality and/or poorly reported,

• … meta-analysis is severely limited by publication bias and selective reporting,

• … only few markers have proven to be clinically useful
Nevertheless, …

• … the situation is still unsatisfactory

• … the medical literature is full of conflicting results,

• … many studies are of poor quality and/or poorly reported,

• … meta-analysis is severely limited by publication bias and selective reporting,

• … only few markers have proven to be clinically useful

• In addition, predictive performance is at most moderate and predictions for individual patients are mostly poor.
Example: Rosenwald DLBCL study

• Prognostic study in patients with diffuse large-B-cell lymphoma (DLBCL) (n=240 patients, median follow-up 2.8 years, 57% deaths)

• International Prognostic Index (IPI): established prognostic model summarizing five clinical covariates (available in 222 patients)

• Prognostic performance (prediction error) is measured by the Brier score comparing actual survival status and predicted survival probability at time $t$. (0.632+ estimate based on bootstrap resampling)
Prediction error (.632+ estimates) in the Rosenwald DLBCL study
Example: Rosenwald DLBCL study

- Prognostic study in patients with diffuse large-B-cell lymphoma (DLBCL) (n=240 patients, median follow-up 2.8 years, 57% deaths)

- International Prognostic Index (IPI): established prognostic model summarizing five clinical covariates (available in 222 patients)

- \( p = 7399 \) microarray features.

- Prognostic performance (prediction error) is measured by the Brier score comparing actual survival status and predicted survival probability at time \( t \) (.632+ estimate based on bootstrap resampling)
High-dimensional genomic information (1)

- Old paradigm: 
  \[ p \text{ (number of predictors)} \ll n \text{ (number of patients / events)} \]

- Current situation: 
  \[ p \gg n \]
High-dimensional genomic information (1)

• Old paradigm:
  \[ p \text{ (number of predictors)} \ll n \text{ (number of patients / events)} \]

• Current situation: \[ p \gg n \]

requires

• stringent criteria for assessment and validation but splitting in training and test set may be very inefficient

• algorithms for fitting sparse models that simultaneously consider high-dimensional genomic and "traditional" clinical information

• further efforts towards reproducibility of research results
High-dimensional genomic information (2)

Approaches for developing prognostic models:

1) univariate analyses

2) dimension reduction

3) regularization (penalization)

4) boosting / ensemble methods
Prediction error (.632+ estimates) in the Rosenwald DLBCL study
Prediction error (.632+ estimates) in the Rosenwald DLBCL study

![Graph showing prediction error over time for different methods: Kaplan-Meier, IPI, microarray (CoxBoost), combined (CoxBoost).]
Prediction error (.632+ estimates) in the Rosenwald DLBCL study
Current limitations of systematic reviews

• … Current systematic reviews … only draw attention to the paucity of good-quality evidence …

• … systematic reviews of prognosis often highlight inconsistent and sometimes contradicting conclusions …

• … all reviews reported associations based on statistical significance only …

• … all these biases indicate that the literature may be largely unreliable and that under these circumstances, statistical significance loses its discriminating ability for the importance of prognostic markers …
So what to do?

Give up!

or

Join the Cochrane Prognosis Systematic Review Methods Group!
Questions

• Are systematic reviews of prognosis studies, done in the "traditional" way, worth the efforts?

• Is it really necessary to know all these poorly designed (conducted, analyzed, reported) studies?

• Shouldn’t we concentrate on high-quality studies where individual patient data (IPD) are available?
Back to the roots!

• Improvement of primary prognostic studies
  – Large protocol-driven prospective studies
  – Mandatory registration
  – Improved and reproducible statistical analyses
  – Improved reporting standards
  – More emphasis on pre-publication validation
  – General availability of individual patient data

• Prospectively planned pooled analyses of high-quality studies based on individual patient data
  – Development of a suitable framework
  – Further methodological challenges (!)
Summing up

• So, all in all, some cautious optimism seems indicated if the "prognostic community" is prepared for active and large-scale cooperation

• But, there is probably still a long way towards an evidence-based assessment and application of prognostic factors