

BRIEF COMMUNICATION

Reliability of Oncology Value Framework Outputs: Concordance Between Independent Research Groups

Joseph C. Del Paggio, Sierra Cheng, Christopher M. Booth,
Matthew C. Cheung, Kelvin K. W. Chan

See the Notes section for the full list of authors' affiliations.
Drs Chan and Cheung contributed equally to this article.

Correspondence to: Dr. Kelvin K. W. Chan, 2075 Bayview Ave, Toronto, ON M4N 3M5, Canada (e-mail: kelvin.chan@sunnybrook.ca).

Abstract

Research groups are increasingly utilizing value frameworks, but little is known of their reliability. To assess framework concordance and interrater reliability between two major value frameworks currently in use, we identified all previously published datasets containing both scores from the American Society of Clinical Oncology Value Framework (ASCO-VF) and grades from the European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS). The intraclass correlation coefficient (ICC) was used to assess interrater reliability. Four eligible studies contained drugs evaluated by both value frameworks, resulting in a dataset of 39 grades/scores for discrete drug indications. ICC was 0.82 (95% confidence interval = 0.70 to 0.90) for ASCO-VF and 0.88 (95% confidence interval = 0.80 to 0.93) for ESMO-MCBS. Absolute concordance was found to be 5% for ASCO-VF and 44% for ESMO-MCBS, increasing to 74% and 80% when deviations within 20 points and 1 grade were considered, respectively. Interrater reliability of ASCO-VF and ESMO-MCBS is, therefore, near perfect, while absolute concordance is poor. This has implications when considering framework outputs in drug funding or treatment decision making.

Clinical value can be challenging to quantify, so value frameworks have been created to objectify therapeutic benefit. Various cohort studies have attempted to evaluate the congruity between two important frameworks—the American Society of Clinical Oncology Value Framework (ASCO-VF) (1) and the European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) (2)—finding discrepancy in their outputs (3), as well as overall fair correlation (4–7). Their reliability, however, is unclear: how consistent are framework outputs when calculated by independent users? The aim of this study was to assess interrater reliability and absolute concordance of ASCO and ESMO frameworks across research groups.

A PubMed search was performed on August 30, 2017, to identify all studies that scored/graded trials based on the ASCO-VF and ESMO-MCBS. Studies were included if both ASCO scores and ESMO grades were systematically applied to a cohort of anticancer therapies, and outputs (ie, scores/grades) were either 1) included in the publication and/or its supplementary appendix, or 2) available to the authorship group because of direct study involvement.

All corresponding trials for drugs subjected to framework analysis were tabulated. Individual ASCO-VF scores and ESMO-MCBS grades assigned by the authors to each respective drug were abstracted by two authors (S.C. and J.C.D.P.); any discrepancies were resolved. Inclusion criteria for the final dataset consisted of drugs for specific indications that had both ASCO scores and ESMO grades, rated independently by at least two different study groups.

The intraclass correlation coefficient (ICC) assesses the degree of consistency between ordinal, interval/continuous, and ratio variables and is defined as the ratio of variability between subjects (ie, collated framework outputs) to total variability (8). ICC values range from slight agreement (0.0–0.2) to near-perfect agreement (0.81–1.0) (9). ICC was assessed for the entire study cohort, as well as for the subset where scores/grades were derived from a single trial. Absolute concordance was assessed by calculating the frequency of studies with identical scores/grades from all assigned scores/grades. As a sensitivity analysis, concordance was also derived from the cohort of drug indications

Received: April 30, 2018; Revised: July 30, 2018; Accepted: August 10, 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

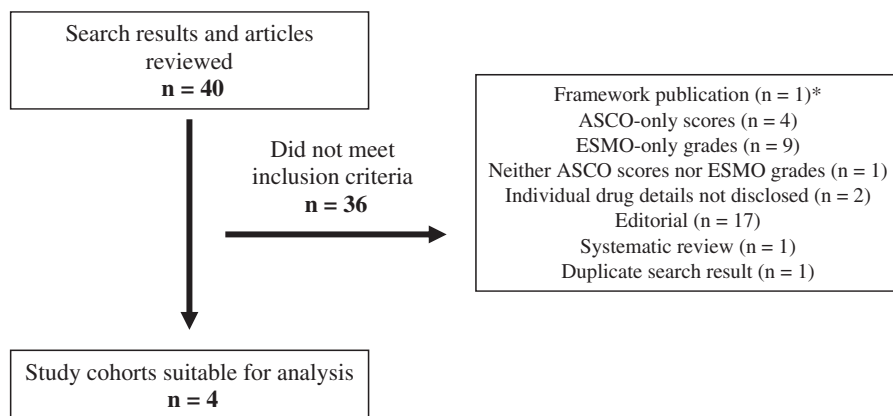


Figure 1. Identification of cohort studies to which American Society of Clinical Oncology (ASCO) scores and European Society for Medical Oncology (ESMO) grades were systematically applied. *ESMO-MCBS framework paper grades were included in the correlative analysis, where applicable.

scored/graded from single trials. The frequencies of deviation in scores/grades were calculated among raters within ± 10 points and ± 20 points and within ± 1 grade and ± 2 grades for ASCO-VF and ESMO-MCBS, respectively. Statistical analyses were conducted with R 3.3.1 (R Project, Vienna, Austria).

The initial search strategy yielded 40 articles, 36 of which did not meet the inclusion criteria (Figure 1). The final cohort consisted of four studies (10%) suitable for data extraction and agreement analysis, two of which were studies previously published by members of this authorship group (5,6). All four studies utilized the revised ASCO-VF (1) and version 1.0 of ESMO-MCBS (2).

The final cohort comprised 39 indications from 36 drugs (Table 1). Approximately 25% of these drug/indication combinations (10/39) consisted of scores/grades derived from more than one trial. A sample of grades from the ESMO authorship group were included in the cohort (2); no “gold-standard” scores were applicable from the ASCO authorship group.

Interrater reliability was found to be near-perfect for both ASCO scores (0.82, 95% CI = 0.7 to 0.9) and ESMO grades (0.88, 95% CI = 0.8 to 0.93) in all settings (Table 2). When drug scores/grades derived from multiple trials were removed from analysis ($n=29$ drug indications), ICCs were similar: 0.89 for ASCO-VF (95% CI = 0.72 to 0.92) and 0.9 (95% CI = 0.83 to 0.95). Using the ESMO-MCBS definition of “substantial clinical benefit” (ie, grades B, A, 4, or 5) to dichotomize authorship grades, ICC values were similar: 0.84 (95% CI = 0.74 to 0.9) for all drug grades and 0.89 (95% CI = 0.80 to 0.94) for drug grades derived from single randomized, controlled trials.

Absolute concordance among all authorship groups’ final ASCO scores and ESMO-MCBS grades was 5% and 44%, respectively (Table 3). For ASCO scores, concordance increased to 46% and 74% when respectively assessed within ± 10 points and ± 20 points; for ESMO grades, concordance increased to 80% and 90% when respectively assessed within ± 1 grade and ± 2 grades. For the sensitivity analysis, concordance was similar when derived from drug indications scored/graded from single trials ($n=29$ drug indications) (Table 3).

The increasing interest in quantifying the value of anticancer agents is becoming justified by mounting evidence that these approved drugs have little to know correlation between their cost and efficacy (5,10–12). As such, understanding framework reliability is essential. In this regard, we have found that:

1) interrater reliability by ICC is near perfect between users for both frameworks, while 2) absolute concordance is poor, but greater for ESMO-MCBS than ASCO-VF.

Although studies have previously shown near-perfect agreement between users assessing a single cohort of trials (6,13), this is the first study, to our knowledge, that systematically assessed outputs from ASCO-VF and ESMO-MCBS across published works. We show similar findings with respect to ICC. Comparable concordance, however, is only established with deviations of 1 grade for ESMO-MCBS and 20 points for ASCO-VF. Although the clinical meaning of these deviations may be apparent with ESMO’s definition of “substantial benefit” (ie, grades 4 or 5 in palliative studies and B or A in curative studies) (2), a change in 20 points is ambiguous by ASCO’s scale, particularly given its condemnation of direct score comparisons (14). Nonetheless, support for this degree of deviation is found in our previously assessed cohort of 111 drug approval trials with a median ASCO score of 22 (6): A deviation of 20 points about this median would exceed the interquartile range identified in this cohort (ie, 8–35). Overall, these degrees of imprecision in scores/grades may have significant clinical implications in framework utilization.

The differences in concordance found in this study are likely a consequence of each framework’s inherent scoring system: continuous ASCO-VF scores vs discrete ESMO-MCBS grades. Despite the concern that scores/grades derived from more than one trial (~25% of the cohort) may result in reduced reliability, ICCs only marginally improved when these studies were removed (Table 2). Nevertheless, the discrepancy in absolute framework outputs remains problematic even when indications evaluated from more than one trial were removed (Table 3). Causes of discrepancy (Table 1) are likely the result of differing interpretations of toxicity and quality-of-life benefits, while improved concordance for ESMO-MCBS is likely due to its relatively finite outputs. Framework authors have recently made efforts to clarify technicalities (14,15), but future iterations should strongly consider the publication of formal measurement studies (eg, interrater reliability) prior to framework distribution in order to adequately evaluate measurement characteristics and minimize user errors in framework utilization.

This study has limitations, the most significant of which is its sample size; however, value frameworks, and studies

Table 1. Trials of drugs evaluated by the American Society of Clinical Oncology (ASCO) Value Framework and the European Society for Medical Oncology (ESMO)-Magnitude of Clinical Benefit Scale by four independent research publications, as well as grades denoted by the ESMO framework authorship group, where applicable*

Drug	Indication	PMID (s) for framework outputs	Becker et al. (4)		Vivot et al. (3)		Cheng et al. (6)		Del Paggio et al. (5)		ESMO authorship grade (2)
			ASCO	ESMO	ASCO	ESMO	ASCO	ESMO	ASCO	ESMO	
Abiraterone acetate	Second-line treatment of prostate cancer	22995653, 21612468	23	4	34.6	4	34.3	4	X	X	4
Ado-trastuzumab emtansine	Second-line treatment of HER2-positive breast cancer	23020162	33.7	5	62.4	5	36.4	5	45	5	5
Afatinib	First-line treatment of non-small cell lung cancer with EGFR mutations	23816960	47.7	4	31.7	4	30.6	4	31	4	4
Bevacizumab	First-line treatment of colorectal cancer	15175435	23.5	1	31	3	X	X	X	X	3
Cabazitaxel	Second-line treatment of prostate cancer	20888992	45.2	2	40.5	2	25.2	2	X	X	2
Cabozantinib	Second-line treatment of medullary thyroid cancer	24002501	55.1	3	37.6	2	-17.3	3	X	X	X
Cobimetinib	First-line treatment of BRAF-mutated melanoma	25265494	28.7	3	52.2	4	X	X	X	X	4
Crizotinib	Second-line treatment of ALK-positive lung cancer	23724913	70	4	X	X	65.2	4	59	4	4
Dabrafenib	First-line treatment of BRAF-mutated melanoma	22735384	52	3	59.6	4	52.0	3	X	X	4
Enzalutamide	Second-line treatment of prostate cancer	22894553	59.7	4	52.6	4	40.6	4	X	X	X
Eribulin mesylate	Third-line treatment of breast cancer	21376385	18.5	2	18.3	2	18.0	2	17	2	2
Erlotinib	Second-line treatment of non-small cell lung cancer	16014882, 20493771	47.9	3	42.1	4	X	X	X	X	1
Everolimus	First-line treatment of renal cell carcinoma	18653228, 20549832	49.6	3	32.8	3	33.6	3	X	X	3
Everolimus	Second-line treatment of postmenopausal hormone-receptor-positive breast cancer	22149876	X	X	X	X	33.2	2	30	2	X
Ipilimumab	Second-line treatment of melanoma	20525992	32	2	28.9	4	37.9	4	X	X	4
Ixabepilone	Second-line treatment of breast cancer	17968020	35.5	0	18.3	2	12.9	2	X	X	X
Lapatinib	Second-line treatment of breast cancer	19786658, 20124187, 22689807	34.9	1	54.6	2	21.7	1	26	4	4
Lenvatinib	Second-line treatment of thyroid cancer	25671254	59.2	2	59.2	2	59.2	2	X	X	X
Necitumumab	First-line treatment of squamous non-small cell lung cancer	26045340	12.0	2	11.3	2	X	X	14	2	X

(continued)

Table 1. (continued)

Drug	Indication	PMID (s) for framework outputs	Becker et al. (4)		Vivot et al. (3)		Cheng et al. (6)		DeL'Paggio et al. (5)		ESMO authorship grade (2)
			ASCO	ESMO	ASCO	ESMO	ASCO	ESMO	ASCO	ESMO	
Nivolumab	Second-line treatment of melanoma	25795410	30.2	2	22.0	3	22.4	2	X	X	X
Nivolumab	Second-line treatment of squamous non-small cell lung cancer	26028407	X	X	X	X	73.1	5	61	4	X
Paclitaxel protein bound	First-line treatment of non-small cell lung cancer	22547591	61.2	4	X	X	22.2	4	29	1	X
Panitumumab	Second-line treatment of EGFR-expressing colorectal cancer	17470858, 20921462	34.3	2	19.8	1	19.8	2	X	X	2
Pazopanib	First-line treatment of renal cell carcinoma	20100962	43.2	2	46.3	3	42.4	2	X	X	3
Pemetrexed for injection	First-line treatment of mesothelioma	128660938	10	2	17.6	3	X	X	X	X	X
Pertuzumab	First-line treatment of HER 2-positive breast cancer	23602601, 25693012, 22149875, 23868905	38.9	5	46.8	4	38.9	4	31.7	4	4
Radium 223 dichloride	Second-line treatment of prostate cancer	23863050, 24836273	60.5	3	60.9	5	39.8	5	X	X	5
Ramucirumab	Second-line treatment of gastro-esophageal cancer	24094768	38.8	1	39.4	1	30.6	1	X	X	2
Regorafenib	Third treatment of colorectal cancer	23177514	3.5	1	3.4	1	4.4	1	3	1	1
Sorafenib	First-line treatment of renal cell carcinoma	17215530	51.3	3	53.6	2	X	X	X	X	3
Sunitinib	First-line treatment of renal cell carcinoma	17215529, 19487381	57.4	4	X	X	46	4	X	X	4
Sunitinib	Second-line treatment of gastrointestinal stromal tumor	17046465	36.1	3	36.1	3	X	X	X	X	3
Temsirolimus	First-line treatment of renal cell carcinoma	17538086	25.1	3	21.6	4	22.7	4	X	X	4
Trabectedin	Second-line treatment of liposarcoma or leiomyosarcoma	26371143	0.4	3	14.6	3	X	X	X	X	X
Trametinib	First-line treatment of BRAF-mutated melanoma	23020132, 22663011	62.6	4	52.7	4	52.6	3	X	X	4
Trifluridine and tipiracil	Third-line treatment of colorectal cancer	25970050	28.2	3	49.8	2	X	X	29.2	2	2
Vandetanib	First-line treatment of thyroid cancer	22025146	37	3	45.7	3	8.3	3	X	X	X
Vemurafenib	First-line treatment of BRAF-mutated melanoma	21639808, 24508103	15.6	2	66.5	4	49.7	4	X	X	4
Ziv-aflibercept	Second-line treatment of colorectal cancer	22949147	15.9	1	16.6	1	15.7	1	15	1	1

*X = no score/grade assigned by authorship group. EGFR = epidermal growth factor receptor.

Table 2. Interrater reliability of framework outputs between four independent authorship groups*

Framework	Setting	ICC (95% CI)
ASCO	All drugs scored (n = 39)	0.82 (0.7 to 0.9)
	Only those drugs scored from a single trial (n = 29)	0.85 (0.72 to 0.92)
ESMO	All grades (n = 39)	0.88 (0.8 to 0.93)
	Only those drugs graded from a single trial (n = 29)	0.9 (0.83 to 0.95)

*ASCO = American Society of Clinical Oncology; ESMO = European Society for Medical Oncology.

Table 3. Score/grade concordance and deviation frequencies for 39 evaluated drug indications as well as the 29 drug indications scored/graded from single trials*

Framework	Concordance between authorship groups	Deviance between authorship group	
		±10 points or ±1 grade	±20 points or ±2 grades
ASCO v2 Scores (all indications, n = 39)	2 (5%)	18 (46%)	29 (74%)
ESMO-MCBS Grades all indications, n = 39)	17 (44%)	31 (80%)	35 (90%)
ASCO v2 Scores (single trial evaluations, n = 29)	2 (7%)	16 (55%)	22 (76%)
ESMO-MCBS Grades (single trial evaluations, n = 29)	14 (48%)	25 (86%)	28 (97%)

*ASCO = American Society of Clinical Oncology; ESMO-MCBS = European Society for Medical Oncology-Magnitude of Clinical Benefit Scale.

utilizing them, are novel. Also, our score/grade deviations (ie, 1–2 and 10–20) are arbitrary, and changing these values would change our findings; nonetheless, these deviations were determined a priori and are felt to represent substantial differences in the respective framework outputs. Finally, given the evolving nature of these frameworks, the scores/grades culled from these published studies may be obsolete with future revisions. For example, since the time of our initial search strategy, the ESMO-MCBS group has published v1.1 of their framework (16); although grades across identical cohorts remain relatively stable (7,16), grading thresholds have been modified to aid in future framework revisions.

In conclusion, value frameworks are remarkably reliable when assessing trials of anticancer therapies. At present, however, the absolute concordance is poor, but greater for ESMO-MCBS than for ASCO-VF. This is an important consideration for users placing an onus on absolute framework scores/grades, particularly if the outputs are used for public policy implementation and patient/doctor decision making.

Notes

Affiliations of authors: Department of Medicine, Division of Medical Oncology, University of Toronto, Toronto, Ontario Canada (JCDP, SC, MCC, KKWC); Departments of Oncology and Public Health Sciences, Queen's University, Kingston, Ontario, Canada (CMB); Division of Cancer Care and Epidemiology, Queen's University Cancer Research Institute, Kingston, Ontario, Canada (CMB); Canadian Centre for Applied Research in Cancer Control, Toronto, Ontario, Canada (KKWC).

Disclosures

None.

Funding

The Canadian Centre for Applied Research in Cancer Control received core funding from the Canadian Cancer Society Research Institute (grant 2105-703549).

References

- 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(24):2925–2934.
- 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(8):1547–1573.
- Vivot A, Jacot J, Zeitoun JD, et al. Clinical benefit, price and approval characteristics of FDA-approved new drugs for treating advanced solid cancer, 2000–2015. *Ann Oncol*. 2017;28(5):1111–1116.
- Becker DJ, Lin D, Lee S, et al. Exploration of the ASCO and ESMO value frameworks for antineoplastic drugs. *J Oncol Pract*. 2017;13(7):e653–e665.
- Del Paggio JC, Sullivan R, Schrag D, et al. Delivery of meaningful cancer care: a retrospective cohort study assessing cost and benefit with the ASCO and ESMO frameworks. *Lancet Oncol*. 2017;18(7):887–894.
- Cheng S, McDonald EJ, Cheung MC, et al. Do the American Society of Clinical Oncology value framework and the European Society of Medical Oncology Magnitude of Clinical Benefit Scale measure the same construct of clinical benefit? *J Clin Oncol*. 2017;35(24):2764–2771.
- Del Paggio JC, Sullivan R, Hopman WM, et al. Re-aligning the ASCO and ESMO clinical benefit frameworks for modern cancer therapies. *Ann Oncol*. 2017;29(3):773–774.
- Kim HY. Statistical notes for clinical researchers: evaluation of measurement error 1: using intraclass correlation coefficients. *Restor Dent Endod*. 2013;38(2):98–102.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
- Booth CM, Del Paggio JC. Approvals in 2016: questioning the clinical benefit of anticancer therapies. *Nat Rev Clin Oncol*. 2017;14(3):135–136.
- Mailankody S, Prasad V. Five years of cancer drug approvals: innovation, efficacy, and costs. *JAMA Oncol*. 2015;1(4):539–540.
- Saluja R, Arciero VS, Cheng S, et al. Examining trends in cost and clinical benefit of novel anticancer drugs over time. *J Oncol Pract*. 2018;14(5):e280–e294.
- Bentley TGK, Cohen JT, Elkin EB, et al. Measuring the value of new drugs: validity and reliability of 4 value assessment frameworks in the oncology setting. *J Manag Care Spec Pharm*. 2017;23(6-a Suppl):S34–S48.
- Cherny NI, Dafni U, Bogaerts J, et al. Reply to: re-aligning the ASCO and ESMO clinical benefit frameworks or modern cancer therapies. *Ann Oncol*. 2017;29(3):774–775.
- Schnipper LE, Schilsky RL. Are value frameworks missing the mark when considering long-term benefits from immuno-oncology drugs? *JAMA Oncol*. 2018;4(3):333–334.
- Cherny NI, Dafni U, Bogaerts J, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol*. 2017;28(10):2340–2366.