Several organizations, including American Society of Clinical Oncology (ASCO), European Society for Medical Oncology (ESMO), Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, American Society of Hematology (ASH), American Society of Clinical Oncology, National Comprehensive Cancer Network (NCCN), have developed frameworks to assess the value of oncologic drugs. We previously developed a methodology for evaluating the validity and reliability of value assessments and applied it in a pilot study. This study aimed to evaluate our methodology’s applicability for assessing a broader array of drugs and frameworks.

METHODS

Overview

Our method is based on two primary outcomes:

1. Convergent validity: how correlated drug rankings are across frameworks.
2. Inter-rater reliability: a measure of how stable framework value estimates are across users.

We chose intra-class correlation coefficients (ICC) with 95% confidence intervals (CI) as the statistical measure. We calculated ICC separately for each framework, overall and by subdomain, assuming that the 8 reviewers represented a random sample from a larger population of reviewers.

Application

We applied the method to drug frameworks of 5 drugs for the 3 indications (total of 15 drugs):

- Advanced breast cancer
- Advanced non-small cell lung cancer
- Castration refractory prostate cancer

Eight panels assessed the drug: 4 oncologists, 2 non-oncologists, 2 health services researchers.

Each assessment produced a single numeric or ordinal score.

RESULTS

The 8 panelsists successfully completed a total of 480 assessments (4 frameworks * 8 panelsists * 15 drugs) resulting in the Table 2 (validity) and the Table 1 (reliability).

### Table 1. ICC AND 95% CI

<table>
<thead>
<tr>
<th>Drug</th>
<th>ASCO</th>
<th>ESMO</th>
<th>ICER</th>
<th>NCCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>0.807</td>
<td>0.926</td>
<td>0.917</td>
<td>0.913</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>0.766</td>
<td>0.914</td>
<td>0.912</td>
<td>0.911</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>0.785</td>
<td>0.876</td>
<td>0.829</td>
<td>0.830</td>
</tr>
</tbody>
</table>

Columns show drug frameworks and each framework is scored on a scale of 0-1 (completely agree). In each panel, Kendall’s W is shown as a measure of concordance across all frameworks and each panel’s pairwise comparison.

Specifically:

- Frameworks produced scores on different scales, so raw scores cannot be directly compared.
- When re-scaled from 0 (worst) to 100 (best), scale ranges varied across frameworks.
- NCCN spanned the narrowest range.
- ESMO scores spanned the broadest range.
- Convergent validity among frameworks was fair to excellent, suggesting that the clinical benefit subdomain and simplicity of trial design were well matched.
- Overall convergent validity was excellent (0.75) only when also excellent among all subdomains.
- Clinical benefit subdomain scores (even when convergence among toxicity scores was poor, <0.40)

For example:

- Clinical benefit concordance was poor (<0.40) or fair (<0.65) in 22 minutes, with fair overall concordance, despite good or excellent concordance among the toxicity and quality of life subdomains.
- Clinical benefit concordance was excellent (>0.75) among drugs with excellent-concordance, despite poor toxicity concordance.

Table 1. ICC AND 95% CI by Panel Type and Framework

### Table 2. Mean Panelsists’ Literature Review and Assessment Completion Times

<table>
<thead>
<tr>
<th>Framework</th>
<th>Literature review for each drug assessed</th>
<th>Completion of each assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCO</td>
<td>28 minutes</td>
<td>25 minutes</td>
</tr>
<tr>
<td>ESMO</td>
<td>22 minutes</td>
<td>14 minutes</td>
</tr>
<tr>
<td>ICER</td>
<td>25 minutes</td>
<td>21 minutes</td>
</tr>
<tr>
<td>NCCN</td>
<td>11 minutes</td>
<td>8 minutes</td>
</tr>
</tbody>
</table>

*Assessments were conducted last among all panels; no single framework had a different schedule: easiest to use; having global panelist rating (e.g., comfort with using framework to assess treatment for a loved one).

CONCLUSIONS

This method allows quantitative analyses of value assessment frameworks’ validity and reliability.

When applied to 15 oncologic drugs in 3 indications, this method successfully allowed us to draw conclusions about the convergent validity and inter-rater reliability of 4 value frameworks.

This framework demonstrated fair-to-excellent convergent validity, and appropriately focused on clinical validity.

- Overall concordance was strongly influenced by concordance among clinical efficacy scores.
- All frameworks except NCCN demonstrated good-to-excellent reliability.

When 2 frameworks produced similar clinical benefit scores, the overall scores were generally more concordant. Clinical benefit score primarily reflects efficacy, which is probably an important driver of clinical decision-making. Thus framework scores may reflect those made in clinical practice.

Assessments were found to be time consuming, so their usefulness in practice may be enhanced with the release of more committee-based assessments from framework developers.

Use of this methodology to determine how drugs may be valued by different frameworks will facilitate provider, enhancer, and patient decision-making.

Copher Thanh H. Neville, MD, MSHS Massachusetts Institute of Technology, Cambridge, MA, USA, 3 Memorial Sloan-Kettering Cancer Center, New York, CA, USA, 4 Hematology-Oncology Medical Group of San Fernando Valley, Encino, CA, USA, 5 Stanford University School of Medicine, Los Angeles, CA, USA, 6 City of Hope National Medical Center, Duarte, CA, USA, 7City of Hope, Woodcliff Lake, NJ, USA, 8Eisai Inc., Woodcliff Lake, NJ, USA.