ACTA ONCOLOGICA
2021, VOL. 60, NO. 9, 1225–1232
https://doi.org/10.1080/0284186X.2021.1942546

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Analyzing the clinical benefit of newer therapies for advanced or metastatic non-small-cell lung cancer: application of the ESMO-magnitude of clinical benefit scale v1.1

Ricardo García-Fumero[a], Cristina Fernández-López[a,b], Miguel Ángel Calleja-Hernández[a,c] and José Expósito-Hernández[a,d]

[a]University of Granada, Granada, Spain; [b]Department of Pharmacy, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain; [c]Department of Pharmacy, Virgen Macarena University Hospital, Seville, Spain; [d]Department of Oncology, Virgen de las Nieves University Hospital, Granada, Spain

ABSTRACT

Background: Despite newer therapies, advanced or metastatic non-small-cell lung cancer (NSCLC) continues to be the leading cause of cancer-related deaths worldwide. Deficits in the design and methods of randomized controlled trials (RCTs) may contribute to reducing the clinical benefit of therapies in oncology. To prioritize treatments based on efficacy results and toxicity data, the *European Society for Medical Oncology* (ESMO) has developed the Magnitude of Clinical Benefit Scale (MCBS). The objective of this study was to apply the ESMO-MCBS v1.1 to a cohort of RCTs on therapies for advanced or metastatic NSCLC.

Material and methods: Phase III and pivotal phase II trials, published between 2013 and 2018, investigating drug therapies for advanced NSCLC were included. *PubMed* was specifically searched for efficacy/toxicity updates. Treatments were graded 5 to 1 on the ESMO-MCBS v1.1, using the lower limit of the 95% confidence interval of the hazard ratio (HR), where scores 5 and 4 represent a substantial clinical benefit. Additionally, scores using the point estimate HR were generated, for comparison. Discrepancies between our grade estimations and the ones published on the ESMO website, as scorecards, were identified.

Results: ESMO-MCBS scores were calculated for 42 positive clinical trials. 54.8% met the ESMO-MCBS thresholds for clinically meaningful benefit (final grade of 4 or 5). That percentage decreased to 40.5% when considering the point estimate of the HR. 50.0% of the trials had no published scorecard on the ESMO website and discrepancies affected 11 (26.2%) studies.

Conclusion: Almost half of the RCTs showing a statistically significant result favoring the experimental arm, failed to demonstrate a substantial clinical benefit according to the ESMO framework.

ARTICLE HISTORY
Received 17 March 2021
Accepted 8 June 2021

KEYWORDS
ESMO-MCBS; non-small-cell lung cancer; advanced or metastatic; clinical benefit; scorecard

Introduction

Globally, lung cancer represents the first cause of cancer deaths accounting for 18.4% of the total estimated number of deaths in 2018 [1]. About 84% of lung cancers are diagnosed at an advanced stage with an estimated 5-year survival rate of 6% [2], more than 80% of diagnosed patients corresponding to non-small-cell lung cancer (NSCLC).

Randomized controlled trials (RCTs) have led to many important therapeutic advances in oncology. Newer therapies like targeted agents and immunotherapy are allowing patients with advanced NSCLC to live longer than ever before, however, this 5-year survival rate for advanced or metastatic NSCLC continues to be very distant from the 57% estimated for localized NSCLC [2].

Thus, there is growing concern about the magnitude of benefit from new treatments in oncology, as too many RCTs could be at high risk of bias due to deficits in their design and methods [3–5]. Furthermore, the trends in anticancer drug costs are compromising access to these drugs that are *already unaffordable in some countries* [6]. For all these reasons, the value of the drug, that is the relation between its benefit and its cost, is an increasingly important issue to address for a high-quality cancer care [7].

In light of this emerging concern, both the *European Society for Medical Oncology* (ESMO) [8,9] and the *American Society of Clinical Oncology* (ASCO) [10,11] have developed scales to provide a framework to assess the clinical benefit of new cancer therapies. The ESMO-Magnitude of Clinical Benefit Scale (MCBS) ranks the clinical benefit in a structured manner, by taking into account reported outcomes in terms of longer survival (progression free survival [PFS], overall survival [OS]) and better survival (quality of life [QoL], toxicity). Furthermore, what is also important, the ESMO-MCBS seems

CONTACT Cristina Fernández-López cristina.fernandezl@iconcologia.net Department of Pharmacy, Catalan Institute of Oncology, Gran Via de l'Hospitalet 199-203, L'Hospitalet de Llobregat 08908

Supplemental data for this article can be accessed here.

to be very reliable in advanced or metastatic diseases throughout all treatment settings in daily practice [12].

The primary objective of this study is to assess the clinical benefit of new therapies studied for advanced or metastatic NSCLC by applying the ESMO-MCBS v1.1 to a cohort of RCTs published between 2013 and 2018. Additionally, we evaluate the reproducibility of the scale in this palliative setting by comparison with the corresponding ESMO-MCBS scorecards.

Methods

Search strategy and study selection criteria

A structured search was conducted to identify phase III and pivotal phase II RCTs published between 2013 and 2018 on chemotherapy, targeted therapies, or immunotherapy agents for patients with advanced NSCLC. MEDLINE (accessed *via Ovid SP*) and EMBASE (*Ovid SP*) were consulted using the following search terms: non-small-cell lung, cancer OR carcinoma, humans, advanced OR metastatic, drug therapy, randomized controlled trials, phase II AND phase III.

The inclusion criterion was the comparison of at least two arms of drug therapies in patients with advanced NSCLC. To a lesser extent, single-arm phase-II trials were included if they were pivotal, that is key studies aimed to demonstrate the efficacy and safety of a new drug to obtain its marketing approval by regulatory authorities. We also considered those RCTs comparing different dosage regimens of the same agent or combination of agents. For trials with two or more experimental arms, the arm selected for evaluation in this review was the one which obtained the best primary endpoint result. When different publications of the same RCT (including further data on survival or quality of life) were available within our period of study, the latest data were considered. *PubMed* was additionally searched particularly for publication of survival updates or quality of life assessments specified within the original study.

Exclusion criteria were: other than stages IIIB or IV NSCLC studies; exploratory (non-pivotal) phase I/II trials; not preplanned subgroup analyses; any intervention study not including drug therapies; non-randomized and non-pivotal clinical trials; meta-analyses or reviews reporting data from multiple RCTs; prematurely stopped RCTs due to futility or unacceptable toxicity and studies in a language other than English. Selected trials were scrutinized to identify potential duplication or overlap.

Data extraction and management

Two investigators (RGF and CFL) independently reviewed all abstracts applying the exclusion criteria and extracted data from the eligible studies. A data abstraction form was developed to record details regarding study design, endpoints (including response rates, PFS, OS, QoL and toxicity), and conclusions. Disagreements were discussed between both investigators to reach a consensus.

ESMO-MCBS scoring

The ESMO-MCBS v1.1 was applied to the selected RCTs that demonstrated either a statistically significant result for the primary outcome or a conclusion that supported non-inferiority, as the ESMO-MCBS v1.1 states [9]. For the noncurative setting reviewed there are two forms (2a and 2 b) available that consider the absolute gain in the predefined primary and secondary endpoints and the lower limit of the 95% confidence interval (CI) of the corresponding hazard ratio (HR). For non-inferiority trials, the form 2c was developed and considers QoL/toxicity data for assigning the score. Another form (form 3) is available for the scoring of single-arm studies. The preliminary score was adjusted according to different ESMO stipulations on toxicity, QoL, long-term survival data, etc. Palliative treatments were eventually graded 5 to 1, where scores 5 and 4 represent a substantial clinical benefit.

As proposed in the ESMO framework [9], the lower limit of the 95% CI of the HR was used to assign the preliminary ESMO-MCBS grade. Additionally, we generated the scores using the point estimate HR for comparison. Subsequently, the preliminary score was upgraded or downgraded, where required, according to the adjustments included in the ESMO-MCBS v1.1 forms [9]. For scoring of single-arm pivotal phase-II studies, we employed the form 3 which does not consider the HR values but the median PFS, overall response rates (ORR) and duration of response rates.

At the same time, to evaluate the ESMO-MCBS daily practicability and reproducibility, we compared our grade estimations with those published scores available on the ESMO website as ESMO-MCBS scorecards (https://www.esmo.org/guidelines/esmo-mcbs/esmo-mcbs-scorecards). Not only discrepancies in the final grade value but also those in the application of the MCBS in terms of form utilization, analyzed studies, and adjustments were detected.

Statistical analysis

Data were collected in an Excel file designed for this review, and imported into SPSS 19.0 (IBM, Chicago, IL) for statistical analysis. Given the non-parametric distribution of medians, a bivariate analysis using the Mann–Whitney and Kruskal–Wallis tests was conducted to evaluate how the ESMO-MCBS v1.1 scores were influenced by median OS and median PFS. Substantial benefit scores of 4 to 5 versus scores 1 to 3 were analyzed. Results were considered significant at P value < 0.05.

Results

As shown in Figure 1, the structured search resulted in 775 studies, but only 92 studies were selected after applying the study eligibility criteria. The 42 trials finally included were those in which the ESMO-MCBS v1.1 scoring could be performed (statistically significant result favoring the experimental arm) and involved a total of 21,051 patients with advanced NSCLC. Out of these studies, 25 (59.5%)

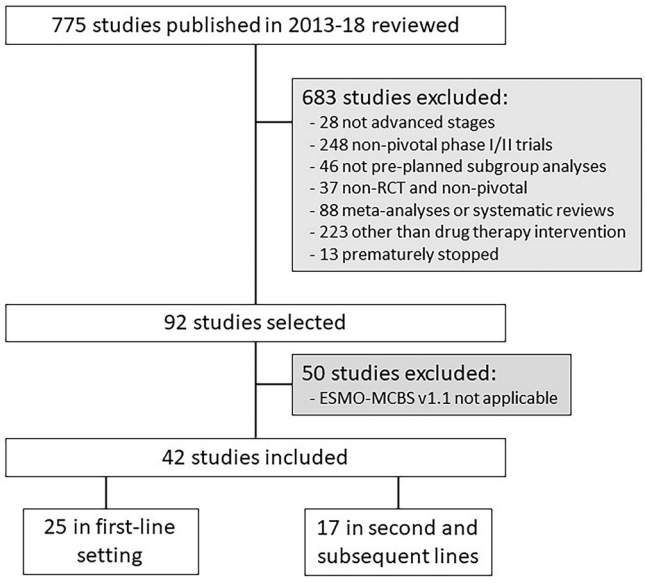

Figure 1. Flow diagram depicting the trial selection process for the review. RCT: randomized controlled trial; ESMO-MCBS: ESMO-Magnitude of Clinical Benefit Scale.

investigated first-line therapies and 17 (40.5%) examined second or subsequent lines of therapy. Phase III trials comprised 37 (88.1%) of all the studies included. There was a single-arm, phase-II clinical trial leading to registration of the examined drug, lorlatinib. The primary endpoint was PFS in 59.5% of trials and OS in 28.6%. QoL data were available for 29 of 42 (69.0%) included trials. Other characteristics are listed in Table 1.

Final ESMO-MCBS grades given to the studies included were based on OS data (form 2a) in 45.2% (19/42) of the trials; on PFS differences (form 2 b) in 42.9% (18/42); on the non-inferiority design of the studies (form 2c) in 9.5% (4/42); and based on ORR (form 3) in 2.4% (1/42) of the total. Detailed scoring and adjustments on ESMO-MCBS v1.1 application to our cohort is available in Table 2 and Supplementary Table S1 (available online) for first-line setting studies and second and subsequent lines of therapy, respectively.

Trials meeting the ESMO-MCBS threshold for a clinically meaningful benefit, attaining a final grade of 4 or 5, represented 54.8% (23/42) of the total. When considering the point estimate of the HR instead of the lower limit of its 95% CI, the percentage of the therapies that met that threshold decreased (40.5% *vs* 54.8%), as shown in Table 3. Stratifying by final scores demonstrates that 7.1% of the cohort (3/42) reached a final score of 1; 9.5% (4/42) a score of 2 and 28.6% (12/42) a score of 3. Grade 4 was achieved by the studied treatment in 18 (42.9%) trials while grade 5 only by 5 (11.9%). However, not all the drugs that obtained such grades of significant benefit did eventually access the market; 4 experimental therapies (nedaplatin, S-1, rmhTNF, and anlotinib) corresponding to 5 trials did not undergo further research. Among the trials that achieved a score of 4 or 5, 3/23 (13.0%) corresponded to cytotoxic agents, 13/23 (56.5%) examined targeted therapies, and 7/23 (30.5%) evaluated immunotherapeutic drugs. Trials on monoclonal antibodies

Table 1. Characteristics of the 42 clinical trials included.

Characteristic	*N*	%
Histology		
Squamous NSCLC	4	9.5
Non-squamous NSCLC	15	35.7
Both histology types	23	54.8
Therapy		
Cytotoxic	7	16.7
Targetet	23	54.8
Monoclonal antibody	3	7.1
Immunotherapy	9	21.4
Primary endpoint		
Overall survival (OS)	12	28.6
Progression-free survival (PFS)	25	59.5
Co-primary endpoints (OS + PFS)	4	9.5
Overall response rate (ORR)	1	2.4
Sample size		
Mean	347	
Median	387	
Range	289−601	
Industry sponsorship		
Yes	39	92.8
No	3	7.2

(bevacizumab, ramucirumab and necitumumab) did not reach this threshold. Furthermore, toxicity/QoL adjustments were needed in 20/23 (87.0%) of trials to achieve the ESMO threshold for substantial clinical benefit.

Median PFS and median OS were slightly higher for clinical trials achieving the ESMO threshold for significant clinical benefit (scores 4 to 5) than for those that did not (scores 1 to 3). However, no statistically significant differences were found in median PFS ($P = 0.734$) and median OS ($P = 0.849$) between both groups of trials.

When comparing our final ESMO-MCBS scores with the pertinent ESMO-MCBS scorecards available on the ESMO website, we found that 21/42 (50.0%) of the trials had no published scorecard; 15 of them on experimental therapies that had granted the market authorization by the *Food and Drug Administration* and/or the *European Medicines Agency,* and continue to be authorized for human use. Discrepancies affected a total of 11 (26.2%) studies for the following reasons: study selection for the ESMO-MCBS scoring ($n = 8$); toxicity and/or QoL adjustment ($n = 4$); and cohort of patients contemplated for scoring ($n = 1$). However, final scores differed in value only in 4 of the 21 trials with published scorecards (9.5% of the total): *PROFILE 1014* on crizotinib, *J025567* on bevacizumab plus erlotinib, *KEYNOTE-024* on first-line pembrolizumab and *KEYNOTE-010* on second-line pembrolizumab 10 mg/Kg. Details on this comparison are summarized in Table 4.

Discussion

In the present study, we have applied the ESMO framework [9] to measure the magnitude of clinical benefit into which the results from phases II and III trials on drug therapy for advanced NSCLC are translated. Our cohort of 42 studies comprises diverse treatment options with palliative intent, from first-line to salvage therapies, published between 2013 and 2018.

Phase III studies represent 90.5% (38/42) of the total, leading to a higher quality of evidence in contrast to phase II

Table 2. Characteristics of trials on first-line therapies for advanced NSCLC stratified by histology and their corresponding final ESMO-MCBS scores.

Trial name or first author	Year(s) of publication	Control arm	Experimental arm	Primary outcome (PO)	PO control arm (mo)	PO gain (mo)	PO HR (95%CI)	QoL assessment	Toxicity adjustment or comment	Final score and form	Led to approval[a]
SQUAMOUS (SQ) ADVANCED NSCLC											
Shukuya T	2015	CsP + DCX	Nedaplatin + DCX	OS	11.4	2.2	0.81 (0.65–1.02)	No	Less grades 3-4 toxic effects. Adj 1	4 (form 2a)	No
SQUIRE	2015	CsP + GEM	CsP + GEM + Necitumumab	OS	9.9	1.6	0.84 (0.74–0.96)	EQ-5D-3L, LCSS	NA	1 (form 2a)	Yes: 1 L, SQ, EGFR
NON-SQUAMOUS (nSQ) ADVANCED NSCLC											
AVAPERL	2013, 2014	Bevacizumab	Bevacizumab + PMX	PFS	3.7	3.7	0.57 (0.44–0.75)	QLQ-LC13, QLQ-C30	NA	2 (form 2 b)	No
PARAMOUNT	2013, 2014	Placebo	PMX	PFS (OS improved)	11	2.9	0.78 (0.64–0.96)	EQ-5D	NA	3 (form 2a)	Yes: maintenance therapy, nSQ
LUX-Lung 3	2013, 2018	CsP + PMX	Afatinib	PFS	6.9	4.2	0.58 (0.43–0.78)	QLQ-LC13, QLQ-C30	Diarrhea G3-4 (0 vs 14.4%). Rash G3-4 (0 vs 16.2%). Adj d	4 (form 2 b)	Yes: 1 L, EGFR
PROFILE 1014	2014, 2018	CbP or CsP + PMX	Crizotinib	PFS (OS improved)	19.2	40.6	0.346 (0.081–0.718)	QLQ-LC13, QLQ-C30, EQ-5D	Adj 1 for QoL	5 (form 2a)	Yes: 1 L, ALK
J025567	2014, 2018	Erlotinib	Bevacizumab + Erlotinib	PFS	9.7	6.3	0.54 (0.36–0.79)	FACT-L	No OS advantage and not improved QoL (Adj c)	2 (form 2 b)	Yes: 1 L, nSQ
LUX-Lung 6	2014, 2015	CsP + GEM	Afatinib	PFS	5.6	5.4	0.28 (0.20–0.39)	QLQ-LC13, QLQ-C30	Improved (adjustment d)	4 (form 2 b)	Yes: 1 L, EGFR
LUX-Lung 7	2016, 2017	Gefitinib	Afatinib	PFS	10.9	0.1	0.73 (0.57–0.95)	EQ-5D	NA	1 (form 2 b)	Yes: 1 L, EGFR
ASCEND-4	2017	CsP or CbP + PMX	Ceritinib	PFS	8.1	8.5	0.55 (0.42–0.73)	EQ-5D-5L, QLQ-LC13, QLQ-C30, LCSS	Improved QoL. Adj d	4 (form 2 b)	Yes: 2 L, ALK
ARCHER 1050	2017, 2018	Gefitinib	Dacomitinib	PFS (OS improved)	26.8	7.3	0.76 (0.58–0.99)	EQ-5D, QLQ-LC13, QLQ-C30	Increased toxicity	3 (form 2a)	Yes: 1 L, EGFR
FLAURA	2018, 2020	Erlotinib or Gefitinib	Osimertinib	PFS (OS improved)	31.8	6.8	0.80 (0.64–1.00)	No	Improved QoL	4 (form 2a)	Yes: 1 L, EGFR
KEYNOTE-189	2018	CbP or CsP + PMX	CbP or CsP + PMX + Pembrolizumab	OS	11.3	>3.0	0.49 (0.38–0.64)	No	NA	4 (form 2a)	Yes: 1 L, nSQ
IMpower 150	2018	CbP + PCX + Bevacizumab	CbP + PCX + Bevacizumab + Atezolizumab	OS	14.7	4.5	0.78 (0.64–0.96)	No	NA	3 (form 2a)	Yes: 1 L, nSQ
BOTH SQ and nSQ ADVANCED NSCLC											
ATLAS	2013	Bevacizumab maintenance	Bevacizumab + Erlotinib	PFS	3.7	1.1	0.71 (0.58–0.86)	No	NA	2 (form 2 b)	Yes: 1 L, nSQ
FASTACT-2	2013	CbP or CsP + GEM	CbP or CsP + GEM + Erlotinib	PFS (OS improved)	15.2	3.1	0.79 (0.64–0.99)	FACT, TOI	Rash G3-4 (0.4% vs 5%). No adjustment	4 (form 2a)	No
OPTIMAL	2015	CbP + GEM	Erlotinib	PFS	4.6	8.5	0.16 (0.10–0.26)	FACT-L, LCS	Improved (adjustment d)	4 (form 2 b)	Yes: 1 L, EGFR
ENSURE	2015	CsP + GEM	Erlotinib	PFS	5.5	5.5	0.34 (0.22–0.51)	No	NA	3 (form 2 b)	Yes: 1 L, EGFR
INFORM	2015	Placebo	Gefitinib	PFS	2.6	2.2	0.42 (0.33–0.55)	FACT-L	Toxic deaths incremental rate = 2% (adj b). Improved QoL (adj d)	3 (form 2 b)	Yes: EGFR
Kubota K	2015	CsP + DCX	CsP + S-1	Non-inferiority	17.1	1.0	1.013 (0.837–1.227)	QLQ-LC13, QLQ-C30	Improved	4 (form 2c)	No
KEYNOTE-024	2016, 2017, 2019	CbP + PMX or CsP + PMX or CbP + GEM or CsP + GEM or CbP + PCX	Pembrolizumab	PFS	14.2	15.8	0.63 (0.47–0.86)	EQ-5D-3L, QLQ-LC13, QLQ-C30	NA	4 (form 2a)	Yes: 1 L, PD-L1 ≥ 50%
ALEX	2017	Crizotinib	Alectinib	PFS	10.4	15.3	0.47 (0.34–0.65)	No	NA	3 (form 2 b)	Yes: 1 L, ALK
Ferry D	2017	CsP 80 mg + GEM	CsP 50 mg + GEM	Non-inferiority	9.5	0.5	1.13 (0.99–1.29)	EQ-5D, QLQ-LC13, QLQ-C30	NA	3 (form 2c)	No
Ouyang X	2018	CsP + VNR	CsP + VNR + Dulanermin	PFS	3.5	2.9	0.403 (0.318–0.512)	No	NA	3 (form 2 b)	No
ALTA-1L	2018, 2020	Crizotinib	Brigatinib	PFS	11.0	13.0	0.49 (0.35–0.68)	No	NA	3 (form 2 b)	Yes: 1 L, ALK

NA: not applicable; adj: adjustment; HR: hazard ratio; MCBS: magnitude of clinical benefit scale; QoL: quality of life; PFS: progression-free survival; OS: overall survival; 1 L: first-line setting; 2 L: second or subsequent line of therapy; CbP: carboplatin; CsP: cisplatin; GEM: gemcitabine; PCX: paclitaxel; PMX: pemetrexed; DCX: docetaxel; rmhTNF: recombinant mutated human tumor necrosis factor.

[a]Trials that led to approval are those which appear alongside the 'EPAR' for EMA or the 'Label' for FDA approvals. Indication for use in advanced NSCLC setting is also reported.

References of the clinical trials included are available online as Supplementary material.

Table 3. Trials that show disparity between the ESMO-MCBS score based on the lower limit of the 95% confidence interval of hazard ratio (final score) and the reported or point estimate hazard ratio ESMO-MCBS score.

Analyzed treatment(s)	Setting	Primary outcome (PO)	PO HR (95%CI)	Toxicity adjustment/ comment	Final score and form	HR point estimate MCBS
Bevacizumab + Erlotinib	First-line, SQ and nSQ	PFS	0.71 (0.58–0.86)	–	2 (form 2 b)	1
PMX	First-line, nSQ, maintenance	PFS (OS improved)	0.78 (0.64–0.96)	–	3 (form 2a)	1
CbP or CsP + GEM + Erlotinib	First-line, SQ and nSQ	PFS (OS improved)	0.79 (0.64–0.99)	Rash G3-4 (0.4% vs 5%). No adjustment	4 (form 2a)	2
Nedaplatin + DCX	First-line, SQ	OS	0.81 (0.65–1.02)	Less G3-4 toxic effects. Adjustment 1	4 (form 2a)	2
Afatinib	First-line, nSQ	PFS	0.73 (0.57–0.95)	–	1 (form 2 b)	1
Dacomitinib	First-line, nSQ	PFS (OS improved)	0.76 (0.58–0.99)	–	3 (form 2a)	1
Osimertinib	First-line, nSQ	PFS (OS improved)	0.80 (0.64–1.00)	Improved QoL	4 (form 2a)	1
CbP + PCX + Bevacizumab + Atezolizumab	First-line, nSQ	OS	0.78 (0.64–0.96)	–	3 (form 2a)	1
CsP/CbP + DCX + rmhTNF	Second or subsequent line, SQ and nSQ	OS	0.75 (0.63–0.89)	Less G3-4 nausea and vomiting	5 (form 2a)	2
Atezolizumab	Second or subsequent line, SQ and nSQ	OS	0.73 (0.53–0.99)	–	3 (form 2a)	1
Atezolizumab	Second or subsequent line, SQ and nSQ	OS	0.75 (0.64–0.89)	Improved	5 (form 2a)	2
Anlotinib	Second or subsequent line, SQ and nSQ	OS	0.68 (0.54–0.87)	Improved	4 (form 2a)	3

NA: not applicable; HR: hazard ratio; MCBS: magnitude of clinical benefit scale; QoL: quality of life; PFS: progression-free survival; OS: overall survival; CbP: carboplatin; CsP: cisplatin; GEM: gemcitabine; PCX: paclitaxel; DCX: docetaxel; rmhTNF: recombinant mutated human tumor necrosis factor.

design. The remaining four phase-II trials included in our review examine therapies that granted a market authorization based on preliminary efficacy results. This tendency of accelerated authorization from regulatory authorities may contribute to reducing the timeframe for new drugs to enter the market at the expense of clear evidence that they improve patients' OS or QoL, even in post-marketing studies [13].

The utilization of MCBS grading forms based on OS and those based on PFS is comparable (45.2% *vs* 42.9%, respectively). Our results reveal that despite the primary endpoint of the included studies was PFS in 59.5% of the total, only 42.9% of trials were evaluated with the form 2 b. The remaining proportion (16.6%, 7 trials) was thus assessed as if their primary endpoint was OS instead, just as the ESMO-MCBS states. An evaluation according to the form designed for OS (form 2a) is required by the ESMO framework when this outcome is presented as a secondary endpoint and shows an advantage. However, this scoring system might overestimate conclusions based on PFS findings in these clinical trials as they are statistically powered to only detect significant differences in PFS, not in OS. This trend toward the use of PFS as the primary endpoint in advanced NSCLC clinical trials has already been confirmed in a recent retrospective cohort study [14], where concerns about how clinical benefits are measured in this setting were also displayed.

To our knowledge, there is no other review on ESMO-MCBS v1.1 application conducted exclusively in advanced NSCLC. Broekman et al. [15] analyzed controversial therapeutic options in advanced-stage ovarian cancer. They could only apply the ESMO-MCB scale to 20% (11/55) of the studies included, in contrast to the 45% (42/92) in which we were able to apply the scale, but concluded that the ESMO threshold for clinical benefit should be considered when designing future clinical trials. Del Paggio et al. [16] reached the same conclusion when evaluating 226 RCTs published between 2011 and 2015 in different cancer types, including NSCLC. They could apply the ESMO-MCBS to 50% of their total cohort and also assessed the proportion of trials that met the ESMO-MCBS threshold for clinical benefit using both the lower limit of 95% CI of the HR and the point estimate. They found that the percentage of trials meeting that threshold was 31% and decreased by 6% when the point estimate was used [16]. In our cohort, 54.8% (23/42) of trials meet the threshold for meaningful benefit but the comparison with the point estimate HR scores also reduces that percentage, in this case to 40.5%, meaning a difference of 14.3%. These findings suggest that the ESMO-MCBS v1.1 could be somewhat permissive when using the lower limit of 95% CI of the HR instead of the point estimate HR.

An increasing debate has been emerging on the validity and reproducibility of the ESMO-MCBS v1.1 [17–19]; toxicity grade adjustments, for example, might be confusing. The ESMO-MCBS v1.1 only applies a toxicity penalty when the primary endpoint and thus 'scoreable' outcome is PFS, and only for high-grade adverse events that compromise global QoL. These apparent differences in toxicity penalties within each grading form constitute one of the unresolved criticisms of the ESMO-MCBS framework [17]. The opportunity exists for the ESMO-MCBS Working Group to consider the

Table 4. Trials showing discrepancies in applying the ESMO framework when compared with the ESMO-MCBS scorecards and reasons for discrepancy.

Analyzed treatment(s)	Setting	Primary outcome (PO)	PO HR (95%CI)	Toxicity adjustment/comment	Final score and form	ESMO-MCBS scorecard value	Reason(s) for discrepancy
Afatinib	First-line, nSQ	PFS	0.58 (0.43–0.78)	Diarrhea G3-4 (0 vs 14.4%); rash G3-4 (0 vs 16.2%)	4 (form 2 b)	4	Analyzed studies
Crizotinib	First-line, nSQ	PFS (OS improved)	0.35 (0.08–0.72)	Improved QoL (adj 1)	5 (form 2a)	4	Analyzed studies and toxicity adjustment
Bevacizumab + Erlotinib	First-line, nSQ	PFS	0.54 (0.36–0.79)	No OS advantage and not improved QoL (adj c)	2 (form 2 b)	3	Analyzed studies and QoL adjustment
Pembrolizumab	First-line, SQ and nSQ	PFS	0.63 (0.47–0.86)	EQ-5D-3L VAS differences show no statistical significance	4 (form 2a)	5	Toxicity/QoL adjustment
Nivolumab	Second or subsequent line, SQ	OS	0.62 (0.47–0.80)	Improved QoL	5 (form 2a)	5	Analyzed studies
Nivolumab	Second or subsequent line, nSQ	OS	0.75 (0.63–0.91)	Improved QoL	5 (form 2a)	5	Analyzed studies
Crizotinib	Second or subsequent line, SQ and nSQ	PFS	0.49 (0.37–0.64)	Improved QoL	4 (form 2 b)	4	Analyzed studies
DCX + Ramucirumab	Second or subsequent line, SQ and nSQ	OS	0.86 (0.75–0.98)	NA	1 (form 2a)	1	Analyzed studies
Pembrolizumab 10 mg/Kg	Second or subsequent line, SQ and nSQ, PDL1 $\geq$ 1%	OS	0.61 (0.49–0.75)	No statistically significant differences observed in the adverse events included on the ESMO-MCBS form	4 (form 2a)	5	Toxicity adjustment
Osimertinib	Second or subsequent line, SQ and nSQ	PFS	0.30 (0.23–0.41)	Improved QoL	4 (form 2 b)	4	Analyzed studies
Lorlatinib	Second or subsequent line, SQ and nSQ	ORR	NA	Selected cohort included EXP2-5 groups	3 (form 3)	3	Cohort selection

NA: not applicable; adj: adjustment; QoL: quality of life; PFS: progression-free survival; OS: overall survival; HR: hazard ratio; MCBS: magnitude of clinical benefit scale; VAS: visual analogue scale.

introduction of toxicity penalties in form 2a; based on OS findings.

As a result of the aforementioned data, discrepancies between scores were found to affect a notable proportion of our cohort (26.2%); however, most of the reasons for discrepancy do not lead to a different final ESMO-MCBS score. Differences in the analyzed studies for scoring can be amended by updating the database, considerably reducing the discrepancy rate observed. Disparities in toxicity or QoL adjustments occur when evaluating pembrolizumab in two different trials. In KEYNOTE-024, no statistically significant differences are shown in QoL assessed by the validated questionnaire EQ-5D-3L visual analog scale (VAS), but two other scales are applied (QLQ-LC13 and QLQ-C30) and show an advantage, although the clinical significance of this advantage is not clear at all. In KEYNOTE-010, though adverse events grades 3 or higher differ between treatment arms, the statistical significance is not available within the publication, and the percentages include any adverse event of grade $\geq$3 but not only those affecting patients' daily well-being, as denoted in the ESMO-MCBS v1.1 forms. Further versions of the ESMO-MCBS that address these limitations in evaluating toxicity profiles are highly expected and desirable.

Although a substantial percentage of discrepancies were found when our scores were compared with the corresponding ESMO scorecards, discrepancies in the value of final scores were minimum. Naturally, the ESMO-MCB scale adds a useful tool for categorizing and processing clinical trial data of the examined drugs. Combined with pharmaceutical costs, the ESMO framework may help clinicians and regulatory authorities to select the most valuable therapeutic option among those competing drugs developed for the same clinical entity [20,21]. Furthermore, it should be considered in the statistical design of future RCTs [15,17] to ensure reaching the thresholds of meaningful clinical benefit and the maximum validity of research data.

A major caveat is that solely 54.8% of all the evaluated clinical trials achieved the ESMO thresholds for meaningful clinical benefit despite the statistically significant difference favoring the experimental arm they had shown. Thus, almost 1 out of 2 positive clinical trials is unable to demonstrate a substantial clinical benefit according to the ESMO framework. As mentioned above, only 5/42 (11.9%) of the clinical trials correspond to not commercialized drugs, leading to a high proportion of commercialized drugs that do not meet the ESMO-MCBS threshold for clinical benefit. Some authors have criticized this issue after evaluating market approvals for cancer drugs in recent periods of time [3,4,13].

One of the main limitations of reviews and meta-analyses is publication bias, a form of selection bias. However, we minimized it by applying an organized searching strategy where two researchers independently selected the studies conforming to predefined inclusion and exclusion criteria. To only assess new therapies that had shown enough efficacy and toxicity data, we discarded phase I and non-pivotal phase II RCTs. Thus, the RCTs included mainly represent commercialized drugs used in daily clinical practice.

Some other limitations might have affected our results. For example, we could not evaluate the effect of permitted cross-over on the OS rates in such studies, which might influence the final ESMO-MCBS grade and lead to suboptimal decisions [22]. Besides, the limited toxicity data available within a publication of a clinical trial prevent from properly adjusting the preliminary grades. In other cases, the lack of QoL estimations results in a less accurate evaluation of the magnitude of clinical benefit, as adjustments related to QoL data cannot be contemplated. We consider that no measure could be taken to minimize these sources of bias.

In conclusion, a great proportion of clinical trials, mostly on commercialized drugs, did not meet the ESMO thresholds for meaningful clinical benefit in our study. Despite the ESMO-MCBS v1.1 constitutes a useful and reproducible instrument for assessing the clinical benefit of drugs for advanced NSCLC, a more detailed approach to toxicity penalties that accounts for those persistent, low-grade adverse events is required, as well as an adapted scoring for those studies based on PFS, ensuring that the limitations of this endpoint, as a surrogate for improved OS, are duly expressed in the final scores.

Acknowledgments

This article is part of the Doctoral Thesis of Ricardo García Fumero, within the Doctoral Program in Pharmacy of the University of Granada (UGR), Spain.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ricardo García-Fumero http://orcid.org/0000-0003-2459-4440
Cristina Fernández-López http://orcid.org/0000-0002-4565-1283
Miguel Ángel Calleja-Hernández http://orcid.org/0000-0001-8449-5490
José Expósito-Hernández http://orcid.org/0000-0003-1591-5476

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
[2] American Cancer Society. Cancer facts & figures 2020. Am Cancer Soc. 2020;17–21.
[3] Naci H, Davis C, Savović J, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. BMJ. 2019;366:l5221.
[4] Hilal T, Gonzalez-Velez M, Prasad V. Limitations in clinical trials leading to anticancer drug approvals by the US food and drug administration. JAMA Intern Med. 2020;180(8):1108–1115.

[5] Hwang T, Ross J, Vokinger K, et al. Association between FDA and EMA expedited approval programs and therapeutic value of new medicines: retrospective cohort study. BMJ. 2020;371:m3434.
[6] Prasad V, De Jesús K, Mailankody S. The high price of anticancer drugs: Origins, implications, barriers, solutions. Nat Rev Clin Oncol. 2017;14(6):381–390.
[7] Goulart BHL. Value: the next frontier in cancer care. Oncologist. 2016;21(6):651–653.
[8] 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.
[9] Cherny NI, Dafni U, Bogaerts J, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. Ann Oncol. 2017;28(10):2340–2366.
[10] 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;33:2563–2577.
[11] 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):2925–2933.
[12] Kiesewetter B, Raderer M, Steger GG, et al. The European Society for Medical Oncology Magnitude of Clinical Benefit Scale in daily practice: a single institution, real-life experience at the Medical University of Vienna. ESMO Open. 2016;1(4):e000066.
[13] Davis C, Naci H, Gurpinar E, et al. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ. 2017;359:j4530.
[14] Fernández-López C, Calleja-Hernández MÁ, Balbino JE, et al. Trends in endpoint selection and result interpretation in advanced non-small cell lung cancer clinical trials published between 2000 and 2012: a retrospective cohort study. Thorac Cancer. 2019;10(4):904–908.
[15] Broekman KE, Jalving M, van Tinteren H, et al. Clinical benefit of controversial first line systemic therapies for advanced stage ovarian cancer – ESMO-MCBS scores. Cancer Treat Rev. 2018;69:233–242.
[16] Del Paggio JC, Azariah B, Sullivan R, et al. Do contemporary randomized controlled trials meet ESMO thresholds for meaningful clinical benefit. Ann Oncol. 2017;28(1):157–162.
[17] Del Paggio JC. Toxicity adjustment in the ESMO-MCBS: a gestalt approach? Ann Oncol. 2018;29(2):520–521.
[18] Wild C, Grössmann N, Bonanno PV, et al. Utilisation of the ESMO-MCBS in practice of HTA. Ann Oncol. 2016;27(11):2134–2136.
[19] Emprechtinger R, Grössmann N, Wild C. ESMO-MCBS v1.1: statistical and patient-relevant shortcomings. Ann Oncol. 2018;29(4):1070–1071.
[20] Giuliani J, Remo A, Bonetti A. The European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS) applied to pivotal phase III randomized-controlled trials of tyrosine kinase inhibitors in first-line for advanced non-small cell lung cancer with activating epidermal growth factor receptor mutations. Expert Rev Pharmacoecon Outcomes Res. 2017;17(1):5–8.
[21] Hammerman A, Greenberg-Dotan S, Feldhamer I, et al. The ESMO-Magnitude of Clinical Benefit Scale for novel oncology drugs: correspondence with three years of reimbursement decisions in Israel. Expert Rev Pharmacoecon Outcomes Res. 2018;18(1):119–122.
[22] Jönsson L, Sandin R, Ekman M, et al. Analyzing overall survival in randomized controlled trials with crossover and implications for economic evaluation. Value Heal. 2014;17(6):707–713.