Good Value for Cost

• Economic data collected as secondary (or primary) endpoint in randomized trials commonly used in evaluation of value for cost
  – Short-term economic impacts directly observed
    • Within-trial analysis
  – Longer term impacts potentially projected by use of decision analysis
    • Long term projection
  – Reported results: point estimates and confidence intervals for estimates of:
    • Incremental costs and outcomes
    • Comparison of costs and effects

Sample Results Table

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incremental Cost</td>
<td>-713</td>
<td>-2123 to 783</td>
</tr>
<tr>
<td>Incremental QALYs</td>
<td>0.13</td>
<td>0.07 to 0.18</td>
</tr>
<tr>
<td>Cost-Effectiveness Analysis</td>
<td>Dominates</td>
<td>Dom to 6650</td>
</tr>
<tr>
<td>Principal Analysis</td>
<td>Dominates</td>
<td>Dom to 9050</td>
</tr>
<tr>
<td>Survival Benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-33%</td>
<td>Dominates</td>
<td>Dom to 5800</td>
</tr>
<tr>
<td>+33%</td>
<td>Dominates</td>
<td>Dom to 4850</td>
</tr>
<tr>
<td>Drug Cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>Dom to 8750</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>Dom to 6350</td>
</tr>
<tr>
<td>Discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>Dominates</td>
<td>Dom to 7000</td>
</tr>
<tr>
<td>7%</td>
<td>Dominates</td>
<td>Dom to 7000</td>
</tr>
</tbody>
</table>
Outline

• Steps in economic evaluation
• Gold standard and its tensions
• 5 Strategic issues

Steps in Economic Evaluation

Step 1: Quantify costs of care
Step 2: Quantify outcomes
Step 3: Assess whether and by how much average costs and outcomes differ among treatment groups
Step 4: Compare magnitude of difference in costs and outcomes and evaluate “value for costs”
  – e.g. by reporting a cost effectiveness ratio, net monetary benefit, or probability that ratio is acceptable
  – Potential hypothesis: Cost per quality-adjusted life year saved significantly less than $75,000
Step 5: Perform sensitivity analysis

Ideal Economic Evaluation Within a Trial

• Conducted in naturalistic settings
  – Compares therapy with other commonly used therapies
  – Studies therapy as it would be used in usual care
• Well powered for:
  – Average effects
  – Subgroup effects
• Designed with an adequate length of follow-up
  – Allows assessment of full impact of therapy
• Timely
  – Can inform important decisions in adoption and dissemination of therapy
Ideal Economic Evaluation Within a Trial (II)

• Measure all costs of all participants prior to randomization and for duration of follow-up
  – Costs after randomization—cost outcome
  – Costs prior to randomization—potential predictor
• Independent of reasons for costs
• Most feasible when:
  – Easy to identify when services are provided
  – Service/cost data already being collected
  – Ready access to data

Design Issues Not Unique To Trials

• A number of design issues apply equally to economic evaluations in clinical trials and to other economic evaluations:
  – Type of analysis that will be conducted
  – Types of costs that will be included
  – Study perspective
• Issues well addressed in literature

Difficulties Achieving an Ideal Evaluation

• Settings often controlled
• Comparator isn’t always most commonly used therapy or currently most cost-effective
• Investigators haven’t always fully learned how to use new therapy under study
• Sample size required to answer economic questions may be greater than sample size required for clinical questions
• Ideal length of follow-up needed to answer economic questions may be longer than follow-up required for clinical questions
Trade-off
• May be only source of information needed for important early decisions about therapy adoption and diffusion

TRADE-OFF: Ideal vs best feasible

5 Strategic Issues
• What medical service use should we collect?
• What is appropriate sample size?
• How naturalistic should study design be?
• Is there a treatment-by-time interaction?
• How should we interpret results from multicenter / multinational studies?

Issue #1: What Medical Service Use Should We Collect?
• Real/perceived problem
  – Don’t have sufficient resources to track all medical service use
  – Don’t always expect to affect all medical service use
Limited Data Collection Resources

- Availability of administrative data may reduce costs of tracking all medical service use
- If administrative data are unavailable:
  - Measure services that make up a large portion of difference in treatment between patients randomized to different therapies under study
    - Provides estimate of cost impact of therapy
  - Measure services that make up a large portion of total cost
    - Minimizing unmeasured services reduces likelihood that differences among them will lead to biased estimates
    - Provides measure of overall variability

Measure as Much as Possible

- Best approach: measure as many services as possible
  - No a priori guidelines about how much data are enough
  - Little to no data on incremental value of specific items in economic case report form
- While accounting for expense of collecting particular data items
  - E.g., collecting 6700 blood gas tests that accounted for 1.8% of lab costs vs 420 cardiac studies that represented 4.3%

Document Likely Service Use During Trial Design

- Can improve decisions by documenting types of services used by patients who are similar to those who will be enrolled in trial
  - Review medical charts or administrative data sets
  - Survey patients and experts about kinds of care received
  - Have patients keep logs of their use of medical services
- Guard against possibility that new therapy will induce alternative patterns of medical service use
Limit Data to Disease-Related Services?

- Little if any evidence about accuracy, reliability, or validity of such judgments
- Easy for judgments to be flawed
- Investigators routinely attribute AEs to intervention, even when participants received vehicle/placebo
- Medical practice often multifactorial: modifying disease in one body system may affect disease in another body system
  - In Studies of Left Ventricular Dysfunction, hospitalizations "for heart failure" (and death) reduced by 30% (p<0.0001)
  - Hospitalizations for noncardiovascular reasons reduced 14% (p = 0.006)

Other Types of Costs

- Other types of costs that sometimes are documented within economic evaluations include:
  - Time costs: Lost due to illness or to treatment
  - Intangible costs
- Types of costs that should be included in an analysis depend on:
  - What is affected by illness and its treatment
  - What is of interest to decision makers
  - e.g., National Institute for Clinical Excellence (U.K.) and Australian Pharmaceutical Benefits Scheme have indicated they have little interest in time costs

General Recommendations

- General Strategy: Identify a set of medical services for collection, and assess them any time they are used, independent of reason for their use
- Decision to collect service use independent of its reason does not preclude ADDITIONAL analyses testing whether designated "disease-related" costs differ
Specific Recommendations, Which Services

- Identify common patterns of medical service use in centers/countries that will participate in trials
  - Speak with experts in multiple centers/countries
  - Focus groups, etc.
- Design case report forms to collect important, common medical service use
- Collect services independent of reason for their use
- Pilot test forms
- Consider collecting costs other than medical service use

Issue #2. What Sized Sample Should We Study?

- A goal of sample size and power calculation for cost-effectiveness analysis is to identify likelihood that an experiment will allow us to be confident that a therapy is good or bad value when we adopt a particular willingness to pay
  - e.g., We:
    - May expect a point estimate for cost-effectiveness ratio of 20,000 per QALY
    - May be willing to pay at most 75,000 per QALY
    - Want an experiment that provides an 80% chance (i.e., power) to be 95% confident (alpha) that therapy is good value

Sample Size Formula

- At most basic level, sample size for cost-effectiveness is calculated using same formula as used for sample size for a difference in any continuous variable:
  \[ n = \frac{2(z_{\alpha/2} + z_{\beta})^2 \cdot \text{sd}_{\text{min}}^2}{\Delta \text{nmb}^2} \]

where
- \( n \) = sample size/group;
- \( z_{\alpha/2} \) and \( z_{\beta} \) = z-statistics for \( \alpha \) (e.g., 1.96) and \( \beta \) (e.g., 0.84) errors;
- \( \text{sd}_{\text{min}} \) = standard deviation for NMB;
- \( \Delta \text{nmb} \) = expected difference in NMB
Sample Size Formula (2)

- Complexities arise because 1) difference being assessed is difference in NMB \((W\Delta Q - \Delta C)\) and 2) standard deviation of NMB is a complicated formula.
- Data needed to calculate sample size include:
  - Difference in cost
  - SD, difference in cost
  - Difference in effect
  - SD, difference in effect
  - \(Z_\alpha\) and \(Z_\beta\)
  - Correlation of difference in cost and effect
  - Willingness to pay

Full Formula

\[ n = \frac{2(z_\alpha + z_\beta)^2 \left( sd_c^2 + (W sd_q \rho sd_d^2) - 2 W p sd_q sd_d \right)}{(W\Delta Q - \Delta C)^2} \]

Correlation of Difference

- When increasing effects are associated with decreasing costs, a therapy is characterized by a negative (win/win) correlation between difference in cost and effect
  - e.g., asthma care
- When increasing effects are associated with increasing costs, a therapy is characterized by a positive (win/lose) correlation between difference in cost and effect
  - e.g., life-saving care
- All else equal, fewer patients need to be enrolled when therapies are characterized by a positive correlation than when therapies are characterized by negative correlation.
Ability to Shift W

- W is to cost-effectiveness analysis as 1 is to OR and RR
  - It is decision threshold we are trying to rule out if we are to have confidence about value
- While we rarely consider comparing OR and RR to a decision threshold other than 1 (noninferiority trials may be exception), we often choose W because in many countries there is no clear consensus on what its value is
- Moving W “nearer to” or “further away from” expected point estimate reduces or increases power to be confident of value
- Caution: “Nearer” and “further away” are not measured on real number line
  → Sample size need NOT decrease as WTP increases

WTP and Point Estimate

- When WTP is greater than expected point estimate, resulting sample size and power are for experiments that allow us to be confident that therapy is good value
  - Because confidence statements from these trials will be that point estimate is less than willingness to pay
- When WTP is less than expected point estimate, resulting sample size and power are for experiments that allow us to be confident that therapy is bad value
  - Because confidence statements from these trials will be that point estimate is greater than willingness to pay

Effect of SDq vs SDc on Sample Size

- Commonly thought that sample size for cost-effectiveness driven more by standard deviation for cost than it is by SD for effect
  - If not, why would we need a larger sample for economic outcome than we do for clinical outcome?
- However, if willingness to pay is substantially greater than standard deviation for cost, percentage changes in QALY SD can have a substantially greater effect on sample size than do equivalent percentage changes in cost SD
Sample Size Tables, SD

<table>
<thead>
<tr>
<th>SDc</th>
<th>N/Group</th>
<th>SDq</th>
<th>N/Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500</td>
<td>306</td>
<td>0.1</td>
<td>114</td>
</tr>
<tr>
<td>5000</td>
<td>340</td>
<td>0.2</td>
<td>340</td>
</tr>
<tr>
<td>7500</td>
<td>389</td>
<td>0.3</td>
<td>710</td>
</tr>
<tr>
<td>10,000</td>
<td>455</td>
<td>0.4</td>
<td>1224</td>
</tr>
<tr>
<td>15,000</td>
<td>634</td>
<td>0.6</td>
<td>2685</td>
</tr>
</tbody>
</table>

$\Delta C=250; \Delta Q=0.05$; unless otherwise specified, $sd_c=5000; \quad sd_q=0.2; \rho=-0.1; \alpha=0.05; \beta=0.8$

Drop Out

- Sample size estimates are appropriate if we expect no dropout from trial
- If anticipate 10% dropout, will want to divide sample size estimates by 0.9

"Typical" Sample Size Table, W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Exp 1 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
</tr>
</tbody>
</table>

* $\Delta C=-10; \Delta Q=0.05; \quad sd_c=5000; \quad sd_q=0.2; \rho=-0.1; \alpha=0.05; \beta=0.8$
### Sample Size Can Be Increasing with Increasing $W$

<table>
<thead>
<tr>
<th>WTP (€k)</th>
<th>Exp 1</th>
<th>Exp 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
<td>560</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
<td>831</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
<td>1527</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
<td>2573</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
<td>3717</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
<td>6059</td>
</tr>
</tbody>
</table>

* $\Delta C = -250; \Delta Q = 0.001; sd_c = 1000; sd_q = 0.05; \rho = -0.3; \alpha = 0.05; \beta = 0.8$

### Sample Size Not Necessarily Monotonic With $W$

| WTP (€k) | Exp 1 | Exp 2 | Exp 3 *
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
<td>560</td>
<td>3383</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
<td>831</td>
<td>3205</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
<td>1527</td>
<td>3167</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
<td>2573</td>
<td>3185</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
<td>3717</td>
<td>3204</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
<td>6059</td>
<td>3229</td>
</tr>
</tbody>
</table>

* $\Delta C = -40; \Delta Q = 0.04; sd_c = 5500; sd_q = 0.58; \rho = 0.1; \alpha = 0.05; \beta = 0.8$

### Six Patterns of Power Associated with $W$

[Graphs showing six patterns of power associated with $W$.]
Economic Vs Clinical Sample Sizes

- Sample size required to answer economic questions often larger than sample size required to answer clinical questions
  - But it need not be
- $\Delta C$ and $\Delta Q$ are a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome
- In same way that we can have more power for joint cardiovascular outcome than either individual outcome alone, we can have more power for cost-effectiveness than we do for costs or effects alone

Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require stipulation of willingness to pay for a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- Can calculate a cost-effectiveness ratio for any outcome (e.g., cost/case detected; cost/abstinence day), but to be informative, outcome must be one for which we have recognized benchmarks of cost-effectiveness
  - Argues against use of too disease-specific an outcome for economic assessment

Issue # 3. How Naturalistic?

- Primary purpose of cost-effectiveness analysis:
  - Inform real-world decision-makers about how to respond to real-world health care needs
- Greater naturalism, in terms of participants, analysis based on intention to treat, and limitation of loss to follow-up, implies greater likelihood that data developed within trial will speak directly to decision question
#3a. Intention to Treat

- Economic questions relate to treatment decisions (e.g., whether to prescribe a therapy), not whether patient received drug prescribed nor whether, once they started prescribed drug, they were switched to other drugs
  - Implication: costs and effects associated with later decisions should be attributed to initial treatment decision
- Thus, trial-based cost-effectiveness analyses should adopt an intention-to-treat design

#3b. Loss to Follow-up

- Trials should be designed to minimize occurrence of missing data
  - Study designs should include plans to aggressively pursue participants and data throughout trial
  - Strategies may include:
    1) intensive outreach to reschedule assessment, followed by
    2) telephone assessment, followed by
    3) interview of a proxy who has been identified and consented at time of randomization

Loss to Follow-up (2)

- Investigators should also ensure that:
  - Follow-up continues until end of study period
  - Data collection isn’t discontinued simply because a participant reaches a clinical or treatment stage such as failure to respond
    - Often occurs in antibiotic, cancer chemotherapy, and psychiatric drug trials
  - Given failure commonly associated with change in pattern of costs, discontinuation likely to bias results
#3c. Protocol-Induced Costs and Effects

- Common concerns:
  - Standardization of care in clinical trial protocols often means care delivered in trials differs from usual care
    - e.g., protocol may require substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice
  - Protocols often prescribe aggressive documentation and treatment of potential adverse effects that differ from usual care
  - Omit these costs???

Omission of Protocol-Induced Costs?

- Criterion for including costs NOT "Would services have been provided in usual care?"
  - Should be: "Could services have affected care / outcomes (and thus costs)?"
- No problem omitting services that cannot affect care / services
  - e.g., Cost of genetic samples that will not be analyzed until after follow-up is completed
- More problematic to omit services that can change treatment and affect outcome
  - "Cadillac" costs may yield "Cadillac" outcomes
  - Would need to adjust BOTH costs and their effects on outcomes

Biases?

- Protocol-induced testing may bias testing cost to null
  - Might be a difference in testing in usual care, but can't be observed if everyone routinely receives a test
- Protocol induced testing may bias treatment cost / outcome in unknown direction
  - Trial's extra testing may lead to:
    - Detection and treatment of outcomes that wouldn't be detected or treated in usual care
    - Earlier detection and treatment of problems when they are less severe and easier to treat
  - Adjustment requires assumptions about what would or wouldn't have been detected in usual care
Specific Recommendations, Naturalism

• Use intention to treat sample for economic analysis
• Be aggressive in maintaining follow-up, including continuing to collect data on those who fail or switch therapy
• Use appropriate analytic methods to address missing data if and when they occur
• When possible, minimize effect of protocol on patient care

Issue #4. Is There a Treatment-By-Time Interaction?

• When trial observes cost-effectiveness for time-limited period (e.g., 2 or 3 years), but therapy will be taken for lifetime, consider likelihood that cost-effectiveness ratio observed in trial will describe longer term therapy
  – Referred to as a treatment-by-time interaction

Likelihood of a Treatment-by-Time Interaction

• Treatment-by-time interaction less likely to be substantial when intervention’s cost and outcome begin at approximately same time and continue to be incurred together over time (e.g., drug therapy for heart failure)
• Interaction more likely to be substantial when:
  – Treatment cost and outcome incurred over time, but:
    • Outcome delayed for a number of years (e.g., risk reduction from cholesterol-modifying therapy) OR increasing with time
  – Treatment cost incurred initially (e.g., surgical removal of tumor) and outcome (e.g., survival) accrued over time
Addressing a Treatment-By-Time Interaction

• Evaluate what was observed during trial (within-trial analysis)
• Develop decision analytic models to make projections beyond period of observation (projection)

Strengths and Weaknesses

• Within-trial analysis and longer term projections have opposing strengths/weaknesses:
  – Within-trial: More certain of what was observed during trial, but follow-up may be too short to capture most important impacts of therapy
  – Long term: Less certain about projection beyond trial, but projection attempts to quantify what may be most important impacts of therapy

Within-Trial Analysis

• Even if primary analysis will be a projection beyond period of observation, should still evaluate costs and outcomes observed during trial
• In within-trial evaluation, should maintain same time horizons for costs and outcomes observed in trial (e.g., if follow-up for trial was for one year, costs and effects should be measured for one year)
• Not always easy to demonstrate cost-effectiveness in a within-trial analysis
  – e.g., no within-trial analysis of cholesterol-modifying therapy has demonstrated reasonable cost-effectiveness
Longer-Term Projection

• To investigate whether cost-effectiveness ratio is homogeneous with respect to time, should also project results for longer periods
• For projection: Maintain a common time horizon for both costs and effects
  – Some studies have used cost difference observed within trial; argued that benefits of therapy extend beyond trial; and incorporated benefits but not costs from beyond trial
  – e.g., West of Scotland Coronary Prevention Study
    • If therapy has downstream benefits that have been inadequately captured, most likely also has downstream costs that have been inadequately captured

Time Horizon for Projection

• Because length of projection (should be) inversely related to certainty of results, should make projections for different time horizons
  – Even if time horizon of lifetime projection is 30-40 years, may observe projected long-term cost-effectiveness asymptotically approaching life-time ratio after only 5 or 10 years of projection
• To increase face validity, if sufficient follow-up during trial, make estimates for differing lengths of follow-up during trial (e.g., first year, first 2 years, etc.)

CER And CI Within Trial and Projected

<table>
<thead>
<tr>
<th>Years of Follow-up</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within the trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Dominated</td>
<td>168,884 to Dominated</td>
</tr>
<tr>
<td>2</td>
<td>282,857</td>
<td>45,577 to Dominated</td>
</tr>
<tr>
<td>3</td>
<td>73,529</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>4</td>
<td>12,074</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>5</td>
<td>15,258</td>
<td>Dominates to 122,772</td>
</tr>
<tr>
<td>Longer term projection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12,246</td>
<td>Dominates to 42,263</td>
</tr>
<tr>
<td>15</td>
<td>8,578</td>
<td>Dominates to 26,721</td>
</tr>
<tr>
<td>20</td>
<td>7,320</td>
<td>681 to 21,841</td>
</tr>
</tbody>
</table>
Specific Recommendations, Treatment-by-Time Interaction

- Evaluate whether a strong treatment-by-time interaction is expected
- If yes, consider development of a decision model to assess potential magnitude of interaction
  - Substantial amounts of data used for decision model should be derived from trial
  - Where necessary, augment data from trial with epidemiologic data on long term outcomes, etc.

Issue # 5. How Should We Interpret Results From Multicenter (Multinational) Trials?

- Problem:
  - Growing concern that pooled (i.e., average) economic results from multinational trials not be reflective of results that would be observed in individual centers/countries that participated in trial
  - Similar issues for any subgroup in trial (e.g., more and less severely ill patients)

Common Sources of Concern

- Transcenter/transnational differences in morbidity/mortality patterns; practice patterns (i.e., medical service use); and absolute and relative prices for service use (i.e., price weights)
- Decision makers may find it difficult to draw conclusions about value of therapies evaluated in multicenter/multinational trials
Bad Solutions

- Use trial-wide clinical results, trial-wide medical service use, and price weights from one country
  - e.g., to tailor results to U.S., just use U.S. price weights, and conduct analysis as if all participants were treated in U.S.
- Use trial-wide clinical results and use costs derived from subset of patients treated in country
- Ignore fact that clinical and economic outcomes may influence one another (cost affects practice which affects outcome; practice affects outcome which affects cost)

Impact of Price Weights vs Other Variation

<table>
<thead>
<tr>
<th>Country</th>
<th>Trial-Wide Effects</th>
<th>Country-Specific Costs and Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price weight</td>
<td>Country-specific Costs</td>
</tr>
<tr>
<td>1</td>
<td>40,818</td>
<td>5921</td>
</tr>
<tr>
<td>2</td>
<td>57,636</td>
<td>91,906</td>
</tr>
<tr>
<td>3</td>
<td>53,891</td>
<td>90,487</td>
</tr>
<tr>
<td>4</td>
<td>69,145</td>
<td>93,326</td>
</tr>
<tr>
<td>5</td>
<td>65,800</td>
<td>**</td>
</tr>
<tr>
<td>Overall</td>
<td>45,892</td>
<td>45,892</td>
</tr>
</tbody>
</table>

** Country-specific resource use / Country-specific price weights
** New therapy dominates

Two Approaches To Transferability

- Two approaches -- which rely principally on data from trial -- have made their way into literature
  - Hypothesis tests of homogeneity (Cook et al.)
  - Multi-level random-effects model shrinkage estimators

Hypothesis Tests Of Homogeneity

• Evaluate homogeneity of results from different countries
  – If no evidence of heterogeneity (i.e., a nonsignificant p-value for test of homogeneity), and test considered powerful enough to rule out economically meaningful differences in costs, can’t reject that pooled economic result from trial applies to all of countries that participated in trial
  – If evidence of heterogeneity, should not use pooled estimate to represent result for individual countries
    • Method less clear about result that should be used instead

Estimation

• Multi-level random-effects model shrinkage estimation assesses whether:
  – Observed differences between countries are likely to have arisen simply because we have divided trial-wide sample into subsets VS
  – Whether they are likely to have arisen due to systematic differences between countries
• Borrows information from mean estimate to add precision to country-specific estimates
• Methods have potential added advantage of providing better estimates of uncertainty surrounding pooled result than naive estimates of trial-wide result

Specific Recommendations, Transferability

• Identify method for use in evaluation of transferability (e.g., homogeneity testing)
• If evaluating homogeneity, test for several pre-specified levels of willingness to pay including: point estimate of pooled cost-effectiveness ratio as well as $USD 25,000, 50,000, 75,000, and $100,000
• Given limited power of methods for assessing transferability, consider aggregating countries into regions and performing tests at regional level
• Consider using same techniques to evaluate homogeneity of economic findings for important clinical subgroups
Summary

- Clinical trials may provide best opportunity for developing information about therapy’s value for cost early in product life
- When appropriate types of data are collected and analyzed appropriately, provide data about uncertainties related to assessment of value for cost of new therapies
- Potentially useful for policy makers, drug manufacturers, health care providers and patients