



## **Guidelines for Clinical Proteomics – The Statistician’s Point of View**

Rolf Holle

Helmholtz Zentrum München  
Institute of Health Economics and Health Care Management

Kopenhagen, 24/02/08



## **Types of biomarkers according to clinical use**

### **Diagnostic marker**

- detect (specific) disease precisely, timely, efficiently, non-invasively
- diagnostic gold standard needed, ROC analysis (sensitivity and specificity)

### **Monitoring marker**

- early detect remission or deterioration (surrogate endpoint)
- repeated measurements during follow-up, gold standard for clinical event

### **Prognostic marker**

- estimate course of disease, identify high/low risk group
- relevant follow-up and outcome measure needed, „survival“ analysis

### **Predictive marker**

- predict response to specific treatment (→ personalized therapy)
- evaluated best within RCT, interaction effect treatment\*biomarker

# Phase I - V model of diagnostic evaluation

## Phase I:

Technical evaluation and optimization of the diagnostic procedure / marker (reproducibility, stability)

## Phase II:

Discriminative ability in selected patient groups and controls (e.g. from biobanks)

## Phase III:

Prospective evaluation in a representative clinical patient sample (diagnostic study)

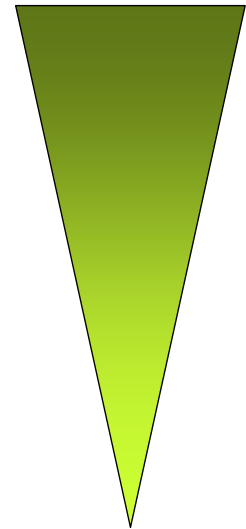
## Phase IV:

Controlled evaluation of improved health outcomes (randomized clinical trial)

## Phase V:

Economic evaluation (model based)

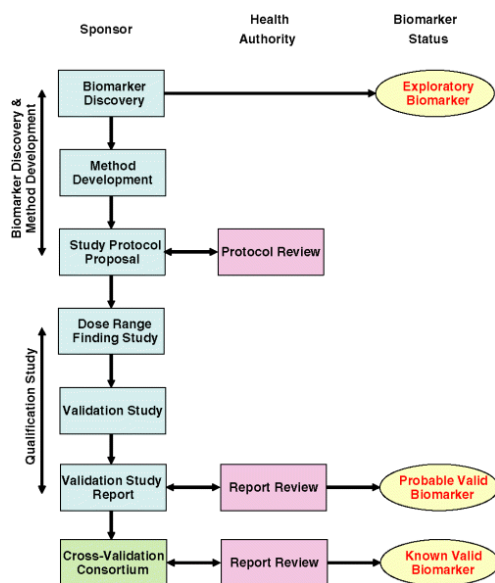
potential innovative marker



routine clinical use

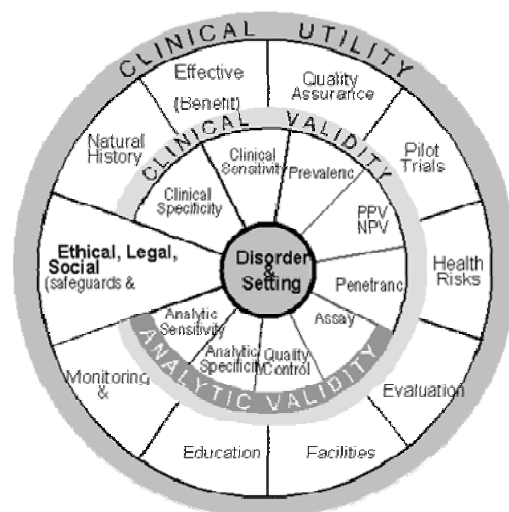
## Validation steps: other views

FDA qualification process



from: Marrer & Dieterle 2007

CDC: the ACCE wheel



<http://www.cdc.gov/genomics/gtesting/ACCE.htm>

# Phase I diagnostic studies

## Technical optimization

- minimize time and resources needed for biomarker measurement
- allow parallel analysis of multiple markers

## Quality assurance and standardization

- identify sources of measurement variation, quantify measurement error
- improve pre-analytical and analytical precision
- make procedures fit for clinical routine use

## Statistics

- reliability/reproducibility analysis: intraclass correlation, coeff. of variation

# Phase II & III diagnostic studies

## Phase II

- cases and controls from biobank/clinical database
- blinded assessment essential (Phase II)

## Phase III

- prospective clinical cohort (phase III)
- complete work-up (i.e. gold standard diagnostic)

## Analysis

- sensitivity, specificity, ROC curves
- positive and negative predictive value
- comparison to available diagnostic tests

# Multiple marker studies

## Aim

- selection of best marker
- optimal combination of markers

## Exploratory analysis of multiple markers results in overoptimism

- strict multi-step approach: searching, tuning, validating

## Best combination is impossible to identify

- „All models are wrong, but some are useful.“ (George Box)
- Simple models are generally more robust, easier to apply, and less costly

# Prognostic factor studies

## Study design

- Define timing of marker measurement
- Ensure standardized treatment (not depending on biomarker)
- Complete follow-up after relevant time period
- Assess add-on prognostic value in addition to available variables

## Clinical relevance of prognostic factors

- Improved predictability of course of disease ???
- Futility of specific treatment ?
- Necessity of intensified treatment ?
- How good should the prognostic information be to guide treatment?
- Is there an adequate treatment alternative?

# Identification of predictive markers

## Paradigm of „personalized (or individualized) medicine“

- several treatment alternatives which may be overall equivalent
- biomarker helps to identify best alternative for each patient

### Example:

- most simple case: one biomarker, two treatments
- subgroup M+ : treatment A better than treatment B
- subgroup M- : treatment B better than treatment A

### Evidence:

- best from RCT of A vs. B
- statistical test for qualitative interaction marker \* treatment  
(Gail & Simon 1985)
- even more complicated if threshold has to be defined

## Checklists study design phase III

STARD Initiative: Towards Complete and Accurate **Reporting** of Studies of **Diagnostic Accuracy** (Bossuyt et al., AnnIntMed, 2003)

- Checklist with 25 items

QUADAS: Tool for the Quality Assessment of Studies of **Diagnostic Accuracy** Included in Systematic Reviews (Whiting et al., BMC MedResMeth, 2003)

- Checklist with 14 items (yes / no /unclear)

Dupuy & Simon: Critical Review of Published **Microarray Studies** for Cancer Outcome and Guidelines on Statistical Analysis and Reporting, JNCI, 2007

- Checklist with 40 Dos and Don`ts

# Design aspects of phase III validation studies

The following questions are derived from different checklists.

## Clinical relevance

### 1. Is a clearly focussed and clinically relevant question given?

- relevant patient group
- different treatment options available
- relevant outcomes (survival, QoL, costs)
  
- decision aid based on one or more biomarkers

# Representative sample

## 2. Is the patient sample representative for a clinically relevant population?

- realistic setting
- well defined inclusion criteria
- careful choice of exclusion criteria
- inception cohort
- uniform timing of marker measurement

QUADAS Items 1, 2

# Marker definition

## 3. Is the diagnostic/prognostic marker clearly defined?

- reproducibility
- definition of cut-off values for test positivity
- handling of intermediate or missing results
- decision rule for multiple markers

QUADAS Item 13

## Reference test or clinical outcome

### 4. Is the reference test (diagnosis) or the outcome variable (prognosis) clearly defined?

- short time interval (diagnosis)
- relevant long term outcome (prognosis)
- handling of drop-outs

QUADAS Items 3, 4, 9, 14

## Independence of test and reference

### 5. Is the index test performed without knowledge of the reference test (or outcome) and vice versa?

- blinded assessment
- completeness of reference or outcome assessment
- treatment independent from marker (prognosis)

QUADAS Items 5, 6, 7, 10, 11

# Comparative study

6. Will available baseline data (i.e. known prognostic variables) be taken into account?

- add-on value compared to routine clinical data
- comparison against established markers

QUADAS Item 12

# Tuning and validation

7. Is there a distinction between exploratory and confirmatory analysis?

- choice of optimal cutpoints
- choice of optimal multivariate model
- internal validation (bootstrap, jack-knife etc.)
- (external) validation sample

Ref.: Simon R et al: Pitfalls in the use of DNA microarray data for diagnostic prognostic classification. JNCI (2003)

and

# Sample size and statistical analysis

## 8. Is the statistical approach and sample size adequate?

- based on hypothesis test or confidence interval
- based on effect estimates from literature
- different strategies for exploration and validation
- statistical hypothesis based on sensitivity and specificity (ROC curve)

## References

- Hayes DF et al: Tumor Marker Utility Grading System – a Framework to Evaluate Clinical Utility of Tumor Markers. J NCI 1996
- LaBaer J: So, You Want to Look for Biomarkers. J Proteome Research 2005
- Marrer E & Dieterle F: Promises of Biomarkers in Drug Development – A Reality Check. Chem Biol Drug Des 2007
- Schrohl AS et al: Tumor Markers – From Laboratory to Clinical Utility. Mol Cell Proteomics 2003