Methodological challenges

Jan Bogaerts, PhD
Methodology Vice Director
EORTC
What is the question?

- The data
- The purpose
  - Regulatory approval
  - How to treat this particular patient?
  - What is the status of clinical understanding: how can we move ahead in this disease?
A clear-enough future (we can still be wrong but have a fairly clear picture what the future will look like)

- Treatment with dramatic side effects ("penicillin effect"). Non-RCTs appear to be optimal design to address this level of uncertainty.

Alternate futures (a few discrete alternatives whose outcomes cannot be reliably predicted)

- Equipoise exists; an RCT the best method to resolve this level of uncertainty.

A range of futures (a range of potential futures can be identified but no natural discrete scenario has emerged)

- Many new drugs. Few data on safety and efficacy. Phase II trials to address this level of uncertainty.

True ambiguity (complete ignorance)

- A new chemical moiety. Further preclinical or phase I testing necessary to help shape our uncertainty in more solid direction.
## Levels of clinical evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level I</td>
<td>Adequately powered, high quality randomised trial, or meta-analysis of randomised trials showing statistically consistent results</td>
</tr>
<tr>
<td>Level II</td>
<td>Randomised trials inadequately powered, possibly biased, or showing statistically inconsistent results</td>
</tr>
<tr>
<td>Level III</td>
<td>Non-randomised studies with concurrent controls</td>
</tr>
<tr>
<td>Level IV</td>
<td>Non-randomised studies with historical controls (i.e. typical single arm phase II studies)</td>
</tr>
<tr>
<td>Level V</td>
<td>Expert committee review, case reports, retrospective studies</td>
</tr>
</tbody>
</table>

*I. F. Tannock, Eur. J. Cancer Supplements Vol. 1, No. 5, Sept. 03, p. 93*
### (Part of) the Oxford 2011 Levels of Evidence

<table>
<thead>
<tr>
<th>Question</th>
<th>Step 1 (Level 1*)</th>
<th>Step 2 (Level 2*)</th>
<th>Step 3 (Level 3*)</th>
<th>Step 4 (Level 4*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>How common is the problem?</strong></td>
<td>Local and current random sample</td>
<td>Systematic review of surveys that allow matching to local circumstances**</td>
<td>Local non-random sample**</td>
<td>Case-series</td>
</tr>
<tr>
<td><strong>Is this diagnostic or monitoring test accurate?</strong> (Diagnosis)</td>
<td>Systematic review of cross sectional studies with consistently applied reference standard and blinding</td>
<td>Individual cross sectional studies with consistently applied reference standard and blinding</td>
<td>Non-consecutive studies, or studies without consistently applied reference standards**</td>
<td>Case-controls poor or non-applied reference studies**</td>
</tr>
<tr>
<td><strong>What will happen if we do not add a therapy?</strong> (Prognosis)</td>
<td>Systematic review of inception cohort studies</td>
<td>Inception cohort studies</td>
<td>Cohort study or control arm of randomized trial*</td>
<td>Case-series quality prognostic study**</td>
</tr>
<tr>
<td><strong>Does this intervention help?</strong> (Treatment Benefits)</td>
<td>Systematic review of randomized trials or n-of-1 trials</td>
<td>Randomized trial or observational study with dramatic effect</td>
<td>Non-randomized controlled cohort/follow-up study**</td>
<td>Case-series, or historical controlled studies**</td>
</tr>
<tr>
<td><strong>What are the COMMON harms?</strong> (Treatment Harms)</td>
<td>Systematic review of randomized trials, systematic review of nested case-control studies, n-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect</td>
<td>Individual randomized trial or (exceptionally) observational study with dramatic effect</td>
<td>Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**</td>
<td>Case-series, or historical studies**</td>
</tr>
<tr>
<td><strong>What are the RARE harms?</strong> (Treatment Harms)</td>
<td>Systematic review of randomized trials or n-of-1 trial</td>
<td>Randomized trial or (exceptionally) observational study with dramatic effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Is this (early detection) test worthwhile?</strong> (Screening)</td>
<td>Systematic review of randomized trials</td>
<td>Randomized trial</td>
<td>Non-randomized controlled cohort/follow-up study**</td>
<td>Case-series, or historical studies**</td>
</tr>
</tbody>
</table>
Oxford levels of evidence: tentative comments

• It is conceptually interesting to consider different questions
  ▪ It does not seem to be very different from the ‘usual’ table

• Why would a systematic review of n-of-1 trials be better quality than a randomized trial?

• One could consider amending/modifying this table with some ideas for rare cancers

• BUT …
Levels of evidence for rare cancers

- There is no logical rationale to say that the levels of evidence would function differently “because” it is harder to get data ...
Considerations for design

- RCT remains the gold standard
- n-of-1 design: this is a sequence of different treatments in one and the same patient.
  - Has the feel of cross-over design
  - Question: how does that work in oncology?
- Play with the type I error (or even type II error). For example:
  - One sided testing: can be acceptable
  - Higher type I error (alpha): this will never be found, because the trial will not be repeated
  - More optimistic alternative hypothesis: this has the same practical effect as increasing the type II error (beta): only a really strong improvement has good chances of being identified. Look more at the confidence interval.
Considerations for design

- Single-arm/non-comparative approaches
- The fact of having some responses is an improvement in itself
- The fact of stopping progression is an improvement in itself
- Robust historical data is available with small between trial variability (not likely, but happens)
There is no current standard ...

I would then (still) suggest a randomized approach with either a Phase II selection design, or a play-the-winner (adaptive randomization) approach

Other cases:

• Maybe it is worthwhile to incorporate in the plans a trial / decision point where disagreement is settled

• If the standard is wait-and-see, that can be randomized against
Example of evolution: how we see Phase II trials

- Trying to improve the Positive Predictive Value
- Accommodate many objectives: moving to an amalgam of approaches

Seymour et al. CCR 2010
Back-Up
Looking for new common ground

• Trials with a high level of patient startup work
  ▪ Screening many to obtain some eligible patients
  ▪ Splitting according to markers
  ▪ High workload to include patients
  ▪ Timelines to enter a patient

• Think about:
  ▪ Trials spanning several phases of development
  ▪ Trials with multiple additional analyses / endpoints
  ▪ TR analysis and planning of such analysis
  ▪ Biobanking
  ▪ Tools to perform complex logistics
Buzzword: Adaptive designs

- We are learning to plan and run these complicated trials in an acceptable way
  - Appropriate use of IDMC
  - Appropriate use of adaptive elements in the design
- Word of warning: adaptive designs are not the solution to manage the unexpected. But adaptive elements can be very interesting to manage the complicated.
- We are already using many adaptive ideas in our trials (all phases).
- Keys here are: think and discuss upfront and monitor during the trial
<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Regulatory Evidence</th>
<th>Study Design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>Clinical benefit</td>
<td>Randomized</td>
<td>Direct measure of benefit, easy, precise</td>
<td>Large studies, crossover / followup Tx affects, noncancer deaths</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Clinical benefit</td>
<td>Randomized, blinded</td>
<td>Patient perspective of direct clinical benefit</td>
<td>Blinding hard, missing data, clinically relevant effect, validated tools lacking</td>
</tr>
<tr>
<td>DFS</td>
<td>Surrogate</td>
<td>Randomized, blinded, blinded review</td>
<td>Smaller, shorter</td>
<td>Not stat. validated as surrogate for OS / not precise, open to bias / many definitions</td>
</tr>
<tr>
<td>RR</td>
<td>Surrogate</td>
<td>Blinded, blinded review</td>
<td>1-arm possible, smaller, shorter, attributable to drug</td>
<td>No direct measure of benefit / no comprehensive measure of drug activity / only subset of benefiting pats.</td>
</tr>
<tr>
<td>CRR</td>
<td>Surrogate</td>
<td>Blinded, blinded review</td>
<td>1-arm possible, smaller, shorter, durable CR = benefit</td>
<td>No direct measure of benefit / no comprehensive measure of drug activity / small subset of benefiting pats.</td>
</tr>
<tr>
<td>PFS</td>
<td>Surrogate</td>
<td>Randomized, blinded, blinded review</td>
<td>Smaller, shorter, SD included, crossover / other Tx not affecting, objective &amp; quantitative</td>
<td>Not stat. validated as surrogate for OS / not precise, open to bias / many definitions / frequent assessments / need to balance timing x arms</td>
</tr>
</tbody>
</table>
Evolution of endpoints leading to EMA oncology approvals

Per F Pignatti presentation at EORTC advanced course, September 2010
Alternative designs (cont’d)

- Bayesian design, formally incorporating historical data into the design
  - Involve prior beliefs which may not be universally accepted
  - If we conduct a small trial, the choice of the prior may carry heavy weight