Assessing the Generalizability of Economic Results from Multinational / Multicenter Clinical Trials

John R. Cook
Merck Research Labs

ISPOR Workshop
May 17, 2003
Washington, DC

Acknowledgments
Joseph Heyse
Merck Research Labs
Michael Drummond
University of York
Henry Glick
University of Pennsylvania

Based on manuscript “Assessing the appropriateness of combining economic data from multinational clinical trials.” Statistics in Medicine 2003 (in press)

Introduction
- Trials are increasingly conducted in multiple centers across several countries.
  - Increases representativeness of patients
  - Speeds the trial enrollment
  - Helps approval process by regulatory agencies
- Economic evaluations are increasingly incorporated into these trials
Introduction

- Do the results apply to each country in the trial?
  - **Issue of Generalizability**

- What do we mean by "generalizability"?
  - **Location**: between study countries/non-study countries
  - **Study conditions**: applicable to non-study conditions
  - **Study population**: sample patients representative
  - **Time**: study results applicable today

Introduction

- Do results apply to each country in the trial?
  - **Clinical versus Economic Results**

- Concern due to differences in:
  - Morbidity/Mortality Patterns
  - Practice Patterns
  - Unit Costing

Introduction

- Decision-makers interested in local results
  - Often have little information from a single country
  - Desirable to combine data across countries
  - Appropriate to do so if results are “homogeneous”

- Statistical methods available used to assess
  **homogeneity** of treatment effect
  - Currently used with clinical endpoints
  - Can be used with economic endpoints
Outline

1. The Scand. Simvastatin Survival Study (4S)
2. Tests of Interaction
   - Quantitative vs Qualitative
   - Methods
3. Assess Interaction in 4S
   - Hospitalization Rate
   - CE Ratio / Net Benefit
4. Assessing the Power of the Tests
5. Pooling Results

Scand. Simv. Survival Study (4S)

- Randomized, double-blind, placebo-controlled, N=4444 patients in:
  - Denmark (N=713)
  - Finland (N=868)
  - Iceland (N=157)
  - Norway (N=1025)
  - Sweden (N=1681)
- Patients had previous MI (80%) or angina (40%)
- Patients followed for 5.4 years (median)
- Received cholesterol lowering therapy

Scand. Simv. Survival Study (4S)

- Data analysis plan developed to assess differences in
  - Resource utilization
  - Cost
  - Cost per life year gained
- External cost data (from Sweden) applied to CV hospitalizations (converted to DRGs)
- Several economic evaluations based on 4S data have previously been published
Outline

1. The Scandinavian Simvastatin Survival Study (4S)
2. Tests of Interaction
   - Quantitative vs Qualitative
   - Methods
3. Assess Interaction in 4S
   - Hospitalization Rate
   - CE Ratio / Net Benefits
4. Assessing the Power of the Tests
5. Pooling Results

Quantitative Interaction

Same direction of the treatment effect - “no cross-over”

Qualitative Interaction

Different direction of the treatment effect - “cross-over”
**Test of Quantitative Interaction**  
(Gail & Simon, 1985)

Suppose there are \( K \) countries, each with mean treatment effect \( D_i \) and standard deviation \( S_i \)

- Compute:  
  \[
  H = \sum_{i=1}^{K} \left[ (D_i - \bar{D})^2 / S_i^2 \right]
  \]
  
  where  
  \[
  \bar{D} = \left[ \frac{\sum_{i=1}^{K} D_i}{K} \right] = \left[ \frac{\sum_{i=1}^{K} |D_i|}{K} \right]
  \]

- Compare \( H \) to \( \chi^2 \) with \( K-1 \) d.f.

Large value of \( H \) implies treatment differences exists among countries/centers

---

**Test of Qualitative Interaction #1**  
(Gail & Simon, 1985)

- Compute \( Q^- \) and \( Q^+ \) for positive and negative differences:
  \[
  Q^- = \sum_{i=1}^{K} \left( D_i^2 / S_i^2 \right) I(D_i > 0)
  \]
  \[
  Q^+ = \sum_{i=1}^{K} \left( D_i^2 / S_i^2 \right) I(D_i < 0)
  \]

- Test Statistic:  
  \[
  Q = \text{Min}(Q^+, Q^-) > C
  \]

Large value of \( Q \) implies differences exist in direction of treatment effect among countries

---

**Test of Qualitative Interaction #2**  
(Piantadosi & Gail, 1993; Pan & Wolfe, 1997)

- Construct confidence intervals for each country \((L_i, U_i)\):  
  \[
  D_i \pm Z_{\alpha^*} \cdot S_i \quad \text{for } i = 1, 2, \ldots, K
  \]

  where  
  \[
  \alpha^* = \sqrt{\frac{1-P_K}{2}}
  \]
  \[
  P_K = 2(1-\alpha)^{\frac{1}{k-1}} - 1
  \]

- Qualitative interaction exists if there are two countries \((i \text{ and } j)\) with intervals such that:  
  \[
  U_i < 0 \quad \text{and} \quad L_j > 0
  \]

- Pan & Wolfe discuss relative merit of methods
Outline

1. The Scandinavian Simvastatin Survival Study (4S)
2. Tests of Interaction
   - Quantitative vs Qualitative
   - Methods
3. Assess Interaction in 4S
   - Hospitalization Rate
   - CE Ratio / Net Benefits
4. Assessing the Power of the Tests
5. Pooling Results

4S CV Hospitalizations per Patient Year by Country

<table>
<thead>
<tr>
<th>Country</th>
<th>( D_i )</th>
<th>( S_i )</th>
<th>Relative Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>-0.381</td>
<td>0.085</td>
<td>0.683</td>
</tr>
<tr>
<td>Finland</td>
<td>-0.252</td>
<td>0.084</td>
<td>0.777</td>
</tr>
<tr>
<td>Iceland</td>
<td>-0.498</td>
<td>0.214</td>
<td>0.608</td>
</tr>
<tr>
<td>Norway</td>
<td>-0.369</td>
<td>0.075</td>
<td>0.691</td>
</tr>
<tr>
<td>Sweden</td>
<td>-0.276</td>
<td>0.056</td>
<td>0.759</td>
</tr>
<tr>
<td>OVERALL</td>
<td>-0.316</td>
<td>0.085</td>
<td>0.729</td>
</tr>
</tbody>
</table>
4S CV Hospitalizations

- Quantitative Interaction
  - Gail & Simon: $H=2.91$ (d.f. = 4) $p=0.57$

- Qualitative Interaction
  - Gail & Simon: $Q=\min(Q', Q'')=0$ $p=1.00$
  - Plantadosi &Gail: not significant (all confidence interval limits $<0$)

Cost-Effectiveness Ratios

- Homogeneity among countries in costs and effects does not imply homogeneity in the ratio

- Challenges with the CE ratio
  - Analytic challenges:
    - Lack of uniqueness with ratio
    - When $\Delta E = 0$
  - Conceptual challenge:
    - What is a qualitative interaction?

Cost-Effectiveness Ratios

- What is a qualitative interaction for the ratio?
  - Requires specification of CE threshold ($\lambda$)
    - CE ratios “below” $\lambda$ are deemed cost-effective
    - CE ratios “above” $\lambda$ are deemed not cost-effective
  - Qualitative interaction exists if some countries are cost-effective, while others are not
  - Modify tests:
    - replace $D_i$ with $(D_i - \lambda)$ or $(D_i - f(\lambda))$
### ISPOR03_22.ppt

**4S CE Ratio**

**Incremental Costs and Effectiveness**

- **Incremental Survival Probability**
  - Range: 0.00 to 0.05
- **Incremental Cost (U.S.$)**
  - Range: 0 to 2500

- **Country-specific treatment effect based on angular transformation** (Cook & Heyse, 2000)

<table>
<thead>
<tr>
<th>Country</th>
<th>$\theta_i$</th>
<th>$\theta_S$</th>
<th>CE Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>71.5</td>
<td>11.0</td>
<td>47,032</td>
</tr>
<tr>
<td>Finland</td>
<td>82.0</td>
<td>6.6</td>
<td>120,963</td>
</tr>
<tr>
<td>Iceland</td>
<td>81.6</td>
<td>16.0</td>
<td>129,474</td>
</tr>
<tr>
<td>Norway</td>
<td>70.6</td>
<td>7.6</td>
<td>43,526</td>
</tr>
<tr>
<td>Sweden</td>
<td>75.0</td>
<td>5.8</td>
<td>58,208</td>
</tr>
</tbody>
</table>

### ISPOR03_23.ppt

**4S CE Ratio**

- **Gail & Simon:** $Q = \min(Q, Q') = 0.352$  
  - $p = 0.75$

**Qualitative Interaction**

- Depends on threshold: assume $\lambda = 75,000$
- **Gail & Simon:** $Q = \min(Q, Q') = 0.352$  
  - $p = 0.75$

### ISPOR03_24.ppt

**4S CE Ratio**

- **Quantitative Interaction**
  - $H = 1.68$ (d.f. = 4)  
  - $p = 0.79$
4S CE Ratio

Qualitative Interaction (Plantadosi & Gail)

- Transform \( \lambda \) (75,000) to angle (78.4°)
- Construct confidence intervals (w/ \( \alpha = .2 \))
  - Denmark (48.2°, 94.8°)
  - Finland (61.1°, 102.9°)
  - Iceland (45.6°, 117.7°)
  - Norway (54.9°, 86.3°)
  - Sweden (63.4°, 86.7°)
- No significant qualitative interaction
  - no CI with lower bound > 78.4°
  - no CI with upper bound < 78.4°

4S Net Benefits (Monetary)

- Approach requires specification of \( \lambda \) value
  - for estimation
  - for all tests of interaction
- Given \( \lambda \), can estimate monetary benefit for each patient:
  \[ \lambda \cdot I(\text{Survived}) - \text{Cost} \]
- Note: with individual patient responses, can test for interaction using ANOVA model

4S Net Benefits (Monetary)

- For \( \lambda = 75,000 \) ($/Survivor):

<table>
<thead>
<tr>
<th>Country</th>
<th>( D_i )</th>
<th>( S_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>1208.8</td>
<td>1964.7</td>
</tr>
<tr>
<td>Finland</td>
<td>-850.2</td>
<td>1306.2</td>
</tr>
<tr>
<td>Iceland</td>
<td>-636.5</td>
<td>3404.0</td>
</tr>
<tr>
<td>Norway</td>
<td>1370.9</td>
<td>1486.8</td>
</tr>
<tr>
<td>Sweden</td>
<td>542.3</td>
<td>1105.7</td>
</tr>
</tbody>
</table>
4S Net Benefits (Monetary)

- Quantitative Interaction
  - Gail & Simon: $H = 1.62$ (d.f. = 4), $p = 0.81$

- Qualitative Interaction ($\lambda = 75,000$)
  - Gail & Simon: $Q = \min(Q', Q'') = 0.459$, $p = 0.72$
  - Piantadosi & Gail: not significant (all CI’s include 0)

Outline

1. The Scandinavian Simvastatin Survival Study (4S)
2. Tests of Interaction
   - Quantitative vs Qualitative
   - Methods
3. Assess Interaction in 4S
   - Hospitalization Rate
   - CE Ratio / Net Benefits
4. Assessing the Power of the Tests
5. Pooling Results

Assessment of Power

- How confident are we in the test results?
  - Tests for interaction typically have low power
  - Cost estimates often have large variance

- Investigate with ex-post power calculation
  - Pan & Wolfe ('97) provide method to assess power for qualitative interaction
  - Power depends on magnitude of ‘important’ differences (what does “truth” look like?)
4S: Assessment of Power

- Estimate power for Net Monetary Benefit:
  - \( \lambda = 75,000 \) ($/Survivor)
  - use observed standard errors for each country

<table>
<thead>
<tr>
<th>Net Monetary Benefit</th>
<th>( \alpha = 0.20 )</th>
<th>( \alpha = 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta = 750 )</td>
<td>9.8</td>
<td>1.0</td>
</tr>
<tr>
<td>( \delta = 1500 )</td>
<td>27.3</td>
<td>5.7</td>
</tr>
<tr>
<td>( \delta = 3000 )</td>
<td>79.3</td>
<td>48.7</td>
</tr>
</tbody>
</table>

Assessment of Power

- Why might the power be low?
  - Sample Size?
    - Hopefully not in 4S
  - Variability in NMB \( (\lambda \Delta S - \Delta C) \)?
    - Don’t always blame variance of cost difference
    - In 4S, \( SE(\lambda \Delta S) \) was 4X-5X as large as \( SE(\Delta C) \)
    - If \( SE(\lambda \Delta S) \) was as small as \( SE(\Delta C) \),
      - power for \( \delta = 750 \) would be 58.2%

Assessment of Power

- If there are many countries in the clinical trial, it may be difficult to detect true heterogeneity
  - sample size per country is small
  - consider pooling countries based on “meaningful” covariate (region, health care system)
- Even if there are few countries, may want to conduct “sensitivity analysis”
  - Pool “qualitatively” equivalent countries
Outline

1. The Scandinavian Simvastatin Survival Study (4S)
2. Tests of Interaction
   - Quantitative vs Qualitative
   - Methods
3. Assess Interaction in 4S
   - Hospitalization Rate
   - CE Ratio / Net Benefit
4. Assessing the Power of the Tests
5. Pooling Results

To Pool or Not to Pool?

- No Interaction detected: Check power and obtain pooled overall estimate.

- Interaction detected:
  - Quantitative Interaction: Inferences about the direction of a treatment effect are O.K.
  - Qualitative Interaction: Rarely observed. Need to investigate reasons.

To Pool or Not to Pool?

- Many approaches to pooling across countries
  - Ignore countries (pool across patients)
  - Simple average across countries
  - Weight country estimates by sample size
  - Weight country estimates by inverse of variance

- Caution: don’t pool CE ratios without applying angular transformation
4S Net Monetary Benefit
Alternative Pooling Methods

For $\lambda = $75,000/survivor

- Ignore countries  $\text{NMB} = $525
- Average across countries  $327
- Weight by sample size  $527
- Weight by inverse of variance  $379

Concluding Remarks

- Focus attention on assessing country-to-country differences
  - Readily available statistical methods exist
  - Less consensus on how to pool if appropriate
  - Test may be informative for decision makers in non-study countries

- Design issues need more consideration.
  - Sample size
  - Country and clinic selection
  - Costing methodology

References

  (see correction in Statistica in Medicine (1998), 17, 2015-2016.)