A Novel Method for Evaluating Value Assessment Frameworks

Tanya G.K. Bentley, PhD,1,2 Joshua T. Cohen, PhD,2,3 Elena B. Elkin, PhD,4 Julie Huynh, MD,5 Arnab Mukherjea, DrPH, MPH,2 Thanh H. Neville, MD,1,4,5 MSHS,6 Matthew Mel, MD,2,3 Ronda Copher, PhD,4 Russell L. Kruth, PhD,6 Ioana Popescu, MD, MPH,5,6 Jennifer M. Zambrano, DNP, CNS, RN1,5 Jackie Lee, BS1, Eunice Chang, PhD,5 Michael S. Broder, MD, MS1,5

1Partnership for Health Analytic Research, LLC, Beverly Hills, CA; 2Tufts Medical Center, Boston, MA; 3Memorial Sloan-Kettering Cancer Center, NY; 4Hematology Oncology of San Fernando Valley, Encino, CA; 5California State University, East Bay, Hayward, CA; 6David Geffen School of Medicine at UCLA, Department of Medicine, Los Angeles, CA; 7City of Hope National Medical Center, Duarte, CA; 8Essai Inc., Woodcliff Lake, NJ

BACKGROUND

• Various frameworks have been developed to assess the value of oncology drugs.
• Organizations who have developed frameworks include:
  – American Society of Clinical Oncology (ASCO)
  – European Society for Medical Oncology (ESMO)
  – Institute for Clinical and Economic Review (ICER)
  – National Comprehensive Cancer Network (NCCN)
• Despite their common goals, it is unclear whether the frameworks actually provide valid and reliable measurements of value and how to assess such validity and reliability in practice.

OBJECTIVE

• We developed a methodology for evaluating the validity and reliability of value assessment frameworks.

METHODS

Overview

• We calculated convergent validity, defined as the correlation among drug rankings across frameworks.
• Kendall's W coefficient of concordance for (Kendall's W) was chosen as the statistical measure.
  1. Calculated mean scores for each drug.
  2. Ranked mean scores of each of the 5 drugs within each framework from highest to lowest.
  3. Compared rankings among the frameworks.
• Kendall’s W ranges from 0 (no agreement) to 1 (complete agreement).
• P-values tested alternative hypothesis of complete agreement (W > 0) against null hypothesis.
• Means were re-scaled to 0-1,00 for easy comparisons.

RESULTS

Application

• We applied the method to 5 drugs for advanced non-small cell lung cancer.
• Each assessment produced a single numeric or ordinal outcome (in aggregate the ‘panelist scores’).
• Used along with NCCN’s published assessments (“published scores”) to evaluate convergent validity across 4 frameworks.

Panelists successfully completed all value assessments for 5 selected drugs.

• Results of application are shown in Figure 2 (validity) and in the Table (reliability).

Specifically:

• Raw scores are on different scales and cannot be compared.
• When re-ranked from 0 (worst) to 100 (best), score ranges varied among frameworks.
• ASCO and ESMO had wider ranges: 31 and 72 points, respectively.
• ICER and NCCN had much narrower ranges: 14 and 19 points, respectively.
• ASCO and ESMO: 16-47
• ESMO: 25-97
• ICER: 80-94
• NCCN: 75-94
• ASCO scores were the lowest, and NCCN scores were highest.
• Kendall’s W=0.703

Figure 2. Ranking of Re-scaled Scores of 5 Lung Cancer Drugs using 4 Frameworks: Overall and by Subdomain

1. Overall
2. Clinical Benefit
3. Toxicity
4. Quality of Life
5. Certainty

CONCLUSIONS

• This method is the first to allow quantitative analyses of value assessment frameworks’ validity and reliability.
• When applied to 5 oncology drugs, this method successfully allowed us to draw conclusions about the convergent validity and inter-rater reliability of 4 value frameworks.
• Overall, reliability was quite good.
• Reliability was better among oncologists and physicians for ASCO and ESMO, but not ICER.
• Individuals who want to conduct their own value assessments in oncology (rather than use a published value) should choose either ASCO or ESMO, because these two frameworks demonstrated high validity and reliability.
• ICC calculations were done assuming the 8 reviewers were organized.
• All reviewers had the same scores for each drug.
• Kendall’s W was chosen as the statistical measure.

Table. ICC (95% CI), Overall and by Panelist Type and Subdomain

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>ASCO</th>
<th>ESMO</th>
<th>ICER</th>
<th>NCCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>All reviewers</td>
<td>0.796</td>
<td>0.804</td>
<td>0.281</td>
<td>0.055</td>
<td>0.799</td>
</tr>
<tr>
<td>(n=8)</td>
<td>(0.517, 0.970)</td>
<td>(0.545, 0.973)</td>
<td></td>
<td>(0.055 - 0.799)</td>
<td></td>
</tr>
<tr>
<td>Oncologists vs. Non-oncologists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncologists</td>
<td>0.835</td>
<td>0.843</td>
<td>0.120</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>(n=4)</td>
<td>(0.526 - 0.979)</td>
<td>(0.520 - 0.980)</td>
<td>(0.000 - 0.759)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-oncologists</td>
<td>0.716</td>
<td>0.806</td>
<td>0.368</td>
<td>0.029</td>
<td>0.861</td>
</tr>
<tr>
<td>(n=4)</td>
<td>(0.331 - 0.959)</td>
<td>(0.477 - 0.974)</td>
<td>(0.029 - 0.861)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians vs. Non-physicians</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians (n=6)</td>
<td>0.855</td>
<td>0.793</td>
<td>0.228</td>
<td>0.222</td>
<td>n/a</td>
</tr>
<tr>
<td>(0.618 - 0.981)</td>
<td>(0.507 - 0.971)</td>
<td>(0.000 - 0.776)</td>
<td>(0.000 - 0.839)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-physicians</td>
<td>0.562</td>
<td>0.769</td>
<td>0.222</td>
<td>0.222</td>
<td>n/a</td>
</tr>
<tr>
<td>(n=2)</td>
<td>(0.938 - 0.983)</td>
<td>(0.973 - 0.973)</td>
<td>(0.000 - 0.839)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By Subdomain

• NCCN is not a distinct component of the framework.
• ICC and CI shown as measures of framework reliability.
• Negative ICC estimate was observed, suggesting that the true ICC is very low; therefore, ICC of zero was assumed.
• All reviewers had the same scores for each drug.

Panelists’ Survey Results

• Panelists’ mean time to complete each assessment:
  1. ASCO and ICER: ~30 minutes
  2. ESMO: 15 minutes
• Mean time to review literature for each drug for conducting assessments: 20-30 minutes
• ESMO instructions were the clearest.
• ASCO was rated most logically organized.
• No single frameworks emerged as:
  - Easiest to use
  - Having highest global panelist rating (e.g., comfort with using framework to assess treatment for a loved one).

Acknowledgements

Presented at ISPOR EU, October 29-November 2, 2016, Vienna, Austria
Research conducted by Partnership for Health Analytic Research, LLC.

This study was sponsored by Essai Inc.