The role of Statistics in Cancer Research

Daniel Sargent, PhD
ALLIANCE Group Statistician

May 2014
Objectives

After the presentation, participants will be able to:

- Restate the overall purpose of statistics in research.
- Discuss the advantages and limitations of randomization as it relates to subject enrollment, trial conduct and application of study results.
- Differentiate between various types and uses of analysis and their role in evaluating treatment efficacy and making trial conduct decisions.
- Compare the statistical advantages and disadvantages of restrictive vs broad eligibility criteria and the impact on overall study design.
The Literary Digest predicted Alf Landon would win the presidency in 1936
1936 PRESIDENTIAL ELECTION RESULTS

- Roosevelt: 523 electoral votes
- Landon: 8 electoral votes
1936 Presidential election:
Sampling

Potential Voters

Readers who sent in postcards
Readers who didn’t send in postcards
Non-readers of Literary Digest
Outline

- What is statistics?
- Need for randomization
- Types of analysis
- Subgroup analysis
- Interim analyses and DSMBs
- Surrogate Endpoints
- Eligibility Criteria
Statistics

- A foundation for scientific decision-making
- The ability to quantify errors
- The ability to generalize from those tested to a population

- These items are much more important than p-values
## Decisions in Clinical Trials

<table>
<thead>
<tr>
<th>Action: Conclude No Difference</th>
<th>Truth: No Difference Between Regimens</th>
<th>Truth: Difference Between Regimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action: Conclude a Difference</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Decisions in Clinical Trials

<table>
<thead>
<tr>
<th>Action: Conclude No Difference</th>
<th>Truth: No Difference Between Regimens</th>
<th>Truth: Difference Between Regimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Negative</td>
<td></td>
<td>True Positive</td>
</tr>
<tr>
<td>Action: Conclude No Difference</td>
<td>Truth: No Difference Between Regimens</td>
<td>Truth: Difference Between Regimens</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td></td>
<td>True Negative</td>
<td>False Negative; Type II error</td>
</tr>
<tr>
<td>Action: Conclude a Difference</td>
<td>False Positive; Type I error</td>
<td>True Positive</td>
</tr>
</tbody>
</table>
Analogy

**Jury Trial (criminal law)**
- Presume innocent
- Goal: convict the guilty
- “Beyond reasonable doubt”
- Requires: evidence (convincing testimony)
- Mistake: convict innocent person
- Mistake: acquit guilty person

**Clinical Trial**
- Assume the null hypothesis
- Goal: detect a true difference
- “Level of significance”
- Requires: evidence (adequate sample size)
- Mistake: False positive (limit to 5%)
- Mistake: False negative (limit to 10%)
Randomized Clinical Trials

The randomized clinical trial (RCT): “a last resort for the evaluation of medical interventions. It is slow, ponderous, expensive, and often stifling of scientific imagination and creative change in ongoing protocols ...; however, no other method for studying the merits of clinical treatment regimens can approach the precision of estimating effects and the strength of inference permitted by sound RCTs.”

Bailar, 1983
Randomization: Advantages

- Reduce Bias in trial enrollees
  - Patient selection by treating physician
  - Self-selection by patients

- Reduce Bias in trial conduct
  - Ineligibles
  - Refusals
  - Unknown confounding factors
Randomization: Caveats

- Randomization does not ensure the study will include a representative sample of all patients with the disease.

- Randomization does help ensure an unbiased evaluation of the relative merits of the two treatments for the types of patients entered.
Types of Analysis

• Intention to treat: All patients included in the group to which they were randomized irrespective of compliance, administrative errors, or other protocol deviations.
Types of Analysis

• Treatment received: Patients included in the group corresponding to treatment they actually received.

• Typically the intention to treat analysis answers the relevant clinical question.
Multiple Comparisons

- Beware of Field of Dreams: ‘If you test it (enough times), it will come up significant’

Example

- 20 Markers, with prevalence from 10 - 50%, measured on 100 patients
- None related whatsoever to response
- Compare response rate in those with and without marker
- Overall response rate 40%
Multiple Comparisons

Results

- Response rates ranging from 22 - 75% in the marker (+) group
- Difference in response rates between (+) and (-) ranged from 0.5 - 31%
- 2 had $p < 0.05$ comparing response rate in (+) and (-) patients
Subgroup Analyses

- Is it expected that the actual treatment effect may differ in a meaningful way between different subgroups?

- Apparent differences can result by chance alone
  - Increased risk of spurious results with greater number of subgroup analyses
Beware of Subset Analysis (1)

5-FU and levamisole as adjuvant treatment for Dukes C colon cancer

1. Mayo Clinic Trial  (Laurie et al, J Clin Oncol 1989)
   More effective for men, older patients

   More effective for women, younger patients

   No difference in by sex or age
Beware of Subset Analysis (2)

- ISIS Cardiac Trial: 17,000 patients
  - Found aspirin > placebo at preventing vascular deaths
  - Subgroups: Didn’t work in:
    - Non-diabetics
    - Systolic BP < 100 or > 175
Beware of Subset Analysis (2)

- ISIS Cardiac Trial: 17,000 pts
  - To determine ‘significance’, compared these differences to difference in astrological signs
  - No patient characteristic separated patients by more than Gemini/Libra vs other
  - Concluded no real subgroup effects
Monitoring Clinical Trials: Efficacy

- Patients enter sequentially over time, therefore information about the treatments increases as the trial progresses.
- How long is it ethical to continue enrollment in the face of mounting evidence?
Interim analyses

- Definition - An analysis conducted prior to the planned final analysis
- Possible actions:
  - Continue as planned
  - Modify the trial
  - Stop
- Controlled by an independent DSMB
Possible Reasons for Stopping Early

- One treatment convincingly superior or inferior
- Treatments convincingly not different
- Excessive toxicity
- Low accrual
- External evidence (e.g., other trials) leading to scientific irrelevance
DSMBs

- Independent panel of experts, including physician, statistician, lay people
- Have access to full data
- Report directly to study sponsor
- Strongly recommended by the FDA for all Phase III trials, becoming required by many IRBs
Potential Problems with Stopping Early

- Inflation of error rates: false positive or false negative
- Biased estimates of treatment effects
- Decreased precision from original plan
- Inability to investigate secondary objectives
- Conclusions may change with further follow-up
- Lack of credibility - failure to influence medical practice
MRC AML12 – Design

Induction chemotherapy (2 courses)

Consolidation therapy (2 courses)

Randomization

Stop
i.e. 4 courses in total

1 more course
i.e. 5 courses in total

Goal: 1000 Patients
AML12 – First DMC review (1997)

- % still alive vs Months from Randomisation
- P = 0.05

<table>
<thead>
<tr>
<th>Course</th>
<th>No. Patients</th>
<th>No. Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Courses</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>5 Courses</td>
<td>102</td>
<td>7</td>
</tr>
</tbody>
</table>
AML12 – Second DMC review (1998)

% still alive

Months from Randomisation

P = 0.003

<table>
<thead>
<tr>
<th>Course</th>
<th>No. Patients</th>
<th>No. Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Courses</td>
<td>169</td>
<td>42</td>
</tr>
<tr>
<td>5 Courses</td>
<td>171</td>
<td>23</td>
</tr>
</tbody>
</table>

64%

80%
<table>
<thead>
<tr>
<th>Timepoint</th>
<th>Deaths/Patients</th>
<th>Statistics (O–E)</th>
<th>Var.</th>
<th>HR &amp; 95% CI</th>
<th>Odds Redn. (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Five courses</td>
<td>Four courses</td>
<td>ODDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>7/102</td>
<td>15/100</td>
<td>-4.6</td>
<td>5.5</td>
<td>57% (29); p = 0.05</td>
</tr>
<tr>
<td>1998</td>
<td>23/171</td>
<td>42/169</td>
<td>-12.0</td>
<td>15.9</td>
<td>53% (18); p = 0.003</td>
</tr>
</tbody>
</table>

Five courses better | Four courses better
### AML12 – Subsequent results

<table>
<thead>
<tr>
<th>Timepoint</th>
<th>Deaths/Patients</th>
<th>HR &amp; 95% CI</th>
<th>Odds Redn.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Five courses</td>
<td>Four courses</td>
<td>Five courses : Four courses</td>
</tr>
<tr>
<td></td>
<td>(O–E) Var.</td>
<td>(SE)</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>7/102 15/100</td>
<td>–4.6 5.5</td>
<td>57% (29); p = 0.05</td>
</tr>
<tr>
<td>1998 (1)</td>
<td>23/171 42/169</td>
<td>–12.0 15.9</td>
<td>53% (18); p = 0.003</td>
</tr>
<tr>
<td>1998 (2)</td>
<td>41/240 66/240</td>
<td>–16.0 26.7</td>
<td>45% (15); p = 0.002</td>
</tr>
<tr>
<td>1999</td>
<td>51/312 69/309</td>
<td>–11.9 30.0</td>
<td>33% (15); p = 0.03</td>
</tr>
<tr>
<td>2000</td>
<td>79/349 91/345</td>
<td>–9.5 42.4</td>
<td>20% (14); p = 0.1</td>
</tr>
<tr>
<td>2001</td>
<td>106/431 113/432</td>
<td>–6.2 53.7</td>
<td>11% (13); p = 0.4</td>
</tr>
<tr>
<td>2002</td>
<td>157/537 140/541</td>
<td>6.7 74.0</td>
<td>–9% (12); p = 0.4</td>
</tr>
<tr>
<td>2003</td>
<td>220/615 215/619</td>
<td>4.3 108.6</td>
<td>–4% (10); p = 0.7</td>
</tr>
</tbody>
</table>

The table shows the comparison between five courses and four courses in terms of deaths per patients, along with the corresponding hazard ratios and 95% confidence intervals. The odds reduction is also provided for each timepoint.
AML12 - Summary

- DMC’s decision supported by subsequent results
- Main reason for not stopping was that treatment effect “too good to be true” – 50% reduction in mortality with just 25% more therapy
- Had trial been stopped, patients would now be recommended 5 courses (toxic, 3-4 weeks in hospital, ~ 3% die)
Eligibility Criteria

• Eligibility criteria determine the study population, frequently a subset of the target population.

• Trade off - Restrictive vs. Broad

Restrictive: Advantages

1. Homogeneous patients
2. Smaller sample sizes
Eligibility Criteria (continued)

Trade off - Restrictive vs. Broad

Restrictive: Disadvantages

1. Results may not be generalizable
2. Fewer patients eligible implies longer study duration
3. Patients treated off study with some treatments anyway
4. Patient access to therapy
Toward More Liberal Entry Criteria

- “Uncertainty principle”: any patient for whom the effect of treatment is uncertain should be included.
- Eligibility criteria should be loosened as trials become larger.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Sample Size</th>
<th>Eligibility Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>Small</td>
<td>Restrictive</td>
</tr>
<tr>
<td>Phase II</td>
<td>↓</td>
<td>↓ Liberal</td>
</tr>
<tr>
<td>Phase III</td>
<td>Large</td>
<td></td>
</tr>
</tbody>
</table>
Endpoints

• Must be
  a. Sensitive to the effect of treatment
  b. Clinically relevant

• Sensitive in what way?
  a. Biologic activity - short term
  b. Clinical efficacy - long term

• Question: When can a short term endpoint be substituted for a long term endpoint? That is, when can we use a surrogate endpoint?
Surrogate endpoint example: AIDS

- HIV Infection associated with progressive depletion of CD4+.
- Multiple studies show CD4+ level predicts onset of AIDS.
- Two large studies (BW 02: 281 pts, and ACTG 016: 351 pts) demonstrated large beneficial effect of ZDV on CD4+
- 1990 US: Placebo controlled trial (ZDV vs. placebo) with survival endpoint cannot be done.
Surrogate endpoint example: AIDS

- 1993 Concorde Study: 1749 patients in Europe:
  - Result 1: Immediate ZDV maintained higher CD4+ count than deferred ZDV.
  - Result 2: Immediate ZDV: 95 deaths
    Deferred ZDV: 76 deaths
- Conclusion: “Results do not encourage early use of ZDV in symptom free HIV infected adults.”
Summary

- Much of statistics, when you really think about it, is common sense
- Computers can calculate numbers, people need to make decisions
- Proper statistical thought provides the data to inform these decisions