Power and Sample Sizes in RCTs

Cape Town 2007

Kenneth F. Schulz, PhD
Power

**Compare**
- No statistically significant difference between treatments ($p > .05$)

**With**
- The decrease in death rate using the new treatment was not statistically significant ($p > .05$), but we had only a 15% chance of detecting a 50% decrease in death rate
Sufficient Power in RCTs

- **Critical** because they are held in such high esteem (in most circles)
  - ... and even though many do not know how to do them

- **Critical** because the inherent properties of an RCT operate to lower sample sizes
  - Complexities of experimental research
  - Informed consent for experimental research
Special Article

THE IMPORTANCE OF BETA, THE TYPE II ERROR AND SAMPLE SIZE IN THE DESIGN AND INTERPRETATION OF THE RANDOMIZED CONTROL TRIAL

Survey of 71 “Negative” Trials

Jennie A. Freiman, A.B., Thomas C. Chalmers, M.D., Harry Smith, Jr., Ph.D., and Roy R. Kuebler, Ph.D.

Abstract: Seventy-one “negative” randomized control trials were re-examined to determine if the investigators had studied large enough samples to give a high probability (>0.90) of detecting a 25 percent and 50 per cent therapeutic improvement in the response. Sixty-seven of the trials had a greater than 10 per cent risk of missing a true 25 per cent therapeutic improvement, and with the same risk, 50 of the trials could have missed a 50 per cent improvement. Estimates of 90 per cent confidence intervals for the true improvement in each trial showed that in 57 of these “negative” trials, a potential 25 per cent improvement was possible, and 34 of the trials showed a potential 50 per cent improvement. Many of the therapies labeled as “no different from control” in trials using inadequate samples have not received a fair test. Concern for the probability of missing an important therapeutic improvement because of small sample sizes deserves more attention in the planning of clinical trials.

97% of articles did not address Freiman articles

- More than two-thirds of the trials had less than a 90% chance of detecting a 50% improvement (p< 0.05)

Recently, better reporting (Moher et al.), but still similar power
Definitions

\( p_1 = \text{True incidence in exposed} \)
\( p_2 = \text{True incidence in unexposed} \)

\( q_1 = \text{Sample incidence in exposed} \)
\( q_2 = \text{Sample incidence in unexposed} \)

\( d = p_2 - p_1 \)

\( H_0: p_1 = p_2 \quad \text{or} \quad d = 0 \)

\( H_1: p_1 \neq p_2 \quad \text{or} \quad d \neq 0 \)
### True Incidence Rates in a Population (i.e. Everyone)

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
<td>700</td>
<td>700</td>
</tr>
<tr>
<td><strong>No Outcome</strong></td>
<td>9,300</td>
<td>9,300</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>10,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

\[ p_1 = 7\% \quad p_2 = 7\% \]
## Samples from the Population of 20,000

**Study #1**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Outcome</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

$q_1 = 10\%$  
$q_2 = 4\%$

**Study #2**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Outcome</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

$q_1 = 4\%$  
$q_2 = 10\%$
95% CIs for Observed Results

- 2 out of 50 – 4% (0.5% to 14%)
- 5 out of 50 – 10% (3% to 22%)
"Truth" and Sample Results

<table>
<thead>
<tr>
<th>Sample</th>
<th>$q_2 - q_1$</th>
<th>Sample</th>
<th>$q_2 - q_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.06</td>
<td>1</td>
<td>+.05</td>
</tr>
<tr>
<td>2</td>
<td>+.24</td>
<td>2</td>
<td>+.43</td>
</tr>
<tr>
<td>3</td>
<td>+.12</td>
<td>3</td>
<td>+.14</td>
</tr>
<tr>
<td>4</td>
<td>-.10</td>
<td>4</td>
<td>+.23</td>
</tr>
<tr>
<td>5</td>
<td>-.18</td>
<td>5</td>
<td>+.32</td>
</tr>
<tr>
<td>6</td>
<td>+.22</td>
<td>6</td>
<td>+.11</td>
</tr>
<tr>
<td>7</td>
<td>-.09</td>
<td>7</td>
<td>+.15</td>
</tr>
<tr>
<td>8</td>
<td>+.05</td>
<td>8</td>
<td>+.07</td>
</tr>
<tr>
<td>9</td>
<td>-.02</td>
<td>9</td>
<td>+.35</td>
</tr>
</tbody>
</table>
Do We Need To Do Many Studies? Statistics Helps Us Here

\( \hat{p} \) = weighted average of \( q_1 \) and \( q_2 \)

Standard error of \( q_2 - q_1 \) is:

\[
s (q_2 - q_1) = \sqrt{\hat{p} (1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}
\]

Sampling distribution of \( q_2 - q_1 \) is approximately normal

\[
Z = \frac{q_2 - q_1}{s (q_2 - q_1)}
\]

\( H_0: p_1 = p_2 \) (null hypothesis)

\( H_A: p_1 \neq p_2 \) (alternative hypothesis)
Pictorial Representation of a Statistical Test

Testing Distribution

- \( Z_\alpha s \)  
  (e.g. \(-1.96s\))

0.0

+ \( Z_\alpha s \)  
  (e.g. \(+1.96s\))
Power as a Function of $d$

Testing Distribution  
Assumed True Distribution
Power as a Function of d

Testing Distribution

Assumed True Distribution
Power as a Function of $d$

Testing Distribution vs. Assumed True Distribution

0.0 vs. $d$
Power as a Function of $d$

Testing Distribution

Assumed True Distribution

0.0

d
Power as a Function of $d$

- Testing Distribution
- Assumed True Distribution
Power as a Function of the p-Value

Testing Distribution

Assumed True Distribution

0.0

d
Power as a Function of the p-Value

Testing Distribution  

Assumed True Distribution

0.0  d
Power as a Function of the p-Value

Testing Distribution

Assumed True Distribution

0.0  d
Power as a Function of the p-Value

Testing Distribution

Assumed True Distribution

0.0  d
<table>
<thead>
<tr>
<th>α (type 1 error)</th>
<th>Power (1 – β)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.50</td>
<td>0.80</td>
<td>0.90</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>100</td>
<td>200</td>
<td>270</td>
<td>480</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>170</td>
<td>300</td>
<td>390</td>
<td>630</td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>280</td>
<td>440</td>
<td>540</td>
<td>820</td>
<td></td>
</tr>
</tbody>
</table>

Source: The Lancet Vol. 365 page 1350 Schulz, K. and Grimes, D.
Distributions and Increasing Sample Size

Testing Distribution

Assumed True Distribution

0.0  d
Increasing Sample Size
An investigator would like 95% power to show a difference between an incidence rate in the exposed of .04 and an incidence rate in the unexposed of .02 when testing at a level of significance of alpha = .05.