Oncology Clinical Research & Race: Statistical Principles

Daniel Sargent, PhD
ALLIANCE Group Statistician

November 2014
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
Modern example: 2012 Presidential Election

- Overall: Obama 51%, Romney 47%
- Whites: Obama 39%, Romney 59%
- African American: Obama 93%, Romney 6%
- Hispanic ethnicity: Obama 71%, Romney 27%

http://www.ropercenter.uconn.edu/elections/how_groups_voted/voted_12.html
Moral: Who you sample matters

- The issues of who is tested when testing therapies are identical to who you ask when doing a political poll
- If you do a study in a non-representative group, you may get the wrong answer
- This is a bias no statistical test can fix – a biased sample will give a biased result
In NCI trials: pre-specify expected enrollment by gender and race/ethnicity

<table>
<thead>
<tr>
<th>Ethnic Category</th>
<th>Females</th>
<th>Males</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic or Latino</td>
<td>10</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>457</td>
<td>582</td>
<td>1039</td>
</tr>
<tr>
<td>Ethnic Category: Total of all subjects</td>
<td>467</td>
<td>593</td>
<td>1060</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial Category</th>
<th>Females</th>
<th>Males</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaskan Native</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Asian</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Black or African American</td>
<td>33</td>
<td>45</td>
<td>78</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>White</td>
<td>426</td>
<td>536</td>
<td>962</td>
</tr>
<tr>
<td>Racial Category: Total of all subjects</td>
<td>467</td>
<td>593</td>
<td>1060</td>
</tr>
</tbody>
</table>
Reality: Trials underpowered for subgroups

- Sample size based on having a high chance to detect a true effect if it is present (power)
- This is based on a single endpoint, in the entire population
- Reducing sample size by half reduces power from 80% to 50%, reducing it by 80% reduces power to 20%
- With the number of minorities presently enrolled on trials, there is no meaningful chance to detect effects in racial sub-groups
What if we just are interested in: “Does the effect differ by race’?’

- Test for ‘interaction’ – does treatment effect differ by a variable of interest (e.g. race)
- Reality: To do this, must reliably estimate treatment effect within each group, then compare
- Very challenging, in general, requires a 4x larger study, even if prevalence 50/50 (much worse for race)
What if we test anyway: 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

   More effective for men, older patients

   More effective for women, younger patients

   No difference 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
Summary of cautions

- Based on current trial design approaches
  - Cannot be confident that a differential treatment effect by race would be identified even if it exists
  - Cannot be confident that a differential treatment effect by race is real if it is identified
- In short, be very cautious
Possible strategies

- Develop sample sizes specifically for racial groups
  - Will be larger overall trials
  - Once the ‘majority’ population results are known, is it ethical to continue to enroll minority patients?

- Meta-analyses (pooling data)
  - If individual trials are underpowered, can we combine several trials to get a reliable answer
  - Yes, if those trials are available
  - Current funding/approval models result in little ‘redundancy’
Possible strategies (cont)

- Pre-specify race-based hypotheses based on prior knowledge
  - If differences are identified, they will be more believable
- Increase proportion of minorities on trials
  - Good from every perspective
Summary: Statistics and Race in Oncology Clinical Trials

- Small populations research is extremely difficult, be it by race, age, or any other variable.
- We have an obligation to look for differences, but also to require a high level of evidence.
- Statistics cannot solve this alone, and greater accrual is the single best path forward.