Biomarker trials: methods of design and evaluation through examples

Toby Prevost
Professor of Medical Statistics
June 2013
Title

Biomarker trials: methods of design and evaluation through examples

Abstract

We will begin with examples of evaluation studies of biomarkers and their combinations: can we diagnose cancer less invasively? ...can we target population screening of diabetes?

We will end with the construction of an adaptive design for early phase trials when there are multiple biomarkers that may predict patients’ treatment response in Psoriasis.

In the middle we will be looking at three key designs for biomarker trials: what are the characteristics of these randomised designs? ...and when can I use each one?
Outline:

Detection of bladder cancer less invasively  
- clinical head-to-head design

Detection of undiagnosed Type 2 diabetes  
- public health trial design

Developing a Family History Tool in primary care

Classifying biomarkers by their role

Three common phase III-IV designs – which when?

Design a multi-marker phase II study in Psoriasis  
- adaptive?
Detection of bladder cancer less invasively

The Lancet – 1999 – Stoeber et al. Immunoassay for urothelial cancers that detects DNA replication protein Mcm5 in urine

“The presence in cells of Mcm5 protein might be predictive of cancer”

Martin Bland – Think again

N=8+28

Figure 2: Relative responses obtained from cellular extracts of the two study populations

The box and whisker plot shows the range, the 25th–75th percentile, and the median for each group. A, urothelial malignancies (n=8); B, normal bladder (n=28). (Equal variance t-test p=0.0018).
Stages of research – study 1

Idea for research

Research study

Design it

Conduct it

Evaluate it

Disseminate

Problem recognised
Stages of research – study 2

Next time

- Idea for research
- Research study
  - Design it
  - Conduct it
  - Evaluate it
- Disseminate
AUC = 0.93 [95% CI: 0.89-0.97]

Boxplot
- Like separation of boxes
- Minimal overlap

Sensitivity & Specificity
- ROC curve
  - S & S at each cut-point
  - nonparametric

Meeting: Cambridge Statistical Consulting Unit

Diagnosis of ... cancer by detection of MCM5 in urine sediments
Stages of research – study 3

Idea for research → Research study

Design it → Conduct it → Evaluate it → Disseminate

Third time lucky?
Evaluation of urinary Mcm5 as a diagnostic agent in u-g tract malignancy

Two continuous markers
- with lower detection limit

Disappointingly low
- Sensitivity & specificity
- Though similar to NMP22
- Comparable AUROCs

Added value
- from combining markers
- significant contribution
- of Mcm5 added to NMP22
- MultiROC “or” method (Shulz)

N=183+1100

2 markers:
- Mcm5 AUC=0.75
- NMP22 AUC=0.72

<table>
<thead>
<tr>
<th>Features of the research programme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage</strong></td>
</tr>
<tr>
<td>Approach</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Size</td>
</tr>
<tr>
<td>Findings</td>
</tr>
</tbody>
</table>
Some desirable attributes of markers wherever used

Diagnosis

Repeateable
Valid
Reliable
Distinguishes groups
Sensitive & specific
Added value

Early detection

Prognosis

Predicts treatment response
Detection of undiagnosed Type 2 Diabetes
Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ

Diabetes risk score: towards earlier detection of Type 2 diabetes in general practice


Developed **Risk score** from:

Logistic regression model predicting 0/1 undiagnosed diabetes

- gender, age, bmi, medications

  Providing estimated coefficients = weights

  - A weighted **combination**

  - Acceptable ROC curve
Example of the Enrichment Design

The Addition Study

(Screening for diabetes)

Marker = Risk score (the combination of factors)
Marker + = the 25% with the highest risk
Marker - = the 75% with the lowest risk
Screening for T2D and population mortality over 10 years: a cluster RCT

1909 deaths

HR = 1.06 (95% CI: 0.90–1.25)

184 057 person-years of follow up

**Figure 3:** Cumulative incidence of death in the screening and no screening control groups in the ADDITION-Cambridge trial

Lancet – 2012 – Simmons, ..., Kinmonth, Wareham, Griffin. Screening for T2D and population mortality over 10 years: a cluster RCT
Developing a Family History Tool in Primary Care
Developed tool/combination from:

Logistic regression model predicting 0/1
Increased risk of >=1 of 4 conditions
Gold standard = pedigree (family tree of conditions)
Conditions = diabetes, heart disease, colorectal & breast cancer

N=618 Stage 1: Six of 12 binary items identified associated

N=529 Stage 2: Validated in a second patient sample
- ROC curve areas: Men 0.90 Women 0.89
Classifying biomarkers by their role
Biomarkers and Surrogate Markers: An FDA Perspective

Typically, “biomarker” is defined as a laboratory measurement that reflects the activity of a disease process.

There are many such markers identified for many diseases of the nervous system, for example, various MRI measures in multiple sclerosis and Alzheimer’s disease treatments, PET scanning of dopamine transporters in Parkinson’s disease, etc. In essentially all cases, these markers quantitatively correlate ... with disease progression.
Biomarkers – classifications and uses

**Biomarker**: often: present/absent +/- M+/M-

**Prognostic**:
- associated with **disease outcome**
  (not specific to a particular treatment)
- use biomarker to risk assess (+,-) to stratify for any treatment

**Predictive**:
- associated with **treatment response**
- M+ benefit from experimental tmt (only)
- can individualise therapy
- this is personalised medicine

Biomarkers – evaluated with 3 common Phase III designs
Three common phase III-IV designs – which when?
References – for biomarker trial design

Clinical Trial Designs for Predictive Biomarker Validation: One size does not fit all.

Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges.

Randomized Clinical Trials With Biomarkers: Design Issues.

Towards validation of statistically reliable biomarkers.

Bayesian adaptive design for targeted therapy development in lung cancer— a step toward personalized medicine.
**Biomarker-Stratified Design** (Full specification)
Recommended when *preliminary evidence of effect* is less robust

**Biomarker-Strategy Design** ("Use" vs "Ignore" biomarker)
Less feasible with low **M+ prevalence**

---

**Clinical Trials: Subgroup analyses**

**Pragmatic RCT Implement policies**

**Compare pooled strategies**
Enrichment Design (targeted/selected)
Requires evidence of **lack of benefit** of experimental treatment in M-

**Choice of design** depends on... evidence for a biomarker role...
- **quality** (reproducibility, validation – relevant, robust, accurate)
- **effect size** of marker-treatment relationship
- **lack of benefit** in M-
- **prevalence** of M+
- **finding** those effective ones from multiple biomarkers
- **practical limits** of sample size, cost, turnaround

‘Combination of **scientific, clinical, statistical and ethical** considerations’
Requires phase II studies to **fill gaps** and **increase potential**
Design a multi-marker phase II study in Psoriasis - adaptive?
The “client” and the Psoriasis example

Aim
Identify biomarkers specific to Psoriasis
- that predict response from treatment singly
- and in combination, sufficiently well
to inform a larger scale trial and given limited resources

Basic design (Prospective non-experimental) - Psoriasis pts on treatment
Evaluate biomarkers and treatment response: singly & together

2 Previous studies

Rheumatoid Arthritis: Davis JM et al Jnl of Immunology 2010;184:7297-304
Early RA group (n=25) / controls (n=15)
- develop immune response score from 17 “cytokine profile”
But many variables / over-fitted model / abandoned methods
Need to improve reproducibility of score \(\rightarrow\) with increased sample size

n=5 patients treated with *infliximab* (treatment)
- decline in mean severity score (response)
- decreases in Th17 / Th1 cells (marker)
Assess \(\rightarrow\) patient-level marker-response in \(\rightarrow\) larger sample
Consider \(\rightarrow\) control treatment to establish marker specific to infliximab
What effect size should be ‘detectable’?
(how much ‘variation in treatment response’ is explained by biomarker)

R-squared 10%

R-squared 20%

R-squared 40%

Markers detected with n=49
Alternatively: a 2-stage adaptive interim design

- Effective $(R^2 = 20\%)$
  - 90% retained
  - n1 = 24
  - n2 = 72
- Moderate $(R^2 = 15\%)$
  - 82% retained
- Modest $(R^2 \approx 5-10\%)$
  - 53-70% retained
  - Patients: 90% retained
- Ineffective $(R^2 = 0\%)$
  - 70% dropped after interim
  - Patients: 82% retained

Significance Fisher’s test

- Effective: 90%
- Moderate: 80%
- Modest: 36-64%
- Ineffective: 4-5%

Power

- Effective: larger sample (n2=72) with focused biomarkers to develop combination

Bauer P, Kohne K Evaluation of experiments with adaptive interim analyses
Biometrics 1994:50:1029-41
Concluding points

Biomarker trials

There are several designs for phase II-III

“One size does not fit all”

Choice of design informed by role and known characteristics of biomarker

Early phase studies allow us to learn:

– which biomarker(s) / which combination / marker prevalence
  specificity of biomarker to treatment / effect size / which later phase design

Consider an adaptive element

– cost-saving on markers (as opposed to subjects) / longer
  – larger n & focus 2^{nd} stage efforts on promising biomarkers

Plan design and analysis together

– analysis approach tested & tailored to objectives
  – based on realistic detectable effect size
  – giving appropriate sample size and adequate power
  – towards markers valid / reproducible / applicable for purpose
What else can be adapted in an adaptive design?

**MAMS** (Multi-arm Multi-stage) **design** (Matt Sydes)
Drop arms
Example: Stampede Trial
Prostate Cancer – 5 active arms
Allocation: 2 (control) : 1 : 1 : 1 : 1 : 1
http://www.trialsjournal.com/content/10/1/39

**Adaptive randomisation** (Don Berry)
http://www.methodologyhubs.mrc.ac.uk/PDF/Design%20and%20analysis%20of%20trials%20involving%20biomarkers-%20Donald%20Berry.pdf

**Seamless Phase II/III design** (Nigel Stallard)
http://www.trialsjournal.com/content/12/S1/A2
Aims to overcome delays when outcomes have long follow-up