VALUE IN HEALTH 20 (2017) 200–205

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Validity and Reliability of Value Assessment Frameworks for New 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, USA;* [2]*Tufts Medical Center, Boston, MA, USA;* [3]*Memorial Sloan Kettering Cancer Center, New York, NY, USA;* [4]*Hematology Oncology of San Fernando Valley, Encino, CA, USA;* [5]*Health Sciences Program, California State University, East Bay, Hayward, CA, USA;* [6]*Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA;* [7]*City of Hope National Medical Center, Duarte, CA, USA;* [8]*Eisai Inc., Woodcliff Lake, NJ, USA*

A B S T R A C T

Background: Several organizations have developed frameworks to systematically assess the value of new drugs. These organizations include the American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO), the Institute for Clinical and Economic Review (ICER), and the National Comprehensive Cancer Network (NCCN). **Objectives:** To understand the extent to which these four tools can facilitate value-based treatment decisions in oncology. **Methods:** In this pilot study, eight panelists conducted value assessments of five advanced lung cancer drugs using the ASCO, ESMO, and ICER frameworks. The panelists received instructions and published clinical data required to complete the assessments. Published NCCN framework scores were abstracted. The Kendall's W coefficient was used to measure convergent validity among the four frameworks. Intraclass correlation coefficients were used to measure inter-rater reliability among the ASCO, ESMO, and ICER frameworks. Sensitivity analyses were conducted. **Results:** Drugs were ranked similarly by the four frameworks, with Kendall's W of 0.703 ($P = 0.006$) across all the four frameworks. Pairwise, Kendall's W was the highest for ESMO-ICER ($W = 0.974$; $P = 0.007$) and ASCO-NCCN ($W = 0.944$; $P = 0.022$) and the lowest for ICER-NCCN ($W = 0.647$; $P = 0.315$) and ESMO-NCCN ($W = 0.611$; $P = 0.360$). Intraclass correlation coefficients (confidence interval [CI]) for the ASCO, ESMO, and ICER frameworks were 0.786 (95% CI 0.517–0.970), 0.804 (95% CI 0.545–0.973), and 0.281 (95% CI 0.055–0.799), respectively. When scores were rescaled to 0 to 100, the ICER framework provided the narrowest band of scores. **Conclusions:** The ASCO, ESMO, ICER, and NCCN frameworks demonstrated convergent validity, despite differences in conceptual approaches used. The ASCO inter-rater reliability was high, although potentially at the cost of user burden. The ICER inter-rater reliability was poor, possibly because of its failure to distinguish differential value among the sample of drugs tested. Refinements of all frameworks should continue on the basis of further testing and stakeholder feedback.

Keywords: convergent validity, inter-rater reliability, oncology, value frameworks.

Introduction

Several organizations have developed frameworks that can be used to systematically assess the value of oncology drugs. These organizations include the American Society of Clinical Oncology (ASCO), the European Society for Medical Oncology (ESMO), the Institute for Clinical and Economic Review (ICER), and the National Comprehensive Cancer Network (NCCN).

Value frameworks aim to help physicians and patients weigh treatment options with clinical decision making by assessing the value of new therapies. In addition, they propose to help public and private payers such as the Centers for Medicare & Medicaid Services and managed care organizations make value-based pricing and resource allocation decisions [1–10]. Such frameworks conceptually define "value" generally on the basis of treatment benefit and cost, although they use different components (e.g., efficacy, toxicity, and quality of life) and analytic approaches. The framework developers intend to apply frameworks across an array of drugs and provide value assessments, either in published form or in software tools, that can facilitate modifications on the basis of user preferences [2,4–10].

Despite their common goals, it is unclear whether the frameworks actually provide valid and reliable measurements of value. To date, published assessments of value frameworks have been primarily conceptual [11–18], including qualitative descriptions of differences in framework subdomains, purposes, audiences, or

* *Address correspondence to:* Tanya G. K. Bentley, Partnership for Health Analytic Research, LLC, 280 South Beverly Drive, Suite 404, Beverly Hills, CA 90212.

E-mail: tbentley@pharllc.com.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2016.12.011

editorial commentaries on how value frameworks may help achieve the goal of providing value-based care. In this pilot study, we sought to provide the first assessment of their convergent validity and inter-rater reliability in practice.

Methods

Overview

We evaluated the convergent validity and inter-rater reliability of the ASCO, ESMO, ICER, and NCCN value frameworks as applied to five systemic therapies (drugs) for advanced lung cancer. This pilot study is the first stage of a larger study involving multiple cancer types. A panel of clinicians and health services researchers assessed five lung cancer drugs using the forms and instructions included in the ASCO, ESMO, and ICER value frameworks. Each assessment produced a single numeric or categorical outcome (in aggregate the "panel scores") that was used along with NCCN's published assessments ("published scores") to evaluate convergent validity—the correlation among rankings—across the four frameworks. The ASCO, ESMO, and ICER frameworks' inter-rater reliability—the degree to which they provide stable and consistent results—was assessed on the basis of panel scores. Sensitivity analyses evaluated the extent to which framework subdomains (e.g., clinical efficacy, toxicity, and quality of life) and panelist training impacted validity and reliability outcomes.

Panelists

Eight panelists were selected to represent a range of potential value framework users, including four oncologists, two non-oncologist physicians (one general internist and one pulmonary/critical care physician), and two nonphysician researchers (one PhD and one DrPH) with experience in oncology health services research.

Drugs

We identified more prevalent and costly cancers [19] and from these we selected those with drugs for which published value scores were available. We considered drugs representing a range of indications (curative and palliative), malignancies (solid and hematologic), and mechanisms (cytotoxic, biologic, and immunologic). After expert review, we developed a final list of 12 cancers and 90 drugs. The present report focuses on our examination of five drugs for a single indication: advanced lung cancer. The study sponsor played no role in determining the included cancers or drugs.

Assessments

The panelists applied each of the three frameworks to each of the five lung cancer drugs, yielding a total of 120 scores. They were provided efficacy and safety data from phase III randomized controlled trials with which to conduct each value assessment. Drug-specific quality-of-life data from the randomized controlled trials were included when available. In cases in which there were published scores with cited literature, we used the same literature. A basic set of instructions describing how to complete assessments using each framework and how to incorporate different types of data (e.g., overall vs. progression-free survival) was provided to the panelists. To simulate real-world assessment conditions, if panelists made arithmetic errors in calculating scores, we did not make corrections. After completing their value assessments for each drug and framework (15 assessments per panelist), the panelists were given a survey to rate the different frameworks and provide comments regarding their experiences.

Each framework produces scores on different scales. Most frameworks produce scores that can be directly ranked, as required in our analysis. The ASCO framework produces "net health benefit" scores ranging from −20 (worst) to +180 (best). The ASCO score is calculated on the basis of the drug's clinical efficacy, toxicity, effects on long-term survival, palliation, quality of life, and treatment-free interval. The ESMO framework produces scores ranging from 1 (worst) to 5 (best) on the basis of efficacy, toxicity, and quality of life. Unlike the other frameworks, the ICER framework does not produce scores that can be ranked. Instead, it comprises multiple components, including comparative clinical effectiveness, cost-effectiveness, and budget impact, each of which requires specific methodology and in-depth analysis. To produce ICER scores that could be ranked, we used ICER's comparative clinical effectiveness component—the evidence rating matrix. This online tool reports final grades from I (worst) to A (best) on the basis of comparative net benefits and the level of certainty associated with these benefits. For our analysis, these grades were converted to a numerical scale from 0 (worst) to 4 (best). The NCCN framework produces scores from 1 (worst) to 5 (best) for each of the four health benefit measures: efficacy, safety, quality of evidence, and consistency of evidence. Affordability was excluded.

Analysis

Mean scores and SDs were estimated for each drug and framework, overall and by subdomain. Means were also rescaled to 0 to 100 for descriptive comparisons. Convergent validity and inter-rater reliability were the primary outcomes. Although multiple types of validity exist, for this evaluation we measured convergent validity: the extent to which each framework produced similar evaluations for the same list of drugs. Convergent validity was evaluated using the Kendalll's coefficient of concordance for ranks (Kendall's W). Kendall's W measures the agreement of ranked items and was calculated on the basis of comparing ranked mean drug scores (i.e., from 1 to 5 to represent best to worst drug scores) among the four frameworks. Kendall's W is defined as follows:

$$R = \sum_{i=1}^{k} \left(R_i - \overline{R}\right)^2,$$

$$W = \frac{12R}{m^2(k^2-k)},$$

assuming m panelists assessed k drugs rank-ordered from 1 to k, and R_i is the total rank for drug i, $\overline{R}$ is the mean of R_i's, and R is thus the sum of squared deviations. Kendall's W ranges from 0 (no agreement) to 1 (complete agreement). P values were reported to test the alternative hypothesis of complete agreement ($W > 0$) against the null hypothesis (no agreement). W may be interpreted using a scale similar to that for intraclass correlation coefficients (ICCs) (described hereafter).

Kendall's W was calculated as follows: overall (across the four frameworks); within each pair of frameworks; for framework subdomains of clinical benefit, toxicity, quality of life, and certainty; and for each individual panelist.

Inter-rater reliability was assessed for the ASCO, ESMO, and ICER frameworks using ICCs and 95% confidence intervals (CIs). ICCs measured the extent to which the value assessments varied among panelists within each framework and represented the reproducibility of the assessments by different panelists. Panelist scores for each framework were used to calculate the ICC, defined as follows:

$$\mathrm{ICC} = \frac{\mathbf{var}(\beta)}{\mathbf{var}(\alpha)+\mathbf{var}(\beta)+\mathbf{var}(\varepsilon)},$$

where var(β) is the variability due to differences in the drugs, var (α) is the variability due to differences in the panelists, and var(ε) is the variability due to differences in the drugs and panelists. ICCs range from 0 to 1, where values less than 0.40 represent poor reliability; from 0.40 to 0.59, fair reliability; from 0.60 to 0.74, good reliability; and 0.75 and higher, excellent reliability [20].

ICC calculations assumed that the eight panelists represented a random sample from a larger population of framework users. Each panelist evaluated the same drugs with three frameworks. In sensitivity analyses, we calculated ICCs with each panelist removed one at a time. ICCs were also calculated between oncologist versus nononcologist as well as physician versus nonphysician panelists, and for the following subdomains relevant in each framework: clinical benefit, toxicity, quality of life, and certainty.

Data were collected using electronic forms, exported into Excel, and analyzed using SAS® version 9.4 (SAS Institute, Cary, NC). All tests were two-sided with a significance level of 0.05.

Results

Overall and subdomain mean scores and SDs for each of the five drugs using the ASCO, ESMO, ICER, and NCCN frameworks are presented in Table 1. Drugs were ranked similarly by the frameworks, with Kendall's W of 0.703 ($P = 0.006$) across all four frameworks (Fig. 1, panel 1). Pairwise, Kendall's W was the highest for ESMO-ICER ($W = 0.974$; $P = 0.007$) and ASCO-NCCN ($W = 0.944$; $P = 0.022$) and the lowest for ICER-NCCN ($W = 0.647$; $P = 0.315$) and ESMO-NCCN ($W = 0.611$; $P = 0.360$). When ranking drugs on the basis of distinct framework subdomains (Fig. 1, panels 2–5), Kendall's W was 0.715 ($P = 0.037$) for clinical benefit (ASCO, ESMO, and NCCN); 0.885 ($P = 0.064$) for quality of life (ASCO and ESMO); 0.633 ($P = 0.081$) for toxicity (ASCO, ESMO, and NCCN); and 0.348 ($P = 0.690$) for certainty (ICER and NCCN). Considering panelist scores only, Kendall's W increased to 0.816 ($P = 0.009$); W remained higher than 0.700 and P less than 0.050 when assessed one panelist at a time for all but two panelists, for whom W was 0.576 ($P = 0.126$) and 0.487 ($P = 0.221$). When rescaled from 0 (worst) to 100 (best), ASCO scores ranged from 16 to 47, ESMO scores from 25 to 97, ICER scores from 80 to 94, and NCCN scores from 75 to 94.

ICCs for the panelists' assessments using the ASCO, ESMO, and ICER frameworks were 0.786 (95% CI 0.517–0.970), 0.804 (95% CI 0.545–0.973), and 0.281 (95% CI 0.055–0.799), respectively (Table 2). For both the ASCO and ESMO frameworks, the ICC was higher for oncologists than for nononcologists (0.835 vs. 0.716 for ASCO; 0.843 vs. 0.806 for ESMO) and for physicians than for nonphysicians (0.855 vs. 0.562 for ASCO; 0.793 vs. 0.769 for ESMO). For the ICER framework, the ICC was lower for oncologists than for nononcologists (0.120 vs. 0.368) and differed by 0.006 between physicians and nonphysicians. When panelists were removed one at a time, the range of ICCs was 0.092 for ASCO, 0.048 for ESMO, and 0.147 for ICER. When considering framework subdomains, the ICC for the ASCO framework was lower for clinical benefit (0.692; 95% CI 0.383–0.952) and quality of life (0.681; 95% CI 0.372–0.950) and higher for toxicity (0.825; 95% CI 0.584–0.976). The opposite was true for the ESMO framework, with a higher ICC for clinical benefit (0.857; 95% CI 0.643–0.981) and quality of life (1.000; CI not applicable because all panelists had same scores for each drug) and a lower ICC for toxicity (0.468; 95% CI 0.172–0.890). ICER's ICC for the certainty subdomain was 0.006 (95% CI 0.000–0.511; Table 2).

Of the three drugs for which there were published ESMO scores [2], the mean panelist score was the same (4.00) for one drug and differed for the other two, with scores of 4.00 versus 3.38 and 4.00 versus 2.00 for published scores versus panelist scores, respectively.

Each value assessment took panelists approximately 30 minutes using the ASCO and ICER frameworks and 15 minutes using the ESMO framework. The mean time needed to review the literature (up to two articles) for each drug ranged from 20 to 30 minutes, excluding panelists' first drugs ever assessed. When asked about their experiences using the ASCO, ESMO, and ICER frameworks on a scale of 1 (strongly agree) to 5 (strongly disagree), panelists agreed that the instructions from the ESMO framework were most clearly written (mean: ASCO, 2.4; ESMO, 1.5; ICER, 2.6) and the ASCO framework was the most logically organized (mean: ASCO, 1.5; ESMO, 2.1; ICER, 2.4). Panelists neither agreed nor disagreed about whether the frameworks were easy to use (mean: ASCO, 2.5; ESMO, 2.3; ICER, 2.8) or whether they would be comfortable using the frameworks for assessing the value of cancer treatment for a loved one (mean: ASCO, 2.6; ESMO, 2.9; ICER, 3.0).

Discussion

Summarizing the value of oncology drugs into a single metric is a major challenge. In this pilot study, we evaluated four frameworks—those developed by ASCO, ESMO, ICER, and NCCN—of which each offers different approaches to overcoming this challenge. No two frameworks use the same subdomains, formulas, or scoring scales. Despite this, they ranked the drugs similarly, providing preliminary evidence of convergent validity and suggesting that they are all measuring a similar concept. The divergent approaches taken by each framework manifest themselves in their inter-rater reliability: there appears to be a trade-off between defining benefit using conceptually familiar categories on one hand and achieving high reliability on the other. The ICER framework uses broad categories (e.g., "substantial" benefit), but fails to clearly distinguish among the drugs tested. In contrast, the ASCO framework, for example, defines benefit using hazard ratios, allowing it to distinguish between the values of different, but similar, drugs. When the differences among the scores are small, differences between panelists are magnified, reducing inter-rater reliability.

Neither the NCCN ratings nor the ICER framework was able to distinguish clearly between several of the drugs included in this study, each grouping the five drugs into just three scores. On a 0 to 100 scale, the ICER framework provided the narrowest band of ratings. This finding may be due in part to the ICER's intuitive approach. To conduct assessments using this framework, users summarize a drug's benefits and risks in their own words and use this information to rate net benefit using broad and conceptually familiar, yet poorly defined, categories. For example, across all panelists, both "small/incremental" and "substantial" were selected at least once for every drug in the ICER value assessments. Users of the ICER framework also select a "conceptual confidence interval" to represent certainty around a drug's net benefit. The CI categories are intentionally subjective, and as such, inter-rater reliability for this certainty subdomain was also poor. Six of eight panelists in our study reported that the ICER components were too subjective to be useful. We could not subject the NCCN scores to inter-rater reliability testing, because we found no published details (i.e., beyond the category definitions) regarding the specific process for applying this framework. If they were replicated, one might expect a similar finding as with the ICER framework, because the range of value scores was only slightly less narrow (ranges of 19 vs. 14 points on a 0–100 scale).

The developers of the ICER framework might argue that the use of broad, conceptual categories is a "feature, not a bug," given that the goal of the ICER framework differs from the oncology-

Table 1 – Framework (mean ± SD) scores for five drugs, overall and by subdomain.

Drug	ASCO (N = 8)	ESMO (N = 8)	ICER (N = 8)	NCCN* (N = 1)
Overall				
A	67.58 ± 22.61	4.00 ± 0.00	3.63 ± 0.35	4.50
B	73.85 ± 6.23	3.38 ± 0.74	3.63 ± 0.23	4.75
C	63.81 ± 11.27	4.88 ± 0.35	3.75 ± 0.27	4.00
D	40.95 ± 10.65	2.75 ± 0.46	3.19 ± 0.53	4.00
E	11.49 ± 8.77	2.00 ± 0.76	3.19 ± 0.46	4.00
Clinical benefit				
A	39.88 ± 16.57	3.00 ± 0.00	–	4.00
B	50.35 ± 0.14	3.00 ± 0.00	–	5.00
C	41.00 ± 0.00	4.00 ± 0.00	–	4.00
D	29.00 ± 0.00	2.00 ± 0.00	–	4.00
E	21.00 ± 0.00	2.00 ± 0.76	–	4.00
Toxicity				
A	0.71 ± 2.56	0.13 ± 0.35	–	4.00
B	3.00 ± 3.79	0.50 ± 0.53	–	4.00
C	13.31 ± 4.20	0.88 ± 0.35	–	4.00
D	2.45 ± 7.60	0.75 ± 0.46	–	4.00
E	−16.52 ± 4.85	0.00 ± 0.00	–	4.00
Quality of life				
A	8.75 ± 3.54	1.00 ± 0.00	–	–
B	6.25 ± 5.18	0.00 ± 0.00	–	–
C	0.00 ± 0.00	0.00 ± 0.00	–	–
D	0.00 ± 0.00	0.00 ± 0.00	–	–
E	0.00 ± 0.00	0.00 ± 0.00	–	–
Certainty				
A	–	–	1.50 ± 0.53	5.00
B	–	–	1.75 ± 0.46	5.00
C	–	–	1.50 ± 0.53	4.00
D	–	–	1.75 ± 0.71	4.00
E	–	–	2.00 ± 0.53	4.00

ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; ICER, Institute for Clinical and Economic Review; NCCN, National Comprehensive Cancer Network.

* Published scores were used for the NCCN framework; panelist scores were used for the other frameworks.

specific missions of the ASCO, ESMO, and NCCN framework developers. The ICER framework seeks to provide a transparent and explicit method for assessing drugs across all therapeutic areas, a startlingly difficult task. If achieving such a goal is to be possible, it would seem to require the use of widely applicable categories with which to assess benefit. The ASCO framework, however, includes evaluations of a drug's impact on long-term survival, palliation, and treatment-free intervals, factors clearly developed keeping in mind patients with cancer. Both the ASCO and ESMO clinical benefit subdomains are measured on the basis of oncology-specific thresholds for overall and progression-free survival outcomes. The NCCN efficacy and safety category definitions also appear oncology-specific.

Better inter-rater reliability may come at the expense of increased cognitive burden for users. Our analysis found that the ASCO framework appears to have excellent inter-rater reliability. We used the revised 2016 version of this framework for this pilot study [4]. A previous version [3] was used in our study's earlier phase, and inter-rater reliability improved with the revision. Nevertheless, panelists reported that completing assessments using the revised ASCO framework was highly burdensome, more so than in the previous version. This was especially true for the toxicity subdomain, which was modified to incorporate more complex criteria and calculations. The optimal balance between inter-rater reliability and usability is not known.

ASCO, ESMO, ICER, and NCCN are science-focused entities engaged in a thoughtful process to improve health care. With the exception of the NCCN framework developers, each of the other framework developers published transparent descriptions of their development processes, which included multiple stakeholders. Evaluating the frameworks' construct validity—the extent to which they actually measure the latent variable "value"—is challenging, because multiple definitions of value have been offered [16,21–24]. The Institute of Medicine defined the six elements of value as effectiveness, safety, patient-centeredness, timeliness, efficiency, and equity [18]. All four frameworks evaluated here address the first two elements of effectiveness and safety. Patient-centeredness appears to have not yet been fully achieved. The concept is defined by the National Health Council as not only quality of life or patient-reported outcomes, but also as patient engagement throughout framework development and value assessment [25]. For these frameworks to achieve true patient-centeredness, comprehensive patient engagement is needed. To an extent, the framework developers acknowledge that their frameworks do not comprehensively measure value, noting, for example, that elements relevant to patients' or providers' individual value systems may not all be included [2,4,6,9]. Subgroup-specific value measurements are provided to the extent that such data are available, often not the case in the oncology setting. Because results may vary for individual patients, these frameworks should be considered one among many tools for real-world treatment decision making.

We note that the focus of this analysis has been on whether value assessments made using the frameworks we looked at are reproducible. Reproducibility is a prerequisite to the frameworks

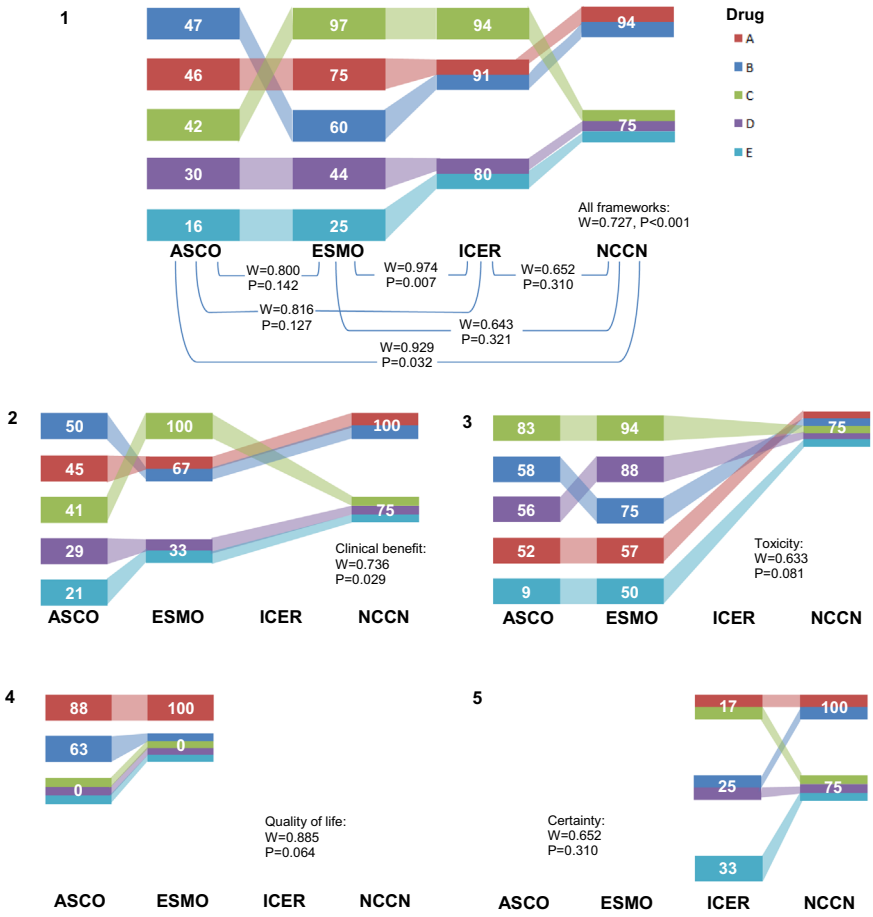

Fig. 1 – Columns represent each framework. Re-scaled mean scores (range: 0-100) are shown in each rectangle. Multi-colored rectangles represent tied scores. In panel 1, Kendall's W is shown as a measure of concordance across all frameworks and each pairwise comparison. In panels 2-5 it is shown for each subdomain. Subdomain scores not shown are not distinct components of the framework. In the certainty subdomain for ICER (panel 5), lower scores represent higher rankings. ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; ICER, Institute for Clinical and Economic Review; NCCN, National Comprehensive Cancer Network.

Table 2 – ICCs (95% CI) by panelist type and subdomain*.

Panelist type/subdomain	ASCO	ESMO	ICER
All panelists (n = 8)	0.786 (0.517–0.970)	0.804 (0.545–0.973)	0.281 (0.055–0.799)
Oncologists vs. nononcologists			
Oncologists (n = 4)	0.835 (0.526–0.979)	0.843 (0.520–0.980)	0.120 (0†–0.759)
Other (n = 4)	0.716 (0.331–0.959)	0.806 (0.477–0.974)	0.368 (0.029–0.861)
Physicians vs. nonphysicians			
Physicians (n = 6)	0.855 (0.618–0.981)	0.793 (0.507–0.971)	0.228 (0†–0.776)
Other (n = 2)	0.562 (0†–0.938)	0.769 (0†–0.973)	0.222 (0†–0.839)
By subdomain			
Certainty			0.006 (0†–0.511)
Clinical benefit	0.692 (0.383–0.952)	0.857 (0.643–0.981)	
Quality of life	0.681 (0.372–0.950)	1.000 (NA‡)	
Toxicity	0.825 (0.584–0.976)	0.468 (0.172–0.890)	

ASCO, American Society of Clinical Oncology; CI, confidence interval; ESMO, European Society for Medical Oncology; ICCs, intraclass correlation coefficients; ICER, Institute for Clinical and Economic Review; NA, not applicable.

* ICC and CI shown as measure of framework inter-rater reliability.

† Negative ICC estimate was observed, which suggested that the true ICC is very low; therefore, ICC of 0 was assumed [26].

‡ All panelists had the same scores for each drug.

having any value, but it is only one component of an overall assessment of the frameworks' contribution to value-based decision making. Although it is important to note that the analyses presented here are based on limited sample size, this pilot study is an early step in evaluating these frameworks. The findings will be further elucidated when the full analyses with more cancer types and drugs are completed. Ultimately, the frameworks must be judged by how they influence decisions made by clinicians and patients. Evidence to address that question requires the conduct of further studies.

We encourage further development of the frameworks, including:

1. continuing to solicit feedback from key stakeholders (e.g., patients, physicians, government agencies, manufacturers, and payers);
2. working to improve inter-rater reliability by testing and modification of frameworks;
3. working with agencies such as the US Food and Drug Administration, the European Medicines Agency, and others to increase the incorporation and reporting of quality-of-life end points in phase III trials;
4. to the extent not done already, disclosing the formulas used in their value assessments; and
5. including major stakeholders—patients in particular—in discussions.

Conclusions

Measuring the value of drugs in a quantifiable way has become of increasing interest. In this pilot study, the ASCO, ESMO, ICER, and NCCN frameworks demonstrated convergent validity, despite differences in conceptual approaches used. Inter-rater reliability for the ASCO framework appears high, although potentially at the cost of user burden. The ICER inter-rater reliability seems poor, which may be a result of its failure to clearly differentiate among the sample of drugs tested. Given the preliminary nature of these findings, the testing and refinement of all frameworks should continue, in particular considering the impact that framework-guided provider, payer, and policymaker decisions have on patients.

Source of financial support: This work was funded by Eisai Inc. (Woodcliff Lake, NJ).

REFERENCES

[1] Centers for Medicare & Medicaid Services. 42 CFR part 511 Medicare program; part B drug payment model; proposed rule. Fed Regist 2016;48:13230–61.
[2] Cherny NI, Sullivan R, Dafni U, et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol 2015;26:1547–73.
[3] Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. J Clin Oncol 2015;33:2563–77.
[4] Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology value framework: revisions and reflections in response to comments received. J Clin Oncol 2016;34:2925–34.
[5] National Comprehensive Cancer Network. NCCN Evidence Blocks™ frequently asked questions. Available from: https://www.nccn.org/evidenceblocks/pdf/EvidenceBlocksFAQ.pdf. [Accessed August 24, 2016].
[6] National Comprehensive Cancer Network. NCCN Evidence Blocks™ user guide. Available from: https://www.nccn.org/evidenceblocks/pdf/EvidenceBlocksUserGuide.pdf. [Accessed August 24, 2016].
[7] Ollendorf D, Chapman R, Khan S, et al. Treatment Options for Relapsed or Refractory Multiple Myeloma: Effectiveness, Value, and Value-Based Price Benchmarks. Boston, MA: Institute for Clinical and Economic Review, 2016.
[8] National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology (NCCN guidelines) with NCCN Evidence Blocks. Available from: https://www.nccn.org/evidenceblocks/. [Accessed August 25, 2016].
[9] Ollendorf D, Pearson SD. ICER evidence rating matrix: a user's guide. 2016. Available from: http://icer-review.org/wp-content/uploads/2016/01/Rating-Matrix-User-Guide-FINAL-v10-22-13.pdf. [Accessed June 15, 2016].
[10] Institute for Clinical and Economic Review. Addressing the myths about ICER and value assessment. Available from: https://icer-review.org/myths/. [Accessed August 10, 2016].
[11] Basch E. Toward a patient-centered value framework in oncology. JAMA 2016;315:2073.
[12] Chandra A, Shafrin J, Dhawan R. Utility of cancer value frameworks for patients, payers, and physicians. JAMA 2016;315:2069.
[13] Dalzell M. Considerations for designing "value calculators" for oncology therapies 2016: Available from: http://www.ajmc.com/journals/evidence-based-oncology/2016/peer-exchange-oncology-stakeholders-summit/considerations-for-designing-value-calculators-for-oncology-therapies. [Accessed June 21, 2016].
[14] Dangi-Garimella S. Lessons to learn from the NICE cancer care model. Available from: http://www.ajmc.com/journals/evidence-based-oncology/2016/july-2016/lessons-to-learn-from-the-nice-cancer-care-model/P-1. [Accessed August 2, 2016].
[15] Feinberg B, Lal L, Swint M. Is there a mathematical resolution to the cost-versus-value debate? Am J Manag Care 2015;21:SP542–4.
[16] Goulart BHL. Value: the next frontier in cancer care. Oncologist 2016;21:651–3.
[17] Neumann PJ, Cohen JT. Measuring the value of prescription drugs. N Engl J Med 2015;373:2595–7.
[18] Young RC. Value-based cancer care. N Engl J Med 2015;373:2593–5.
[19] American Cancer Society. Cancer facts and figures 2016. 2016. Available from: http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/. [Accessed June 21, 2016].
[20] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–90.
[21] Porter ME. What is value in health care? N Engl J Med 2010;363:2477–81.
[22] Eaton KD, Jagels B, Martins RG. Value-based care in lung cancer. Oncologist 2016;21:903–6.
[23] Ken Lee KH, Matthew Austin J, Pronovost PJ. Developing a measure of value in health care. Value Health 2016;19:323–5.
[24] Riva S, Pravettoni G. Value-based model: a new perspective in medical decision-making. Front Public Health 2016;4:118.
[25] National Health Council. The patient voice in value: the NHC patient-centered value model rubric. 2016. Available from: http://www.nationalhealthcouncil.org/sites/default/files/Value-Rubric.pdf. [Accessed December 9, 2016].
[26] Taylor PJ. An introduction to intraclass correlation that resolves some common confusions. Faculty paper, Programs in Science, Technology and Values, Critical and Creative Thinking, and Public Policy, University of Massachusetts, Boston, MA. Available from: http://www.faculty.umb.edu/pjt/09b.pdf. [Accessed June 15, 2016].