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

JC Del Paggio[1], Sierra Cheng[1], Christopher M. Booth[2,3], Matthew C. Cheung[1]*, Kelvin K. W. Chan[1,4]*

[1]Department of Medicine, Division of Medical Oncology, University of Toronto, Toronto, ON, Canada

[2]Departments of Oncology and [3]Public Health Sciences, Queen's University, Kingston, ON, Canada

[4]Canadian Centre for Applied Research in Cancer Control (ARCC), Toronto, Canada

BACKGROUND

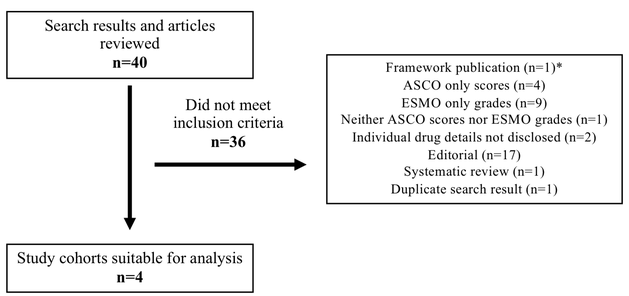
- Little is known of the reliability of modern value frameworks, despite their increasing utilization.

OBJECTIVE

- This study aims to assess the concordance and inter-rater reliability of previously published scores/grades of drugs using the American Society of Clinical Oncology Value Framework (ASCO-VF) and European Society for Medical Oncology Meaningful Clinical Benefit Scale (ESMO-MCBS).

METHODS

- All previously published data sets containing both ASCO-VF scores and ESMO-MCBS grades were identified through a *PubMed* search of studies utilizing value frameworks.
- The final data set included drugs evaluated by both ASCO-VF and ESMO-MCBS grades, independently rated by least two different studies (Figure 1); these scores/grades were abstracted for statistical evaluation.
- Intraclass correlation coefficient (ICC) was used to assess inter-rater reliability.



*ESMO-MCBS framework paper grades were included in the correlative analysis, where applicable.

Figure 1. Identification of randomised controlled trials of systemic therapy in breast cancer, non-small-cell lung cancer, colorectal cancer, and pancreatic cancer published 2011–15

RESULTS

Drug	Indication	PMID(s) for Framework Outputs	Becker et al.[1]		Vivot et al.[2]		Cheng et al.[3]		Del Paggio et al.[4]		ESMO Authorship Grade[5]
			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	20888992	45.2	2	40.5	2	25.2	2	X	X	2

Table 1. Sample abstraction of drugs scores/grades evaluated by ASCO-VF and ESMO-MCBS, respectively, by four independent research publications, as well as grades denoted by the ESMO framework authorship group, where applicable.

Framework	Authorship Concordance	Deviation in Scores/Grades	
		±10 points or ±1 grade	±20 points or ±2 grades
ASCO-VF (v2)	2 (5%)	18 (46%)	29 (74%)
ESMO-MCBS	17 (44%)	31 (80%)	35 (90%)

Table 2. Score/grade concordance and deviation frequencies for 39 evaluated drugs.

Framework	Setting	ICC	95% CI
ASCO	All drugs (n=39)	0.82	0.7 to 0.9
	Only drugs evaluated from single RCT (n=29)	0.85	0.72 to 0.92
ESMO	All drugs (n=39)	0.88	0.8 to 0.93
	Only drugs evaluated from single RCT (n=29)	0.9	0.83 to 0.95

Table 3. Inter-rater reliability of framework outputs between four independent authorship groups.

- Out of 40 eligible studies, 4 contained drugs that were evaluated by both frameworks (Figure 1).
- The final data set contained 36 drugs for 39 indications (Table 1).
- 29 drug indications were evaluated by the authorshop groups using a single RCT; more than one RCT was used to evaluate 10 drug indications.
- 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 (Table 2).
- ICC was 0.82 (95% CI 0.70 - 0.90) for ASCO-VF and 0.88 (95% CI 0.80 - 0.93) for ESMO-MCBS (Table 3).
- Dichotomizing ESMO-MCBS grades by "substantial clinical benefit" (i.e., grades B, A, 4, or 5) yielded similar ICCs.

REFERENCES

1. Becker D.J., et al. Exploration of the ASCO and ESMO Value Frameworks for Antineoplastic Drugs. Journal of Oncology Practice 13, e653-e665 (2017).
2. Vivot A., et al. Clinical Benefit, Price and Approval Characteristics of FDA-approved New Drugs for Treating Advanced Solid Cancer, 2000-2015. Annals of Oncology 28(5):1111-1116 (2017).
3. Cheng S., 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? Journal of Clinical Oncology 35(24):2764-2771 (2017).
4. Del Paggio J.C., et al. Delivery of meaningful cancer care: a retrospective cohort study assessing cost and benefit with the ASCO and ESMO frameworks. Lancet Oncology 18(7): 887-894 (2017).
5. Cherny, N.I., 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). Annals of Oncology 26(8):1547-73 (2015).

DISCUSSION

- Our results show that despite excellent ICCs, absolute concordance is poor for ASCO-VF and fair for ESMO-MCBS.
- ESMO-MCBS concordance improves substantially with a deviation of 1 grade above and below the output, while ASCO-VF requires a deviation of 20 points in order to achieve comparable concordance.
- The complexity in tabulating framework outputs partially explains the differences in absolute concordance: ASCO-VF involves points that cumulate to 120 or above vs ESMO-MCBS which has discrete grading thresholds (e.g. 1 to 5 for palliative trials, and C to A for curative trails).
- Discrepant scores/grades were also seen due to authorship groups evaluating the same drug indication using a differing RCTs; however, this minimally changed the ICC (Table 3).
- Continued updating of the ASCO-VF and ESMO-MCBS is a major limitation of the data generated in this study.

CONCLUSIONS

- Current value frameworks are remarkably reliable in their outputs.
- Absolute concordance in framework outputs, however, is poor.
- Future iterations of each framework should strongly consider formal measurement studies (e.g. inter-rater reliability) prior to publication in order to adequately evaluate their measurement characteristics and prevent errors in framework utilization.
- These are important considerations for users placing an onus on the absolute grades/scores of these frameworks, particularly if the outputs are used for public policy implementation and patient/doctor decision-making.