The Cost-Effectiveness of Cost-Effectiveness Analysis in Comparative Effectiveness Research

Henry Bergquist¹, Jalpa A. Doshi¹, Lee Fleisher¹, Henry A. Glick¹, Bruce Kinosian¹, Sean McElligott¹, Eleanor M. Perfetto² (Co-PI), Daniel E. Polsky¹, Mark V. Pauly¹ (Co-PI), J. Sanford Schwartz¹ (PI), Richard Willke² (Co-PI)

Affiliations: 1 University of Pennsylvania, 2 Pfizer
Funding: Penn-Pfizer Alliance

Motivation

- Comparative effectiveness research (CER) seeks to assess an intervention’s incremental clinical benefit
- Economic efficiency requires consideration of real resource costs and clinical benefit
  - Cost-effectiveness analysis (CEA), as expressed by the incremental cost-effectiveness ratio (ICER), is a well developed method to assess an intervention’s incremental costs and benefits
- Decision making based on CER alone can lead to adoption of expensive interventions with only small incremental clinical benefit
- But explicit use of cost information and formal CEA for clinical and policy decision making is contentious in the U.S.

Concerns and Contradictions in US CER / CEA

Clinical and Policy Decision Making

- Wilensky: Politically unwise to consider costs explicitly in CER
- Garber: Using CER alone to make decisions like ordering from a menu without prices
  - May be some circumstances in which costs can be included in CER indirectly
- Current Law: Limits conduct and use CEA in CER (and considerable political pressures not to do so)
  - Selby: “Correct” that PCORI not consider costs
- Federal agencies: USPSTF and Medicare prohibited from consideration of costs and cost-effectiveness in recommendations and policies; ACIP, VA and NIH expert guideline panels are not so prohibited

Study Aims

- To address public/political cost-effectiveness “fatigue”, identify predictors of agreement between CER decision making and CEA decision making (i.e., identify circumstances when CEA information provides little incremental value to CER information)
  - Are there systematic characteristics of interventions that can be identified a priori that predict when CER decision making is sufficient?
- Empirical analysis: Examine a set of CEA studies to see how frequently and under what circumstances consideration of cost information in conjunction with clinical information lead to the same choice as a decision based on clinical information alone

Study Sample

- Study sample drawn from the Tufts University Center for the Evaluation of Risk in Health CEA Registry

<table>
<thead>
<tr>
<th>Exclusion Criterion</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial sample</td>
<td>6793</td>
</tr>
<tr>
<td>Non-US studies</td>
<td>3718</td>
</tr>
<tr>
<td>Studies conducted prior to 1990</td>
<td>87</td>
</tr>
<tr>
<td>Missing either ICER or QALYs</td>
<td>1065</td>
</tr>
</tbody>
</table>

TOTAL: 1923 studies

Agreement

- Main outcome: Binary variable representing agreement and disagreement between adoption recommendation from CER and adoption recommendation from CEA
- CER Recommendation: Adopt therapy with the larger point estimate for effectiveness
  - In most formal research, effectiveness measures will be disease-specific clinical outcomes
    - e.g., changes in HbA1c, mm/Hg of blood pressure, or mmol/l of cholesterol
  - In current study, effectiveness measure is QALYs derived from the denominator of the cost/QALY ratio
Agreement (2)

- Cost-effectiveness analysis: Compares difference in cost with difference in effect between pairs of therapies
  \[ \text{ICER} = \frac{(C_1 + C_0)}{(E_1 + E_0)} \]
- Ratio generally interpreted as the extra payment per extra unit of effectiveness for more effective therapy
- CEA Recommendation: Adopt therapy that is good value based on the point estimate for the ICER (2010 US$ per QALY) and a pre-specified WTP threshold
  - $100,000 per QALY with $50,000 sensitivity analysis
  - e.g., Adopt more effective therapy if it does more and costs less than alternative or has an ICER<100,000
  - Adopt less effective therapy if more effective therapy has an ICER>100,000

Don’t Distinguish Between “Types” of Agreement

Explanatory variables

- Type of intervention (9)
  - Surgical (index), care delivery, device, diagnostic, health education, medical procedure, pharmaceutical, screening, other
- Disease category (12)
  - Cardiovascular (index), cancer, digestive, endocrine, environmental, infectious, maternal and perinatal, musculoskeletal, neuropsychiatric, respiratory, sensory organs, other
- Source of funding
  - Industry vs other

Explanatory variables (2)

- Prevention stage
  - Primary: Methods used to prevent disease or illness
  - Secondary: Methods used to diagnose and treat disease in early stages before causing significant morbidity
  - Tertiary: Methods used to reduce negative impact of disease by restoring function and reducing disease-related complications

Explanatory variables (3)

- “Publicness” of disease
  - Google trends: Relative search volume for the 12 disease categories
- Year of study
- Research “intensity”
  - Clinicaltrials.gov: Mapped 93,722 clinical studies by 289 MeSH terms into the 12 disease categories
  - Used the number of studies in each disease category to define inverse probability weights that were used in all models
  - Diseases with smaller numbers of studies received more weight

Logistic Regression Agreement Models

- 3 Logistic regression models
- Model 1: Agreement as a function of
  - Type of intervention
  - Disease category
  - Prevention stage
  - Funding source
  - “Publicness of disease”
- Model 2: Model 1 + year fixed effects
- Model 3: Model 2 + interactions between diseases and Google trends “Publicness”
- Report odds ratios
  - OR < 1 CER/CEA more likely to disagree; OR > 1 CER/CEA more likely to agree
Unadjusted Agreement ($100k), Overall and By Type of Intervention

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>~Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, N, (%)</td>
<td>1338</td>
<td>585</td>
</tr>
<tr>
<td>Type of intervention (p=0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>Care delivery</td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td>Device</td>
<td>47</td>
<td>13</td>
</tr>
<tr>
<td>Least agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>552</td>
<td>256</td>
</tr>
<tr>
<td>Screening</td>
<td>193</td>
<td>107</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>55</td>
<td>44</td>
</tr>
</tbody>
</table>

Unadjusted Agreement, Disease Category

<table>
<thead>
<tr>
<th>Disease category (p&lt;0.0001)</th>
<th>Agree</th>
<th>~Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>293</td>
<td>95</td>
</tr>
<tr>
<td>Sensory organs</td>
<td>44</td>
<td>15</td>
</tr>
<tr>
<td>Infectious &amp; parasitic</td>
<td>203</td>
<td>71</td>
</tr>
<tr>
<td>Least agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>40</td>
<td>28</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>80</td>
<td>57</td>
</tr>
<tr>
<td>Maternal &amp; perinatal</td>
<td>16</td>
<td>12</td>
</tr>
</tbody>
</table>

Unadjusted Agreement, Prevention Stage, Funding Source, and Year

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>~Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevention Stage (p=0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>244</td>
<td>120</td>
</tr>
<tr>
<td>Secondary</td>
<td>356</td>
<td>177</td>
</tr>
<tr>
<td>Tertiary</td>
<td>738</td>
<td>288</td>
</tr>
<tr>
<td>Funding Source (p&lt;0.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>295</td>
<td>77</td>
</tr>
<tr>
<td>Other</td>
<td>1043</td>
<td>508</td>
</tr>
<tr>
<td>Year (p=0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre 2000</td>
<td>403</td>
<td>173</td>
</tr>
<tr>
<td>Post 1999</td>
<td>935</td>
<td>412</td>
</tr>
</tbody>
</table>

Odds Ratios from Agreement Logits *

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>0.546</td>
<td>0.534</td>
<td>0.422</td>
</tr>
<tr>
<td>Disease category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>0.650</td>
<td>0.622</td>
<td>0.693</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>0.581</td>
<td>0.539</td>
<td>0.473</td>
</tr>
<tr>
<td>Neuropsychiatric</td>
<td>0.791</td>
<td>0.705</td>
<td>0.576</td>
</tr>
<tr>
<td>Prevention stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>1.818</td>
<td>1.835</td>
<td>2.082</td>
</tr>
<tr>
<td>Tertiary</td>
<td>1.753</td>
<td>1.872</td>
<td>1.900</td>
</tr>
</tbody>
</table>

$50,000 Sensitivity Analysis

- Pharmaceutical studies and musculoskeletal and cancer studies: Remain significant for lower agreement
- Screening: Significantly less agreement in all 3 models
- Infectious and parasitic diseases: Significantly greater agreement
- Neuropsychiatric diseases and prevention stage generally no longer significant

Increased Disagreement for Pharmaceuticals

- Not sure why, but
  - Reasonable spread of pharmaceutical studies across disease areas
    - Reduces likelihood that result due to pharmaceutical studies being clumped in a few disease areas for which WTP might be substantially higher than the $100k threshold we adopted
  - Pharmaceutical studies tend to concentrate in tertiary prevention
    - Tertiary care has greater agreement, but pharmaceuticals have less agreement in this environment of greater agreement

* >1 = more agreement; <1 = more disagreement; robust standard errors clustered at the article level
† p < 0.1, ‡ p < 0.05, § p < 0.01
Limitations

- Are CEA analyses included in our study representative of clinical decisions for which CER analyses will be performed?
  - Publication bias?
- Agreement between CER and CEA measured by comparing incremental QALYs and incremental cost per QALY ratios
- Could not account for uncertainty
- Common WTP threshold for all therapies

Incremental QALYs vs Cost per QALY

- Does the 70% agreement we observed between QALY gains (CER) and cost per QALY ratios (CEA) translate to agreement for other outcomes such as biomarkers?
  - e.g., Simply knowing drug A reduces cholesterol more than drug B does not imply drug A’s resulting increase in QALYs makes it good or bad value
- Depends in part on whether studies in which cost-effectiveness has been reported are a representative sample of studies in which CER will be performed

Uncertainty

- CEA Registry does not report variability of the difference in costs or effects or of the cost-effectiveness ratio
- Addition of variability generally thought to increase agreement
  - Point estimates indicate disagreement, but one or both estimates not significant (no significant difference in effectiveness or CI for CER that includes WTP) and we cannot be confident of disagreement
- But can decrease agreement
  - Point estimates indicate agreement, but nonsignificance of one or both estimates reduces confidence of agreement

Common WTP Threshold

- Given US thresholds generally unknown, difficult to evaluate use of different WTP thresholds for different diseases
- Allowing different diseases to have different thresholds generally thought to increase agreement
  - e.g., if treatments for musculoskeletal or neuropsychiatric diseases or primary prevention have WTP thresholds >100,000
- But also can decrease agreement
  - Do some diseases have lower values of WTP?
  - Interaction with uncertainty?
  - Possible to have less certainty of value as WTP approaches ∞

“Correct” That PCORI Not Consider Costs?

- No evidence that adopting the more effective therapy saves health care $ 
  - In 72% of the studies in our sample, the more effective therapy was associated with higher costs 
  - 28% with lower costs probably overstates the likelihood of savings in health care $ 
    - A number of studies derived savings from non-health care $ (e.g., work loss)

“Correct” That PCORI Not Consider Costs? (2)

- Rationale: PCORI should “put the emphasis on clinical outcomes” and local public and private decision makers can develop economic evidence 
  - Can’t be efficient 
  - Quality of evidence will be mixed at best
- Does development of clinical but not economic evidence make controlling costs harder rather than easier?
  - “But PCORI reported its the most effective therapy...”
  - Future legislation?: “Insurers must cover the most effective therapy as determined by PCORI”
- Should PCORI collect economic data, but not use it in making its recommendations? 
  - Would increase efficiency and allow quality monitoring
Conclusions

• Had hoped to be able to develop measurable criteria that allowed us to confidently avoid some CEA so as to avoid CEA fatigue
• Did find that economics data are more likely to raise questions for CER studies of pharmaceuticals, musculoskeletal conditions, and neuropsychiatric conditions as well as for primary prevention
  – Don’t appear to be very strong results or rule out many cost-effectiveness analyses

Conclusions (2)

• Large amount of agreement between CER and CEA when QALYs are the outcome measure (i.e., part of the CEA calculation), but:
  – May not translate to studies that use some other CER metric
  – Unclear our findings imply we can avoid CEA for politically visible therapies which probably cause the greatest fatigue
• Still a noble aim, and more research is needed

Unadjusted Agreement, Publicness, Disease Burden, and Research Intensity

<table>
<thead>
<tr>
<th>Variable, Mean (SD)</th>
<th>Agree</th>
<th>~Agree</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Publicness”</td>
<td>0.34</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>Research Intensity</td>
<td>0.012</td>
<td>0.010</td>
<td>0.01</td>
</tr>
</tbody>
</table>