Validity and Reliability of Four Value Frameworks for Cancer Drugs

Tanya G.K. Bentley, PhD[1], Joshua T. Cohen, PhD[2], Elena B. Elkin, PhD[3], Julie Huynh, MD[4], Arnab Mukherjea, DrPH, MPH[5], Thanh H. Neville, MD, MSHS[6], Matthew Mei, MD[7], Ronda Copher, PhD[8], Russell Knoth, PhD[8], Ioana Popescu, MD, MPH[6], Jackie Lee, BS[1], Jenelle M. Zambrano DNP, CNS, RN[1], Michael S. Broder, MD, MSHS[1]

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

BACKGROUND

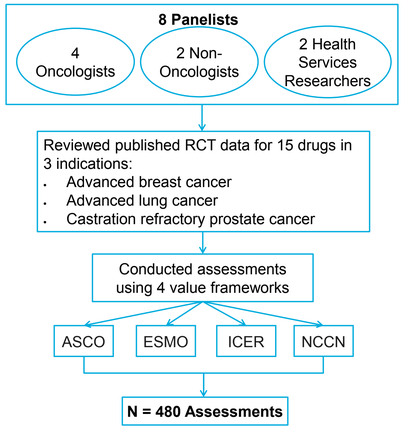
- The American Society of Clinical Oncology (ASCO), European Society for Medical Oncology (ESMO), Institute for Clinical and Economic Review (ICER), and National Comprehensive Cancer Network (NCCN) have developed frameworks to assess the value of oncology drugs.
- The extent to which value assessment frameworks provide valid and reliable measurements of a drug's or regimen's value remains unclear.
- We expanded on a pilot study to evaluate the convergent validity and inter-rater reliability of 4 value assessment frameworks.

METHODS

Overview

- A panel of 8 clinicians and health services researchers assessed the value of 15 new drugs for advanced lung and breast cancers and castration refractory prostate cancer using the ASCO, ESMO, ICER, and NCCN value frameworks. **[Figure 1]**
- They were provided with instructions and published phase III RCT clinical data to complete the assessments.
- For each assessment, panelists assigned each drug a numeric or letter score.
- The 8 panelists completed a total of 480 assessments (4 frameworks * 8 panelists * 15 drugs).
- After completing their assessments, panelists rated the frameworks and provided comments on their experiences using the frameworks.

Figure 1. Study Design



Analysis

Mean Scores

- We estimated mean scores and standard deviations for each drug and framework, overall and by subdomain, and re-scaled the means to 0-100 for descriptive comparisons.

Validity

- We used Kendall's W coefficient to measure **convergent validity**: extent to which frameworks produced similar evaluations for same list of drugs.
- For the 3 cancer types, we calculated W:
  - Overall, across the 4 frameworks
  - Pairwise, within each pair of frameworks
  - By subdomain: clinical benefit, toxicity, quality of life, certainty
  - By panelist characteristics: oncologists vs. non-oncologists; physicians vs. non-physicians
  - By individual panelists

Reliability

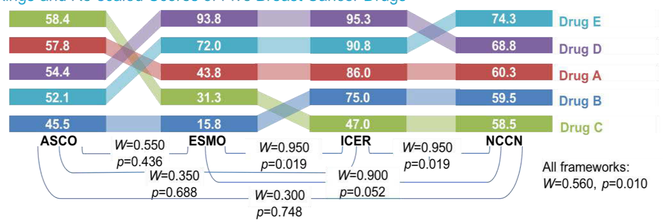
- We assumed the 8 panelists represented a random sample from a larger population of framework users.
- Intraclass correlation coefficients (ICC) used to measure framework **inter-rater reliability**: extent to which independent panelists arrived at same score for each assessment.
- We calculated ICC across cancers using 95% CI and conducted sensitivity analyses for:
  - Each panelist removed one at a time
  - By panelist characteristics: oncologists vs. non-oncologists; physicians vs. non-physicians
  - By subdomain: clinical benefit, toxicity, quality of life, and certainty

RESULTS

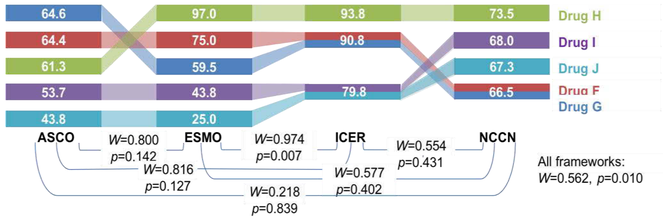
- Results appear in Figure 2 (drug rankings), Table 1 (convergent validity), and Table 2 (reliability).
- The frameworks demonstrated fair-to-excellent convergent validity.
- Clinical efficacy had the greatest influence on convergence.
- The ASCO, ESMO, and ICER frameworks demonstrated good-to-fair reliability. The reliability of the NCCN framework was poor.

Figure 2. Overall Kendall's W Among 4 Frameworks: Rankings and Re-Scaled Scores of 15 Cancer Drugs in 3 Indications

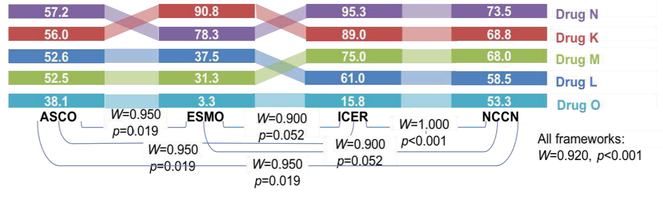
1. Rankings and Re-scaled Scores of Five Breast Cancer Drugs



2. Rankings and Re-scaled Scores of Five Lung Cancer Drugs



3. Rankings and Re-scaled Scores of Five Prostate Cancer Drugs



- Each value assessment took panelists ~25 minutes using the ASCO framework, 14 minutes with ESMO, 21 minutes with ICER, and 8 minutes with NCCN (the framework for which all assessments were done last).
- Mean time to review literature was consistent among cancer types, ranging from 11-28 minutes.
- Panelists generally agreed the frameworks were logically organized and easy to use.
- Panelists neither agreed nor disagreed on whether they would be comfortable using the frameworks for assessing the value of cancer treatment for a loved one.

Table 1. Kendall's W subdomain ranges

	High *(p)*		Low *(p)*	
Breast	Certainty (ICER, NCCN)	0.908 (0.046)	Clinical benefit (ASCO, ESMO, NCCN)	0.345 (0.436)
Lung	Toxicity (ASCO, ESMO, NCCN)	0.944 (<0.001)	Certainty (ICER, NCCN)	0.230 (0.827)
Prostate	Quality of life (ASCO, ESMO)	0.986 (0.003)	Toxicity (ASCO, ESMO, NCCN)	0.200 (0.711)

Table 2. ICC (95% CI) by Panelist Type and Subdomain[a]

	ASCO	ESMO	ICER	NCCN
All panelists (n=8)	0.800 (0.660 - 0.913)	0.818 (0.686 - 0.921)	0.652 (0.466 - 0.834)	0.153 (0.045 - 0.371)
Oncologists vs. non-oncologists				
Oncologists (n=4)	0.807 (0.638 - 0.920)	0.842 (0.699 - 0.936)	0.769 (0.582 - 0.903)	0.210 (0.020 - 0.501)
Other (n=4)	0.786 (0.605 - 0.911)	0.816 (0.655 - 0.924)	0.603 (0.353 - 0.817)	0.156 (0[b] - 0.427)
Physicians vs. non-physicians				
Physicians (n=6)	0.825 (0.686 - 0.926)	0.831 (0.698 - 0.929)	0.641 (0.439 - 0.830)	0.156 (0.031 - 0.395)
Other (n=2)	0.740 (0.375 - 0.905)	0.691 (0.302 - 0.884)	0.482 (0.023 - 0.784)	0.198 (0[b] - 0.597)
By subdomain				
Clinical Benefit	0.829 (0.704 - 0.927)	0.809 (0.673 - 0.917)	n/a	0.149 (0.041 - 0.368)
Toxicity	0.755 (0.592 - 0.891)	0.597 (0.406 - 0.800)	n/a	0.194 (0.067 - 0.432)
Quality of Life	0.671 (0.490 - 0.844)	0.818 (0.686 - 0.921)	n/a	n/a
Certainty	n/a	n/a	0.062 (0[b] - 0.247)	0.022 (0[b] - 0.129)

n/a: Subdomain is not a distinct component of the framework.

[a] ICC and CI shown as measures of framework reliability.

[b] Negative ICC estimate was observed, which suggested that the true ICC is very low; therefore, ICC of zero was assumed.

CONCLUSIONS

- This analysis represents one of the first quantitative assessments of the convergent validity and inter-rater reliability of the ASCO, ESMO, ICER, and NCCN value assessment frameworks.
- The different approaches built into these frameworks suggest that stakeholders have yet to agree on exactly how to define value.
- Perhaps because of the lack of clear conceptual agreement, convergent validity among the frameworks was only fair for 2 of the diseases studied, although it was excellent for the third. Replication with a larger sample using multiple conditions might help further explain this divergence.
- Concordance increased with clinical benefit concordance and simplicity of drug trial data.
- Inter-rater reliability was good or excellent for all frameworks except NCCN's, whose simpler approach made it user-friendly but more susceptible to small differences between users.
- Mean scores produced by a committee will be more reliable than those produced by an individual.
- It remains to be seen if any framework will get incorporated into actual practice and bring us closer to using value-based treatment decisions to improve patient care and outcomes. Future research should evaluate oncologists' use and perceptions of these frameworks.

Corresponding author: Tanya G.K. Bentley, PhD; tbentley@pharllc.com

Presented at 2017 ASCO Annual Meeting; June 2-6, 2017; Chicago, Illinois, USA

This study was sponsored by Eisai, Inc.