ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL RESEARCH

Single-arm trials supporting the approval of anticancer medicinal products in the European Union: contextualization of trial results and observed clinical benefit

J. Mulder[1*], S. Teerenstra[1,2], P. B. van Hennik[1], A. M. G. Pasmooij[1], V. Stoyanova-Beninska[1], E. E. Voest[3,4] & A. de Boer[1,5]

[1]Dutch Medicines Evaluation Board, Utrecht; [2]Department for Health Evidence, Biostatistics Section, Radboud University Medical Center, Nijmegen; [3]The Netherlands Cancer Institute, Amsterdam; [4]Oncode Institute, Amsterdam; [5]Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Available online 11 April 2023

Background: Single-arm trials (SATs) can sometimes be used to support marketing authorization of anticancer medicinal products in the European Union. The level and durability of antitumor activity of the product as well as context are important aspects to determine the relevance of trial results. The aim of this study is to provide details on the contextualization of trial results and to evaluate the magnitude of benefit of medicinal products approved based on SATs.

Materials and methods: We focused on anticancer medicinal products for solid tumors approved on the basis of SAT results (2012-2021). Data were retrieved from European public assessment reports and/or published literature. The benefit of these medicinal products was evaluated via the European Society for Medical Oncology (ESMO)-Magnitude of Clinical Benefit Scale (MCBS).

Results: Eighteen medicinal products were approved based on 21 SATs—few medicinal products were supported by >1 SAT. For the majority of clinical trials, a clinically relevant treatment effect was (pre)specified (71.4%) and most often an accompanying sample size calculation was provided. For 10 studies, each testing a different medicinal product, a justification for the threshold for a clinically relevant treatment effect could be identified. At least 12 out of 18 applications included information to facilitate the contextualization of trial results, including six supportive studies. Of the pivotal SATs analyzed ($n = 21$), three were assigned an ESMO-MCBS score of 4, which corresponds to 'substantial' benefit.

Conclusions: The clinical relevance of the treatment effects shown by medicinal products for solid tumors tested in SATs is dependent on the effect size and context. To better facilitate regulatory decision making, prespecifying and motivating a clinically relevant effect and aligning the sample size to that effect is important. External controls may facilitate in the contextualization process, but the associated limitations must be addressed.

Key words: single-arm trials, oncology, European Medicines Agency, clinical benefit, contextualization

INTRODUCTION

Randomized controlled trials (RCTs) are referred to as the 'gold standard' in testing medicinal products.[1] These trials have several advantages over clinical trials with other designs due to their design features. For example, randomization facilitates subjects in the experimental and control groups being comparable at baseline. Randomization and blinding are useful techniques to determine whether there is a cause—effect relation between treatment and outcome.[2,3] RCTs are the preferred trials to be included in applications for marketing authorization, as laid down in Directive 2001/83/EC. In this directive, it is stated that clinical trials relevant to the indication "shall be done as 'controlled clinical trials' if possible, randomised; any other design shall be justified".[4] Yet, it is not always possible to conduct an RCT, and, consequently, clinical trials with other designs need to be considered for registrational purposes.[5] The latter includes the use of single-arm trials (SATs).

Tenhunen et al. identified that, between 2010 and 2019, the European Commission (EC) approved 22 medicinal products for the treatment of solid tumors or hematological malignancies on the basis of SAT results.[6] Many of the medicinal products included in their study received 'conditional marketing authorization' (CMA).[6] This type of approval was

*Correspondence to: Dr Jorn Mulder, College ter Beoordeling van Geneesmiddelen, Medicines Evaluation Board, Graadt van Roggenweg 500, 3531 AH Utrecht, The Netherlands. Tel: +31-8-82-24-80-44

E-mail: j.mulder@cbg-meb.nl (J. Mulder).

introduced in the past to address an unmet medical need, and is based on less complete data than are usually required for standard approval.[7] It should be mentioned, however, that SATs can also support standard approvals—albeit less common. Examples are the approvals of engineered autologous T-cell immunotherapies.[8,9] However, demonstrating that an investigational medicinal product provides clinical benefit can be challenging when it is tested solely in an SAT. Trials like these are associated with different forms of bias, including selection bias.[10,11] Besides, surrogate endpoints such as objective response rate (ORR) are commonly used in SATs, at least when focusing on cancer research.[12,13] ORR is not a direct measure of clinical benefit. Yet, it is a measure of (antitumor) activity, as spontaneous regression occurs infrequently in cancer.[13]

Some guidance exists on the use of SATs for regulatory purposes. It is stated in the "Guideline on the clinical evaluation of anticancer medicinal products" of the European Medicines Agency (EMA) that resorting to a non-randomized design should be justified by, among others, a large treatment effect on ORR and duration of response (DoR), effects that will likely translate into clinical benefit.[5] Moreover, in the same guideline, it is stated that contextualization of results is an important topic for SATs, particularly for less evident cases.[5] Indirect comparisons with available therapies are often made for these purposes.[14,15] While it is not the task of regulatory agencies to ensure comparative efficacy,[16] there is a general need to ensure that new medicinal products are not worse—in terms of efficacy and/or safety—than standard of care. Importantly, the aspects described above, such as the size and durability of the treatment effect and context, will help to determine the clinical relevance of trial results.

The aim of this study was to provide details on how clinical benefit of anticancer medicinal products tested in SATs was determined, including the methods used to contextualize the trial results. In addition, we were interested in how many of the authorized medicinal products based on SATs showed 'substantial' benefit. We started with investigating whether a threshold for the relevant treatment effect was (pre)specified in the pivotal trials—for example, in a power calculation. Subsequently, we determined if applicants submitted additional evidence to contextualize the SAT results. Finally, by limiting this study to medicinal products for the treatment of solid tumors, we evaluated the magnitude of benefit of the medicinal products included in our study via a validated tool, the European Society for Medical Oncology (ESMO)-Magnitude of Clinical Benefit Scale (MCBS).

MATERIALS AND METHODS

Medicinal products

An overview of all human medicines that were granted approval by the EC was retrieved from the EMA database (https://www.ema.europa.eu/en/medicines). Products were identified on the basis of their Anatomic Therapeutic Chemical (ATC) codes, that is, L01-04 for antineoplastic and immunomodulating agents. We focused on medicinal products for the treatment of solid tumors authorized between 2012 and 2021—a 10-year period. The inclusion criterion for our analysis was initial approvals based on an SAT(s). Approvals based on RCTs were excluded. Approvals of generic and biosimilar products were also excluded.

Data sources

The main data source was the European public assessment reports (EPARs). These reports were obtained from the EMA database (https://www.ema.europa.eu/en/medicines). EPARs contain information on the scientific evaluation conducted by the Committee for Medicinal Products for Human Use (CHMP)—a committee of the EMA. The scientific evaluation forms the basis for the EC decision on approval. Another data source was published literature on pivotal clinical trials. Relevant publications were identified via PubMed and/or ClinicalTrials.gov.

Data collection

Data were retrieved from EPARs and/or scientific publications. We focused on pivotal trials, meaning that clinical pharmacology and dose-finding studies were not included. We collected the following information on the pivotal trials: the study design, dosing regimen, study population, planned sample size, statistical methods, primary/secondary endpoints, clinical outcomes, and type of authorization. It was also determined whether applicants made additional efforts to contextualize the results of the SAT(s), i.e. the use of external evidence to facilitate the interpretation of trial results. This concerned analyses (e.g. within-patient analysis) and/or evidence such as publications and additional studies that were included in the EPAR as supportive evidence. In addition to EPARs, scientific publications, including publicly available protocols that were supplementary to these publications, were used to complement information on the statistical methods.

Determining clinical benefit

The ESMO created the ESMO-MCBS, a validated tool to evaluate the magnitude of clinical benefit.[17] The ESMO-MCBS scores already assigned to clinical trials (i.e. ESMO publications or EMSO-MCBS scorecards) were identified. The remaining SATs included in our analysis were assigned an ESMO-MCBS score independently by two researchers (VSB and JM). This was done according to EMSO instructions.[18] Scientific publications were used for this purpose. In case a CMA was converted to standard marketing authorization (SMA) at the time of data analysis, an ESMO-MCBS score was assigned to the confirmatory trial. For non-curative therapies, ESMO-MCBS scores $\geq$4 represent substantial benefit.[19]

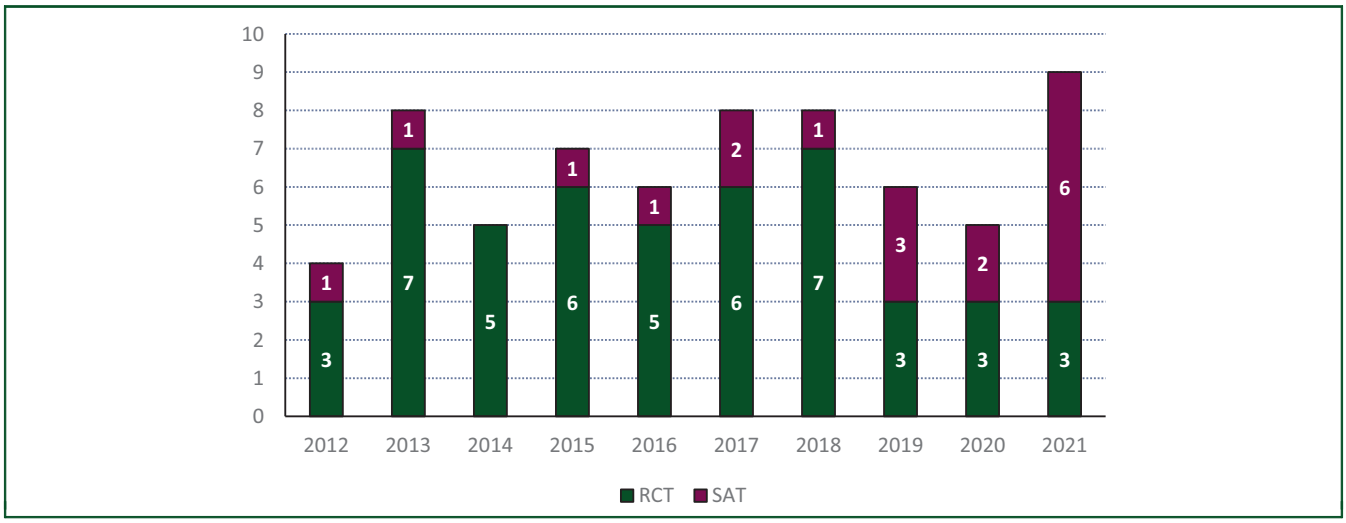

Figure 1. Number of medicinal products for the treatment for solid tumors approved by the European Commission per year. Generic and biosimilar medicinal products were excluded. In purple the number of approvals based solely on single-arm trials (SATs) and in green the number of approvals based on randomized controlled trials (RCTs).

RESULTS

Approval of medicinal products for the treatment of solid tumors

A total of 731 medicinal products received EC approval between 2012 and 2021. Of these, 66 (9.0%) were granted approval for the treatment of solid tumors—excluding generics or biosimilars (Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2023.101209). Over the recent years, the proportion of approvals for solid tumors based on SATs increased compared to prior years (Figure 1). In total, 18 (2.5%) medicinal products were approved based on 21 SATs (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2023.101209). Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2023.101209, shows the intended patient populations for which the medicinal products were approved. Half of the medicinal products were approved (also) for the treatment of advanced non-small-cell lung cancer (NSCLC). The approvals of alectinib, avapritinib, and crizotinib were on the basis of results from an SAT(s) with top-line results from an RCT—albeit not always in a similar treatment setting (e.g. different line of therapy). However, as the SATs remained the pivotal trial(s) supporting these applications, the three products were retained in our analyses.

All 18 medicinal products approved based on an SAT(s) were granted CMA. At the time of data analysis, eight CMAs were converted to SMAs (Supplementary Table S3, available at https://doi.org/10.1016/j.esmoop.2023.101209). For one of the CMAs, i.e. rucaparib, the benefit—risk balance was no longer considered favorable by the CHMP based on the confirmatory trial. The marketing authorization holder (MAH) requested to remove the indication.

Single-arm trials and thresholds for clinically relevant treatment effect

Most approvals were supported by one pivotal trial. The approvals of alectinib, osimertinib, and rucaparib were supported by two SATs. For the approvals of entrectinib and larotrectinib, integrated analyses by pooling data across clinical trials were used for the evaluation of efficacy (three trials each). For all trials or integrated analyses, the primary endpoint was ORR (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2023.101209).

For the majority of clinical trials or integrated analyses [15 out of 21 (71.4%)], a clinically relevant treatment effect was (pre)specified and most often an accompanying sample size calculation was provided (Table 1). The test for a relevant effect was often defined as the lower bound of the 95% confidence interval (CI) for ORR exceeding a (predefined) value, which is equivalent to testing a null hypothesis corresponding to that value. Protocols confirmed the results, which were publicly available (i.e. supplementary to publication) for all SATs except for those studying alectinib (both trials), avelumab, osimertinib (AURA 2), and rucaparib (CO-338-010). For the trials investigating entrectinib, larotrectinib, pemigatinib, and selpercatinib, a clinically relevant lower boundary of the 95% CI for ORR was defined, but the null and alternative hypotheses were not explicitly mentioned in the EPARs/publications. For the trials testing ceritinib, crizotinib, lorlatinib, and rucaparib, no power calculations were carried out based on the information presented in the EPARs and/or publications. Regarding trial CO-338-017, one of the SATs testing rucaparib, some sample size assumptions were made for subgroup allocation (part 1 and 2) and comparison (part 1) of the trial, but no calculations were made based on expected treatment effects.

For 10 out of 21 trials (47.6%), each testing a different medicinal product, justification for the threshold to the statistical test could be extracted from EPARs/publications/protocols (Table 1). Mostly, the treatment effect of available therapies was used as a benchmark ($n = 5$). Other justifications were 'consistent with the response rates seen with approved targeted therapies in genetically defined patient populations who have progressed on prior therapies'

Table 1. Statistical aspects of SATs

Medicinal product	Trial(s)	Therapeutic area	Biomarker-based indication	Available therapies in treatment setting	Sample size calculations[a]	Lower bound of the 95% CI for ORR to be ruled out	Justification	Ref.
Alectinib	NP28761	Lung cancer	Yes	Chemotherapy	Yes	35%	Not provided	20
	NP28673	Lung cancer		Chemotherapy	Yes	35%	Not provided	21
Amivantamab	EDI1001	Lung cancer	Yes	Chemotherapy or immunotherapy	Yes	12%	Single-agent chemotherapy as the benchmark	22
Avapritinib	BLU-285-1101	Sarcoma	Yes	Tyrosine kinase inhibitors	Yes	10%	Benchmarked against available therapies	23
Avelumab	EMR100070-003	Skin cancer	No	Chemotherapy	Yes	20%	Absence of literature documenting treatment outcomes for second-line patients	24
Cemiplimab	2810-ONC-1540	Skin cancer	No	EGFR inhibitors and/or chemotherapy	Yes	laCSCC 25%	Based on previous studies	25
					Yes	mCSCC 15%	Based on previous studies	26
Ceritinib	CLDK378X2101	Lung cancer	Yes	Chemotherapy	No	Not specified	Not applicable	27
Crizotinib	A8081001	Lung cancer	Yes	Chemotherapy	No	Not specified	Not applicable	28
Dostarlimab	4010-01-001	Endometrial cancer	Yes	Chemotherapy or bevacizumab	Yes	20%	Expected ORR for conventional therapy	29
Entrectinib[b]	ALKA-372-001, RXDX-101-01, and RXDX-101-03	Lung cancer	Yes	Crizotinib	Yes (on precision and implicitly on power)	50%	Observed with standard-of-care ROS1 fusion-positive NSCLC treatment	30
		Cancer		No appropriate available therapies	Yes (on precision and implicitly on power)	30%	Not provided	31
Larotrectinib[b]	LOXO-TRK-14001, LOXO-TRK-15002, and LOXO-TRK-15003	Cancer	Yes	No appropriate available therapies	Yes	30%	Consistent with the response rates seen with approved targeted therapies in genetically defined patient populations who have progressed on prior therapies	32
Lorlatinib	B7461001	Lung cancer	Yes	(Platinum-based) chemotherapy/immunotherapy	No	Not specified	Not applicable	33
Osimertinib[c]	AURA extension	Lung cancer	Yes	(Platinum-based) chemotherapy or tyrosine kinase inhibitor rechallenge	Yes (based on precision)	Not specified	Not applicable	34
	AURA 2			(Platinum-based) chemotherapy or tyrosine kinase inhibitor rechallenge	Yes (based on precision)	Not specified	Not applicable	35
Pemigatinib	INCB 54828-202	Bile duct cancer	Yes	Chemotherapy	Yes	15%	Proportions of patients with an objective response reported by previous studies	36
Pralsetinib	BLU-667-1101	Lung cancer	Yes	(Platinum-based) cytotoxic chemotherapy and/or immunotherapy	Yes	48%	Not provided	37
				Chemotherapy ± ramucirumab or immunotherapy	Yes	23%	Not provided	
Rucaparib[b]	CO-338-010	Ovarian cancer	Yes	Chemotherapy	No	Not specified	Not applicable	38
	CO-338-017			Chemotherapy	No	Not specified	Not applicable	39

Continued

Table 1. Continued

Medicinal product	Trial(s)	Therapeutic area	Biomarker-based indication	Available therapies in treatment setting	Sample size calculations[a]	Lower bound of the 95% CI for ORR to be ruled out	Justification	Ref.
Selpercatinib	LOXO-RET-17001	Lung cancer	Yes	Chemotherapy ± ramucirumab or immunotherapy	Yes	NSCLC 30%	Consistent with the response rates seen with approved targeted therapies in molecularly defined populations who failed prior therapies	40
		Thyroid cancer		Treatment options in these settings are limited—tyrosine kinase inhibitors rechallenge—or even lacking	Yes	MTC 20%	The limited treatment options	41
					No	TC Not specified	Not applicable	41
Trastuzumab deruxtecan	DS8201-A-U201	Breast cancer	Yes	HER2-targeted therapy in combination with chemotherapy	Yes	20%	Not provided	42
Vismodegib	SHH4476g	Skin cancer	No	Radiation therapy or chemotherapy	Yes	mBCC 10%	No therapeutic options exist for these patients and spontaneous responses have not been reported in this disease[d]	43
					Yes	aBCC 20%	No therapeutic options exist for these patients and spontaneous responses have not been reported in this disease[d]	43

Therapeutic areas are depicted in color: lung cancer in orange, skin cancer in yellow, cancer (general) in blue, and remaining areas in green. Information was retrieved from EPARs and complemented by scientific publications and protocols, if available and necessary.

aBCC, advanced basal cell carcinoma; CI, confidence interval; HER2, human epidermal growth factor receptor 2; laCSCC, locally advanced cutaneous squamous cell carcinoma; mBCC, metastatic basal cell carcinoma; mCSCC, metastatic cutaneous squamous cell carcinoma; MAH, marketing authorization holder; MTC, medullary thyroid cancer; NSCLC, non-small-cell lung cancer; ORR, objective response rate; ROS1, c-ros oncogene 1; TC, thyroid cancer.

[a]Sample size calculations were based on power unless otherwise specified.

[b]Integrated analysis was carried out based on two or three trials.

[c]Separated and integrated analysis was carried out for trials AURA and AURA2.

[d]Based on protocol.

https://doi.org/10.1016/j.esmoop.2023.101209

Table 2. Information provided by applicants to contextualize SAT results

Medicinal product	Information included in the European public assessment report
Amivantamab	Supportive study 61186372NSC100
Avapritinib	A comparison of trial versus natural history data Supportive study BLU-285-1002
Avelumab	Best response on the last prior anticancer drug therapy for metastatic disease Supportive study 100070-Obs001
Cemiplimab	Supportive study Dermatologic Cooperative Oncology Group
Crizotinib	Indirect comparison versus other treatment[a] Results to previous treatment
Dostarlimab	Best overall response from last platinum-containing prior anticancer therapy
Entrectinib	Supportive study WO40977
Larotrectinib	Comparison of larotrectinib with available systemic treatment for cancer
Lorlatinib	A comparison between time to tumor progression on lorlatinib and the time to tumor progression on last treatment before lorlatinib
Pemigatinib	An analysis of second-line treatment
Rucaparib	Results from prospective studies in platinum-sensitive disease that included third-line treatment
Trastuzumab deruxtecan	A literature-based analysis to understand the historical context Supportive study Unicancer

[a]Data of the indirect comparison were not shown in the EPAR.

($n = 2$), 'limited treatment options' ($n = 1$), and 'absence of literature documenting treatment outcomes for second-line patients' ($n = 1$).

Pralsetinib and selpercatinib were tested in trials that included patients with RET fusion-positive NSCLC who previously received platinum-based chemotherapy. The specified clinically relevant lower bound of the 95% CI for ORR was different between the two trials, namely 23% and 30%, respectively (Table 1). Larotrectinib and entrectinib were tested in clinical trials that included patients with *NTRK* gene fusion-positive tumors. For both applications, the lower bound of the 95% CI for ORR was 30% for the integrated analysis across clinical trials (Table 1).

Contextualization

The type and amount of information that was included for contextualization purposes varied between the applications for marketing authorization for the 18 medicinal products. At least 12 out of 18 applications (71.4%) included some additional information for contextualization purposes (Table 2, Supplementary Table S4, available at https://doi.org/10.1016/j.esmoop.2023.101209). Six out of 18 applications included supportive studies (Table 2, Supplementary Table S4, available at https://doi.org/10.1016/j.esmoop.2023.101209). One of these supportive studies concerned a bibliographic reference, namely the Dermatologic Cooperative Oncology Group (DeCOG) study. Other supportive studies were of a retrospective nature, and included real-world data from various sources. From the supportive studies included in the applications of trastuzumab deruxtecan and entrectinib, i.e. the Unicancer study and WO40977, respectively, matched populations were generated. In the latter study, a comparative analysis with a matched crizotinib arm derived from real-world data was conducted.

Evaluating benefit

Figure 2 and Supplementary Table S5, available at https://doi.org/10.1016/j.esmoop.2023.101209, show the ESMO-MCBS scores for the pivotal SATs ($n = 21$), either assigned by us or already published by ESMO. For all the SATs included in our study, three SATs were assigned an ESMO-MCBS score of '4'. Fifteen SATs were assigned an EMSO-MCBS score of '3', two SATs were assigned an ESMO-MCBS score of '2', and one SAT was assigned an ESMO-MCBS score of '1'. ESMO-MCBS scores of '4' were assigned as a result of the score upgrades for quality of life (QoL), meaning the investigators reported improvements in QoL.

Five out of eight CMAs were converted to SMA based on an RCT, i.e. reaching a comprehensive level of evidence. Of these RCTs, four were assigned an ESMO-MCBS score of '4' and one was assigned a score of '2' (Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2023.101209).

DISCUSSION

In specific situations, medicinal products may receive (expedited) regulatory approval on the basis of results from SATs. In this study, we analyzed pivotal SAT-based applications for anticancer medicinal products in the European Union between 2012 and 2021. In this period, 18 medicinal products for the treatment of solid tumors received an approval based on 21 SATs. At least 12 out of 18 applications included additional information to contextualize the results from the pivotal trials, which included supportive studies, external evidence, information on response to prior therapy, and/or a within-patient comparison. Of all the SATs or integrated analyses supporting the 18 EC approvals, three were assigned an ESMO-MCBS score of '4', that is, a score indicating substantial benefit.

SATs are generally initiated to determine whether an investigational product has sufficient activity to continue development.[44,45] Often statistical testing is used to determine whether the treatment effect is above a prespecified threshold, which is reflected in whether the null hypothesis related to the threshold is rejected.[46] Our results indicate that a justification for this threshold was not always reported in the EPAR (or scientific publication). Tenhunen et al. reported that the threshold for 'success' in pivotal SATs is relatively uniform—20% ORR—and often not scientifically justified.[6] Our study does not confirm their results, as thresholds varied—ranging from 10% to 50%. This might, however, be explained by the partial differences in datasets. The threshold for success is often based on historical data or clinical judgment, which reflects ORRs by available treatment or standard of care.[47] However, determining this threshold can be challenging. For instance,

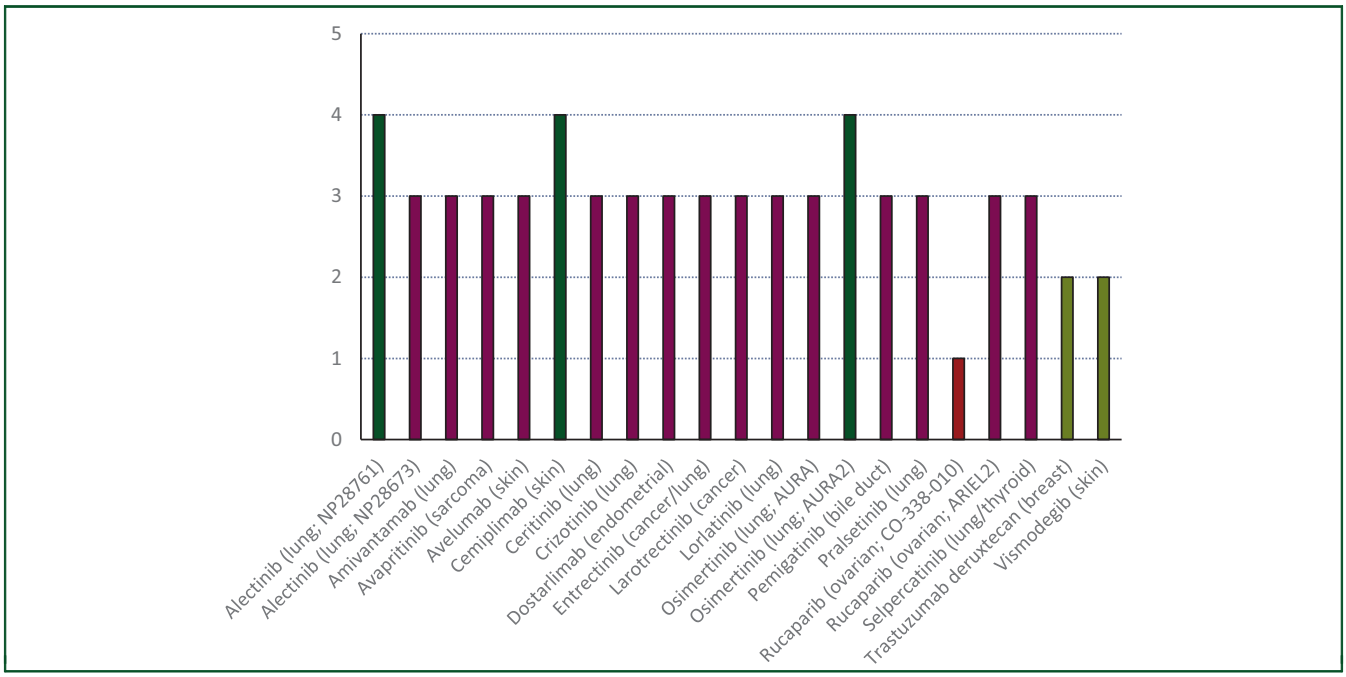

Figure 2. European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) scores assigned to pivotal single-arm trials. In dark green the trials are depicted that were assigned a high ESMO-MCBS score. Scores were either made publicly available by the ESMO or were assigned by us. Supplementary Table S3, available at https://doi.org/10.1016/j.esmoop.2023.101209, provides additional scoring details and reference to publications or scorecards for the scores assigned by ESMO.

historical data can be inconsistent with regard to the observed ORRs. Studies with doxorubicin plus ifosfamide in soft tissue sarcoma showed varied ORRs (i.e. 16%-35%).[48] Also, historical data might be absent or derived from studies that differ in, but not limited to, design or study population in comparison to the SAT.[48] The latter being particularly relevant for biomarker-driven SATs, which concerns the majority of SATs included in our study. Overall, it is important to select an appropriate threshold before conducting an SAT, but even more to provide argumentation why having a lower bound of the 95% CI above this threshold constitutes a clinically relevant outcome.

Our results show that the ORR to be ruled out at a particular significance level is not always ambitious. For example, thresholds based on historical data were sometimes lower than thresholds in absence of treatment options. Also, similar historical data sometimes led to different thresholds. Another point to mention is that the ORR used for sample size/power calculations, i.e. the effect under the alternative hypothesis, is rarely justified to be clinically relevant or corresponding to an effect one would not want to miss (data not shown), the latter, for instance, in the context of a go/no-go decision for proceeding with drug development program.[49] Simply rejecting the null hypothesis may not be sufficient for regulatory decision making. As already highlighted a few decades ago, the meaningfulness of ORR depends on whether this translates into 'true' benefit (e.g. improvement in survival).[50] Oxnard et al. showed that an ORR statistically exceeding 30% (or higher) is associated with regulatory approval, at least for monotherapies tested in SATs.[51] However, not only ORR but also DoR will be important for regulatory decision making. For example, the CHMP was of the opinion that the activity of pralsetinib, indicated by a high percentage of durable responses in the pivotal trial, would translate into clinical meaningful benefit.[52] The observed ORR of entrectinib shown in the integrated analysis was below the assumed ORR used for the sample size calculation. However, the observed ORR, in combination with DoR, was considered of clinical relevance by the CHMP.[53] In contrast, the ORR and DoR shown by retifanlimab in the pivotal SAT were not considered clinically relevant by the CHMP. In fact, the criterion for 'success' was not met in this SAT—i.e. ruling out an ORR of 13%—and the applicant withdrew the application for marketing authorization.[54] If it is justified to use an SAT for regulatory purposes, it will be key to motivate which effect would constitute an (minimal) important effect from a clinical point of view, not merely ruling out a, sometimes unimpressive, historical ORR.

During the approval process, context may be sought via indirect comparisons with (well-)documented outcomes for clinical trials testing available therapies. This is also relevant considering that new data may have become available after initiation of the SAT. We demonstrate that applications frequently include information for contextualization purposes, including results from supportive studies. There are, however, limitations associated with cross-trial comparisons,[55] which necessitate caution when interpreting these results. For example, differences between study populations may lead to inappropriate comparisons.[56] One approach to (partly) overcome these limitations is to use patient-level data to generate a matched external control.[56]

Interestingly, a recent study carried out by Schröder et al. demonstrated that external controls generated from electronic health record-derived databases were successful in replicating a control arm from an RCT in metastatic colorectal cancer.[57] However, matched comparisons with external controls are rare—at least in our dataset. Only two comparative matched analyses with standard of care were carried out. External controls, however, cannot be corrected for confounders that are unknown or unmeasured.[57] There is some regulatory guidance available to reduce potential bias with external controls.[2,5] However, after addressing all the limitations as much as possible, the issue remains that, if there is a high chance for residual bias, the outcome in an SAT has to be convincing to compensate for the potential bias. Importantly, the quality of data will likely determine the extent to which external controls can be used for regulatory decision making.[58]

Pignatti et al. highlighted that the definition of clinical value is different between stakeholders, which may lead to different conclusions.[59] While the CHMP concluded that the benefit of the medicinal products included in our analysis was clinically relevant, stakeholders other than regulators might appreciate benefit differently. For instance, the ESMO considers benefit as 'living longer and/or living better', which resonates in the ESMO-MCBS form for SATs.[19,17] This is evident by our results, as the benefit of the majority of products was 'modest' on the basis of the ESMO-MCBS scores. Tibau et al. stated that large treatment effects in combination with an improvement in QoL (or data from post-marketing studies) are needed for SATs to be assigned a high ESMO-MCBS score.[60] However, QoL is not always a secondary endpoint in clinical trials, and one of the shortcomings of the ESMO-MCBS is that it does not take into account delayed publications or publication bias for QoL.[61] Besides, the CHMP repeatedly stated in assessment reports that no firm conclusion can be drawn from QoL data generated by SATs.[62–65] Thus, QoL is of lesser importance in regulatory decision making on SATs.

There are other tools to evaluate the benefit of approved anticancer medicinal products. For instance, a committee of the Dutch Society of Medical Oncology created the PASKWIL criteria for non-randomized trials, for which the ESMO-MCBS was used as a basis.[66] In comparison to the ESMO-MCBS, QoL and safety are not incorporated in this instrument, and benefit is based on predefined ORR and DoR thresholds.[66] Other criteria are that the medicinal product is authorized by the EC, the disease is rare, the patient population is adequately selected, and there is a biological rationale for therapy.[67] As tools are created on a national level that do not completely align with the EMSO-MCBS, there might be a need to fine-tune what can be considered benefit on an European level. Consistency among tools may warrant further discussion among stakeholders so as to prevent potential inequality in care.

All medicinal products included in our study received a CMA. When the MAH intends to fulfill the specific obligation(s) associated with the CMA, the benefit—risk balance will be re-assessed on a more complete dataset, preferably results from an RCT. However, Tenhunen et al. showed that post-authorization measures associated with CMAs are not always to submit results from an RCT.[6] Of course, the level of evidence to be generated in the post-marketing setting depends on, amongst others, feasibility to conduct large trials. Recently, Fashoyin-Aje et al. informed that a 'comprehensive strategy' for confirmatory trials is needed, which focusses on the so-called on-ramp (e.g. trial design, patient population, etc.) and off-ramp considerations (i.e. verify clinical benefit).[67] The authors highlight that, for accelerated approvals, efforts should be made to timely and adequately address remaining uncertainties regarding the benefit—risk balance. Similarly, Bloem et al. highlight that RCTs should be ongoing when a CMA is granted, ensuring rapid access to a more complete dataset.[68] Important to mention is that re-assessment of the ESMO-MCBS score is possible when results from confirmatory trials are published. This may lead to an improvement in ESMO-MCBS score—as also seen in our study. Furthermore, extended follow-up for the SATs themselves may also improve the EMSO-MCBS score. For example, we previously assigned an ESMO-MCBS score of '2' to the SAT investigating cemiplimab.[69] However, our current research shows a score of '4' (from an ESMO-MCBS scorecard), which is based on a more recent publication.[70]

While this study provides insights into the contextualization process of SAT results, it is limited to SATs supporting initial approvals. While extensions of therapeutic indication(s) can in principle be based on SATs, this is rare and such applications are not included in our analysis. For an extension of indication, there is already existing knowledge on the benefits and risks of the concerned medicinal product due to the initial marketing authorization, which might impact decision making. In addition, we did not include withdrawals of SAT-based applications, as these numbers ($n = 4$) were too limited for a meaningful analysis. It can also be considered a limitation that we restricted our research to publicly available documents. However, we assume that all information relevant to the benefit—risk assessment is incorporated in the EPARs, as it is a reflection of the core documents included in an application, as well as in literature and/or protocols, the latter being available for most SATs. Another limitation is that confirmatory trials were ongoing for some of the products included in our study. The ESMO-MCBS score could, therefore, not yet be re-assessed for these products. Finally, we focused only on SAT-based applications submitted to the EMA. It would be interesting to compare regulatory decision making between agencies, such as the Food and Drug Administration and EMA.

In conclusion, we found that 18 medicinal products were approved for the treatment of solid tumors based on one or more SAT(s). For the majority of clinical trials or integrated analyses supporting these approvals, a threshold to be ruled out was (pre)specified, and most often accompanied by a sample size calculation based on an assumed ORR. However, a justification for the threshold and the assumed ORR could not be identified for all cases. The majority of

applications included additional information for contextualization purposes. Determining the benefit—risk balance of medicinal products tested in SATs is challenging and benefit can be appreciated differently by various stakeholders. The clinical relevance of the treatment effects shown by medicinal products tested in SATs is dependent on the activity, its durability and context, especially if other therapies are available that provide benefit. As general recommendations, prespecifying and motivating a clinically relevant effect and aligning the sample size to that effect is of importance for regulatory decision making. External controls may facilitate in the contextualization process, but the limitations associated with such comparisons must be (adequately) addressed. Preferably, such comparisons should be preplanned. It is of relevance that information on these aspects is presented in the EPAR, as this provides transparency on regulatory decision making toward stakeholders. Finally, it is considered of value to further discuss among stakeholders what can be considered clinical benefit in the context of SATs and thus when approval on the basis of lower levels of evidence is justified. This is considered of importance, as SATs will likely continue to form the basis of authorization of part of the new medicinal products.

FUNDING

None declared.

DISCLOSURE

The authors have declared no conflicts of interest.

DISCLAIMER

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of, or reflecting the position of the EMA or one of its committees, working parties, or any of the national agencies.

REFERENCES

1. Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet*. 2015;385:1502-1503.
2. ICH Harmonised Tripartite Guideline Choice of Control Group and Related Issues in Clinical Trials E10. Available at https://database.ich.org/sites/default/files/E10_Guideline.pdf. Accessed March 18, 2022.
3. Sabbald B, Roland M. Understanding controlled trials. Why are randomized controlled trials important? *BMJ*. 1998;316:201.
4. European Commission. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community Code Relating to Medicinal Products for Human Use. Available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/dir_2001_83_consol_2012/dir_2001_83_cons_2012_en.pdf. Accessed March 17, 2022.
5. European Medicines Agency. Guideline on the Clinical Evaluation of Anticancer Medicinal Products. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-evaluation-anticancer-medicinal-products-man-revision-6_en.pdf. Accessed March 8, 2022.
6. Tenhunen O, Lasch F, Schiel A, Turpeinen M. Single-arm clinical trials as pivotal evidence for cancer drug approvals: a retrospective cohort study of centralized European marketing authorization between 2010 and 2020. *Clin Pharmacol Ther.* 2020;108:653-660.
7. Commission Regulation (EC) No 507/2006 of 29 March 2006 on the Conditional Marketing Authorisation for Medicinal Products for Human Use Falling within the Scope of Regulation (EC) No 726/2004 of the European Parliament and of the Council. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R0507&from=IT. Accessed March 17, 2022.
8. European Medcines Evaluation. Kymriah Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/kymriah-epar-public-assessment-report_en.pdf. Accessed March 17, 2022.
9. European Medcines evaluation. Yescarta Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/yescarta-epar-public-assessment-report_en.pdf. Accessed March 17, 2023.
10. George SL. Selection bias, phase II trials, and the FDA accelerated approval process. *J Natl Cancer Inst*. 2003;95:1351-1352.
11. Glassman RH, Kim G, Kahn MJ. When are results of single-arm studies dramatic? *Nat Rev Clin Oncol*. 2020;17:651-652.
12. Seymour L, Ivy SP, Sargent D, et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clin Cancer Res.* 2010;16:1764-1769.
13. George SL. Response rate as an endpoint in clinical trial. *J Natl Cancer Inst*. 2007;99:98-99.
14. Carrigan G, Whipple S, Capra WB, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. *Clin Pharmacol Ther.* 2020;107:369-377.
15. Yap TA, Jacobs I, Andre EB, Lee LJ, Beaupre D, Azoulay L. Application of real-world data to external control groups in oncology clinical trial drug development. *Front Oncol*. 2022;11:695936.
16. DeLoughery EP, Prasad V. The US Food and Drug Administration's use of regular approval for cancer drugs based on single-arm studies: implications for subsequent evidence generation. *Ann Oncol*. 2018;29:527-529.
17. Cherny NI, Dafni U, Bogaerts J, et al. ESMO-magnitude of clinical benefit scale version 1.1. *Ann Oncol*. 2017;28:2340-2366.
18. European Society for Medical Oncology. ESMO-Magnitude of Clinical Benefit Scale V1.1 Instructions. Available at Version-1-1-Instructions. Available at https://www.esmo.org/content/download/117394/2059186/1/ESMO-MCBS-Version-1-1-Instructions.pdf. Accessed April 9, 2022.
19. 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;26:1547-1573.
20. Shaw AT, Gandhi L, Gadgeel S, et al. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. *Lancet Oncol*. 2016;17:234-242.
21. Ou SH, Ahn JS, De Petris L, et al. Alectinib in crizotinib-refractory ALK-rearranged non-small-cell lung cancer: a phase II global study. *J Clin Oncol*. 2016;34:661-668.
22. Park K, Haura EB, Leighl NB, et al. Amivantamab in EGFR exon 20 insertion-mutated non-small-cell lung cancer progressing on platinum chemotherapy: initial results from the CHRYSALIS phase I study. *J Clin Oncol*. 2021;39:3391-3402.
23. Heinrich MC, Jones RL, von Mehren M, et al. Avapritinib in advanced PDGFRA D842V-mutant gastrointestinal stromal tumour (NAVIGATOR): a multicentre, open-label, phase 1 trial. *Lancet Oncol*. 2022;21:935-946.
24. Kaufman HL, Russell J, Hamid O, et al. Avelumab in patients with chemotherapy-refractory metastatic merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. *Lancet Oncol*. 2016;17:1374-1385.
25. Migden MR, Khushalani NI, Chang ALS, et al. Cemiplimab in locally advanced cutaneous squamous cell carcinoma: results from an open-label, phase 2, single-arm trial. *Lancet Oncol*. 2020;21:294-305.

26. Migden MR, Rischin D, Schmults CD, et al. PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma. *N Engl J Med*. 2018;379:341-351.
27. Shaw AT, Kim DW, Mehra R, et al. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med*. 2014;370:1189-1197.
28. Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med*. 2010;363:1693-1703.
29. Oaknin A, Tinker AV, Gilbert L, et al. Clinical activity and safety of the anti-programmed death 1 monoclonal antibody dostarlimab for patients with recurrent or advanced mismatch repair-deficient endometrial cancer: a nonrandomized phase 1 clinical trial. *JAMA Oncol*. 2020;6:1766-1772.
30. Drilon A, Siena S, Dziadziuszko R, et al. Entrectinib in ROS1 fusion-positive non-small-cell lung cancer: integrated analysis of three phase 1-2 trials. *Lancet Oncol*. 2020;21:261-270.
31. Doebele RC, Drilon A, Paz-Ares L, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. *Lancet Oncol*. 2020;21:271-282.
32. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD. Efficacy of larotrectinib in TRK Fusion-positive cancers in adults and children. *N Engl J Med*. 2018;378:731-739.
33. Solomon BJ, Besse B, Bauer TM, et al. Lorlatinib in patients with ALK-positive non-small-cell lung cancer: results from a global phase 2 study. *Lancet Oncol*. 2018;19:1654-1667.
34. Jänne PA, Yang JCH, Kim DW, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med*. 2015;372:1689-1699.
35. Goss G, Tsai CM, Shepherd FA, et al. Osimertinib for pretreated EGFR Thr790MET-positive advanced non-small-cell lung cancer (AURA2): a multicentre, open-label, single-arm, phase 2 study. *Lancet Oncol*. 2016;17:1643-1652.
36. Abou-Alfa GK, Sahai V, Hollebecque A, et al. Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. *Lancet Oncol*. 2020;21:671-684.
37. Gainor JF, Curigliano G, Kim DW, et al. Pralsetinib for RET fusion-positive non-small-cell lung cancer (ARROW): a multi-cohort, open-label, phase 1/2 study. *Lancet Oncol*. 2021;22:959-969.
38. Kristeleit R, Shapiro GI, Burris HA, et al. A phase I-II study of the oral PARP inhibitor rucaparib in patients with germline BRCA1/2-mutated ovarian carcinoma or other solid tumors. *Clin Cancer Res*. 2017;23:4095-4106.
39. Swisher EM, Lin KK, Oza AM, et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol*. 2017;18:75-87.
40. Drilon A, Oxnard GR, Tan DSW, et al. Efficacy of selpercatinib in RET fusion-positive non-small-cell lung cancer. *N Engl J Med*. 2020;383:813-824.
41. Wirth LJ, Sherman E, Robinson B, et al. Efficacy of selpercatinib in RET-altered thyroid cancer. *N Engl J Med*. 2020;383:825-835.
42. Modi S, Saura C, Yamashita T, et al. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. *N Engl J Med*. 2020;382:610-621.
43. Sekulic A, Migden MR, Oro AE, et al. Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *N Engl J Med*. 2021;366:2171-2179.
44. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials*. 1989;10:1-10.
45. Mariani L, Marubini E. Design and analysis of phase II cancer trials: a review of statistical methods and guidelines for medical researchers. *Int Stat Rev*. 1996;64:61-88.
46. Khan I, Sarker S-J, Hackshaw A. Smaller sample sizes for phase II trials based on exact tests with actual error rates by rading-off their nominal levels of significance and power. *Br J Cancer*. 2012;20:1801-1809.
47. Wason JM, Jaki T. A review of statistical designs for improving the efficiency of phase II studies in oncology. *Stat Methods Med Res*. 2016;25:1010-1021.
48. Taylor JMG, Braun TM, Li Z. Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm Phase II design. *Clin Trials*. 2006;3:335-348.
49. Senn S. Minimally Important Differences: Definitions, Ambiguities and Pitfalls. Available at https://www.ideal.rwth-aachen.de/wp-content/uploads/2014/02/Minimally-Important-Differences-v2.pdf.
50. Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treat Rep*. 1985;69:1375-1381.
51. Oxnard GR, Wilcox KH, Gonen M, Polotsky M, Hirsch BR, Schwartz LH. Response rate as a regulatory end point in single-arm studies of advanced solid tumors. *JAMA Oncol*. 2016;2:772-779.
52. European Medicines Agency. Gavreto Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/gavreto-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
53. European Medicines Agency. Rozlytrek Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/rozlytrek-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
54. European Medicines Agency. Zynyz Withdrawal Assessment Report. Available at https://www.ema.europa.eu/en/documents/withdrawal-report/withdrawal-assessment-report-zynyz_en.pdf. Accessed January 9, 2023.
55. Ray EM, Carey LA, Reeder-Hayes KE. Leveraging existing data to contextualize phase II clinical trial findings in oncology. *Ann Oncol*. 2020;31:1591-1593.
56. Rahman R, Ventz S, McDunn J, et al. Leveraging external data in the design and analysis of clinical trials in neuro-oncology. *Lancet Oncol*. 2021;22:e456-e465.
57. Schröder C, Lawrance M, Li C, et al. Building external control arms from patient-level electronic health record data to replicate the randomized IMblaze370 control arm in metastatic colorectal cancer. *JCO Clin Cancer Inform*. 2021;5:450-458.
58. Mishra-Kalyani PS, Amiri Kordestani L, Rivera DR, et al. External control arms in oncology: current use and future directions. *Ann Oncol*. 2022;33:376-383.
59. Pignatti F, Wilking U, Postmus D, Wilking N, Delgado J, Bergh J. The value of anticancer drugs — a regulatory view. *Nat Rev Clin Oncol*. 2022;19:207-215.
60. Tibau A, Molto C, Borrell M, et al. Magnitude of clinical benefit of cancer drugs approved by the US Food and Drug Administration based on single-arm trials. *JAMA Oncol*. 2018;4:1610-1611.
61. Gyawali B, de Vries EGE, Dafni U, et al. Biases in study design, implementation, and data analysis that distort the appraisal of clinical benefit and ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS) scoring. *ESMO Open*. 2021;6:100117.
62. European Medicines Agency. Libtayo Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/libtayo-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
63. European Medicines Ageny. Tagrisso Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/tagrisso-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
64. European Medicines Agency. Vitrakvi Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/vitrakvi-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
65. European Medicines Agency. Pemazyre Assessment Report. Available at https://www.ema.europa.eu/en/documents/assessment-report/pemazyre-epar-public-assessment-report_en.pdf. Accessed January 9, 2023.
66. BOM. Voorstel PASKWIL-Criteria Niet-Gerandomiseerde Studies. Available at https://medischeoncologie.nl/artikelen/2021/mei/editie-4/we-hebben-de-lat-hoog-gelegd. Accessed January 9, 2023.
67. Fashoyin-Aje LA, Mehta GU, Beaver JA, Pazdur R. The on- and off-ramps of oncology accelerated approval. *N Engl J Med*. 2022;387:1439-1442.
68. Bloem LT, Bot RE, Mantel-Teeuwisse AK, et al. Pre-approval and post-approval availability of evidence and clinical benefit of conditionally approved cancer drugs in Europe: a comparison with standard approved cancer drugs. *Br J Clin Pharmacol*. 2022;88:2169-2179.

69. Mulder J, Pasmooij AMG, Stoyanova-Beninska V, Schellens JHM. Breakthrough therapy-designated oncology drugs: are they rightfully criticized? *Drug Discovery Today*. 2020;25:1580-1584.

70. Cemiplimab ESMO-MCBS Scorecards. Available at https://www.esmo.org/guidelines/esmo-mcbs/esmo-mcbs-scorecards/scorecard-360-1. Accessed January 9, 2023.