Practical Statistical Reasoning in Clinical Trials

Paul Wakim, PhD
Center for the Clinical Trials Network
National Institute on Drug Abuse

7 February 2012
Disclaimer

The views expressed by the speaker during this seminar do not necessarily reflect the official views of the National Institute on Drug Abuse.
Slides are available on CTN’s Dissemination Library: 
http://ctndisseminationlibrary.org/
Search for “statistical reasoning” or “wakim”

References are listed at the end of these slides
Any questions before we start?
Trial Monitoring and Interim Analyses
Trial Monitoring and Interim Analyses

- Participants’ safety
- Regulatory
- Trial performance
- Data quality
- Sample size re-calculation
- Interim analyses for efficacy, futility, and/or harm

Interim Analyses
Why are trial monitoring and interim analyses important?

1) Participants’ safety and well-being
2) Trial integrity
3) Optimal use of resources
4) Ethical considerations
What to monitor?

1) Adverse events (AEs) and Serious Adverse Events (SAEs)
2) Regulatory compliance
3) Recruitment
4) Availability of primary outcome
5) Treatment exposure
6) Retention (follow-up visits)
7) Data quality
Interim Analyses
4 Main Points About Interim Analysis

1. It is a statistical analysis of the response variables performed while the trial is proceeding.

2. It is used to decide whether the study has come to an *early conclusion* without the need to either randomize unnecessarily additional participants, or expose them senselessly to a therapy that is proving to be inferior.

3. Because repeated examination of accumulating data increases the probability of declaring a treatment difference even if there is none, statistical adjustments have to be made.

4. None of the statistical techniques available for interim analyses should be used as the sole basis in the decision to stop or continue the trial.

Based on Proschan et al. (2006) & Friedman et al. (2010)
Possible reasons for terminating a trial earlier than scheduled

1) Serious adverse effects
2) Greater than expected beneficial effect
3) Improbable statistically significant difference by the end of the trial
4) Severe uncorrectable logistical, data quality or recruitment problems
5) Primary research question answered elsewhere or no longer sufficiently important

Friedman et al. 2010
Interim Analyses

• Sample size re-calculation (or re-estimation)

• Interim analyses for efficacy, futility and/or harm
Sample Size Re-Calculation

• Based on nuisance parameters only (no statistical penalty)

• Based on nuisance parameters and observed treatment effect (statistical penalty)

Proschan et al. 2006
Sample Size Re-Calculation
Based on Nuisance Parameters *Only*

Are the values of variances, correlations, drop-out rate, or proportions in the control group, that we assumed at the beginning of the trial consistent with what we actually see so far?

And consequently, is the sample size we calculated initially adequate based on these values?
## Sample Size Re-Calculation
Based on Nuisance Parameters *Only*

<table>
<thead>
<tr>
<th>Result</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current N is adequate</td>
<td>Keep N the same</td>
</tr>
<tr>
<td>N should be higher</td>
<td>Increase N</td>
</tr>
<tr>
<td>Lower N is adequate</td>
<td>Keep N the same or decrease N?</td>
</tr>
</tbody>
</table>
Sample Size Re-Calculation Based on Nuisance Parameters *Only*

<table>
<thead>
<tr>
<th>Result</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower N is adequate</td>
<td>Keep N the same</td>
</tr>
</tbody>
</table>

**Pros:**
- Insure adequate power for primary analysis (just in case)
- Help in interaction and safety analyses
- Help in secondary and sub-group analyses

**Cons:**
- May unnecessarily subject participants to risk
- May waste resources that could be spent on other research
- May unnecessarily delay publishing important results
Sample Size Re-Calculation
Based on Nuisance Parameters Only

<table>
<thead>
<tr>
<th>Result</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower N is adequate</td>
<td>Decrease N</td>
</tr>
</tbody>
</table>

**Pros:**
- End the trial sooner and publish results
- Save resources

**Cons:**
- Not enough power for primary analysis (just in case)
- Less data for interaction and safety analyses
- Less data for secondary and sub-group analyses
Sample Size Re-Calculation Based on Nuisance Parameters and Observed Treatment Effect

Should the sample size be changed based on the values of the nuisance parameters and the treatment effect observed so far?

This is controversial. Criticism has been about potential bias, loss of efficiency, and the possibility of increasing the sample size to detect clinically meaningless differences.

What’s the general question?

Based on the data observed so far, is the experimental treatment:
• clearly beneficial (better than control); or
• clearly futile with no hope of efficacy; or
• clearly inferior (worse than control)?

If so, may stop the trial for ethical reasons and to save resources.
Interim Analyses for Efficacy, Futility and/or Harm (statistical penalty)

Sequential designs (aka group sequential tests or repeated significance tests):
• Group sequential methods
• Flexible group sequential (alpha-spending) methods

Stochastic curtailment tests:
• Conditional power tests (frequentist)
• Predictive power tests (mixed Bayesian-frequentist)
• Predictive probability tests (fully Bayesian)

Dmitrienko et al. (2005)
Group Sequential Methods

\[G_1 \rightarrow G_1 \rightarrow G_1 \rightarrow G_1 \rightarrow G_1\]

Time

Moyé 2006
Group sequential procedures are simply processes that analyze groups of patients sequentially. ... each group’s data is added to the data that has been collected and is already available from the previous groups.

Moyé 2006

Group sequential design enables early trial stopping if there is harm, suggestion of futility, or overwhelming evidence of efficacy.

Zhu et al. 2011
Group Sequential Method
Two-Sided Stopping Boundaries

Based on Jennison & Turnbull (2000) and CTN DSC1-Duke Clinical Research Institute

- Efficacy
- Inconclusive
- Futility
- "Harm"
Flexible Group Sequential (Alpha-Spending) Methods (e.g. Lan-DeMets method)

Same as group sequential methods, but without pre-specifying the number or spacing of interim looks.

• Allow for unplanned and unequally-spaced interim looks

• Provide flexibility on how to “spend” the Type I error (or alpha) during the course of the trial

• Guarantee that at the end of the trial, the overall Type I error will be the pre-specified value of alpha

Based on Friedman et al. (2010), Dmitrienko et al. (2005) & Zhu et al. (2011)
Stochastic Curtailment Tests

Based on Lan & Wittes (1988)
Conditional Power Tests
(frequentist approach)

Conditional power (CP) is the probability that the final study result will be statistically significant, given the data observed thus far and a specific assumption about the pattern of the data to be observed in the remainder of the study, such as assuming the original design effect, or the effect estimated from the current data, or under the null hypothesis.

Lachin (2005)
Predictive Power Tests
(mixed Bayesian-frequentist approach)

They average the conditional power over the posterior distribution of the treatment effect, which is itself based on its prior distribution and the data observed so far.

Based on Dmitrienko et al. (2005)
Predictive Probability Tests
(Bayesian approach)

They are completely based on the posterior probability of a clinically important treatment effect (rather than statistical significance) given the already observed data.

Based on Dmitrienko et al. (2005)
One Cautionary Note

When performing a sample size re-calculation based on nuisance parameters only, without performing an interim analysis on futility, one may increase the sample size and extend the trial when in fact, an interim analysis would have revealed futility. In other words, spend more money testing a futile treatment.
Another Cautionary Note

Because the decision to stop the trial may arise from catching the treatment effect at a random high, truncated RCTs (tRCTs) may overestimate the true treatment effect.

Briel et al. (2009)

Truncated RCTs were associated with greater effect sizes than RCTs not stopped early.

Bassler et al. (2010)
The Importance of Timing

Logistically:
• Decision needs to be made before the end of recruitment

Statistically:
• Too early: the results may not be robust enough
• Too late: recruitment may be completed
References


Questions or Comments