ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Health Policy Analysis

A Comparison of 7 Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies

Ashley Jaksa, MPH, Anthony Louder, PhD, Christina Maksymiuk, PhD, Gerard T. Vondeling, MSc, Laura Martin, MS, Nicolle Gatto, PhD, MPH, Eric Richards, MPH, MSc, Antoine Yver, MD, MSc, Mats Rosenlund, PhD, MPH

ABSTRACT

Objectives: The development of accelerated approval programs for high morbidity and unmet need conditions has driven the use of single-arm studies in drug development. Regulatory and health technology assessment (HTA) agencies are recognizing that high-quality external control arms (ECAs), built using real-world data, can reduce uncertainties arising from single-arm studies. This review compared 7 case studies of regulatory and HTA agencies' evaluations of oncology ECAs.

Methods: Food and Drug Administration multidisciplinary reviews for oncology submissions from 2014 to 2021 were screened to identify 7 cases (2 blinatumomab indications, avelumab, and erdafitinib, entrectinib, trastuzumab deruxtecan, and idecabtagene vicleucel) with ECAs to support efficacy claims. Regulatory (Food and Drug Administration, European Medicines Agency, Health Canada) and HTA (pan-Canadian Oncology Drug Review, National Institute for Health and Care Excellence, Federal Joint Committee, Haute Autorité de Santé, and Pharmaceutical Benefits Advisory Committee) submissions for these cases were reviewed. The decision makers' ECA critiques and the level of influence on the decision were analyzed and categorized.

Results: Across case studies, selection bias and confounding were the most common ECA critiques. Nevertheless, agreement in critiques between and among regulators and HTA bodies was low. ECA influence on agencies' decisions also varied.

Conclusions: Evaluating the same ECA evidence, agencies focused on methodologic issues (ie, selection bias and confounding), but were often not aligned on their critiques. Further research is needed to fully characterize how agencies evaluate ECAs. This study is a first step in critically evaluating agencies' critiques of ECAs and highlights the need for future guidance development around ECA design and generation.

Keywords: external control arms, health technology assessment, oncology, real-world evidence, regulatory approval.

VALUE HEALTH. 2022; 25(12):1967–1976

Introduction

Accelerated approval programs for high morbidity and high unmet need diseases have driven the use of single-arm studies, studies that do not include placebo or active comparator arms (ie, no concurrent control), for drug development. Single-arm studies speed up patient access to innovative treatments because they often require fewer patients than randomized controlled trials (RCTs) and use intermediate or surrogate endpoints (eg, objective response rate [ORR]). Oncology drug development has been increasingly relying on single-arm studies: from 1992 to 2017, 67% of the Food and Drug Administration's (FDA) accelerated approvals were based on single-arm trials.[1] Similar trends have been observed in health technology assessment (HTA) submissions. From 2000 to 2016, 22 submissions to the National Institute for Health and Care Excellence (NICE) in the UK were based on nonrandomized data, and oncology drugs accounted for more than half of these submissions.[2]

A consequence of using single-arm studies is generally less evidence on the therapeutic benefits of the product at market launch. In the absence of the preferred evidence standard (ie, large-scale RCTs with a true state of clinical equipoise) in submission packages, real-world evidence (RWE)—or evidence generated from real-world data (RWD)—is increasingly being considered by decision makers. External control arms (ECAs) generated from RWD, or "data relating the patient health status and/or the delivery of healthcare routinely collected"[3] from insurance claims, electronic health records (EHRs), registries, etc, are emerging to contextualize single-arm trial data by exploring what would happen if single-arm trial patients did not receive the study drug. An ECA is "an umbrella term referring to any control that is not a randomized control" and includes control data collected concurrently or historically from previous clinical trials or RWD.[4] There are numerous challenges to decision makers' adoption of RWE, especially in the context of ECAs, as a trusted source of evidence. One main concern is the potential to introduce bias that is often reduced through randomization; for

example, 2 common methodological challenges are confounding, or systematic differences in patient characteristics across treatment groups that can affect effectiveness and selection bias, or individuals in a study differing systematically from the population of interest, leading to a systematic error in an association or outcome.

Using regulatory case studies, multiple groups have examined methodological challenges for designing ECAs and have suggested strategies to mitigate them.[4-9] Nevertheless, best practice guidance for ECA development from regulators and HTA agencies is lacking, making ECA design principles unclear and the ECA's routine acceptance by regulators and HTAs uncertain. Although the FDA and the European Medicines Agency (EMA) issued high-level recommendations for the characteristics of a persuasive ECA (eg, well-defined natural history, similarities between treatment and control arms, objective endpoint, and large effect size),[10,11] details on how these criteria are defined and evaluated by the agencies are unclear. With a lack of detailed guidance on ECA design and use, uncertainties around agency agreement on quality and what is considered submissible exist.

In the absence of clear recommendations, we can look to recent case studies to evaluate how regulatory and HTA agencies evaluated and accepted ECA evidence generated from RWD. The objective of this descriptive evaluation is to compare a sampling of regulatory and HTA evaluations of ECAs supporting oncology drug decision making. This case study approach can begin to help us understand how agencies evaluate ECA evidence in oncology submissions.

Methods

Case Study Selection

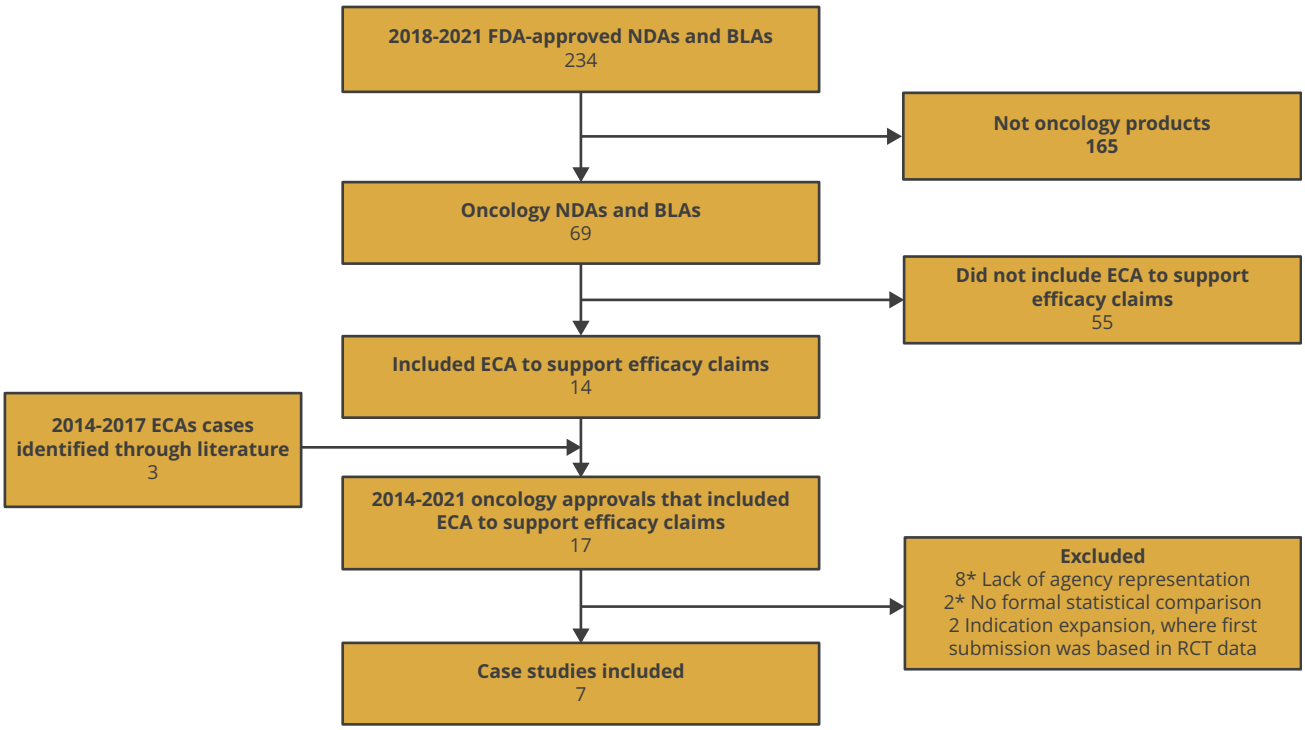
To identify FDA approvals that included RWE-based ECAs, we conducted an exhaustive search of FDA new drug approvals and biologic products licensed—made publicly available on Drugs@FDA website—between 2018 and 2021 (Fig. 1). The FDA does not post documentation on drug or biologic therapies that were not approved; thus, these nonapproved therapies could not be included. To identify relevant cases from 2014 to 2017, we identified peer-reviewed published articles that describe FDA approvals that included RWE-based ECA.[12-14] Once oncology ECAs submitted to the FDA were identified, we searched the respective agencies' website for the matching therapy/indication regulatory and HTA approval documentation in the European Union, Canada, and Australia. Regulatory agencies represented were the EMA and Health Canada (HC); HTA agencies represented were NICE (UK), Federal Joint Committee (G-BA) (Germany), Haute Autorité de Santé (HAS) (France), pan-Canadian Oncology Drug Review (pCODR) (Canada), and the Pharmaceutical Benefits Advisory Committee (PBAC) (Australia). Most source documents were available in English, others were translated.

FDA Multiple Disciplinary Reviews were evaluated to determine that the ECAs were used to support efficacy claims. Criteria used to select case studies included agency representation, formal statistical comparison between the ECA and single-arm study, diversity in tumor type, orphan designation, and year of approval. For indication expansions, therapies that did not have previous RCT evidence were prioritized. Seven case studies (from 6 therapies), with sufficient variety, were selected from 17 candidates (Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.05.016).

Document Search and Data Abstraction

Regulatory and HTA documents were evaluated using a prespecified data collection protocol and summarized by 2 authors (A.J. and C.M.). Discrepancies were discussed and resolved with a third author (A.L.), as necessary. The documents

Figure 1. Case study search process.



*not mutually exclusive

retrieved for the analysis are presented in Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.05.016. Key data elements were evaluated from each source included:

1. Date of review/assessment, decision, and rationale for decision
2. ECA population including inclusion/exclusion criteria
3. ECA primary endpoint and details
4. Statistical or methodological issues
5. ECA results and interpretation by the agency
6. ECA evidence classification (eg, sufficient evidence)

Qualitative Data Analysis

Although agencies make decisions based on the totality of evidence, some evidence sources can contribute more weight than others. For each regulatory and HTA decision, the ECA's level of influence was assessed by 2 reviewers (A.J. and C.M.) who evaluated the agencies' decision rationale. Two levels of influence were defined: high and low. A high level of influence indicated that the ECA was a major part of the clinical and/or economic evidence used for decision making. Low influence indicated that either the agency did not comment on the results of the ECA, implying the results were not taken into consideration, or the ECA results did not appear to influence the agency's decision. A high level of influence can affect the decision in either a positive or negative direction, and the level of influence does not imply quality of the ECA. For example, an ECA can be poorly executed and have a high impact on the agency issuing a negative decision. From the statistical and methodological issues noted in the agencies' assessment of the ECA, we identified recurring critiques of the key ECA elements (Table 1). The critique themes aligned well with the typical challenges associated with nonrandomized evidence (eg, unmeasured confounding). Counts of critiques and average number of critiques for each case study were calculated.

Results

The 7 case studies selected were blinatumomab for Philadelphia chromosome-negative relapsed or refractory B-cell precursor acute lymphoblastic leukemia (ALL) (Ph− ALL) and B-cell precursor ALL with minimal residual disease greater than or equal to 0.1% (MRD+ ALL), avelumab for metastatic Merkel Cell carcinoma (mMCC), erdafitinib for metastatic urothelial carcinoma that has susceptible FGFR3 or FGFR2 genetic alterations (FGFR2/3+ mUC), entrectinib for ROS1 positive nonsmall cell lung cancer (ROS1+ NSCLC), trastuzumab deruxtecan for unresectable or metastatic HER2 positive breast cancer, and idecabtagene vicleucel for relapse or refractory multiple myeloma. Across the 7 case studies, 20 regulatory reviews and 29 HTA decisions were evaluated, including 7 HTA resubmissions (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.05.016). All agencies reviewed blinatumomab for Ph− ALL, avelumab, and entrectinib. The manufacturer of blinatumomab MRD+ ALL did not submit a dossier to HAS. Although we prioritized cases with a formal statistical comparison of the ECA to the single-arm study, we included avelumab, because it represented a unique case type that used a naive comparison. Only the FDA and HC evaluated erdafitinib, but it was selected because it did not had EMA and FDA orphan status. At the time of analysis, PCODR, PBAC, and G-BA did not have assessments for trastuzumab deruxtecan and PCODR, NICE, PBAC, and G-BA did not have assessments of idecabtagene vicleucel.

Table 1. Categories of common ECA critiques.

ECA critique category	Definition
Generalizability	
SoC inconsistent over time	Treatment practices have changed over time, and thus, the generalizability of the external control group is questionable.
ECA nongeneralizable to clinical practice	ECA patient population was derived from outside the country of interest, ECA patient population and market authorization did not match, or other differences in ECA population compared with clinical practice.
Mitigation of confounding	
Unmeasured confounding	All-important known confounders were not available in the data and/or were not included in the adjustment analysis.
Unjustified confounders	Confounders used in adjusting were not justified—no rationale provided regarding why the variable was considered a confounder.
Naive comparison	No adjustment for confounders was executed (eg, propensity score matching).
Other	
Selection bias	Individuals or groups in a study differ systematically from the population of interest leading to a systematic error in an association or outcome. Includes differences related to start of follow-up time (eg, immortal time bias).
Incorrect adjusting methods	Incorrect adjustment methods were used.
Inconsistent outcomes definitions	Outcome variables were defined differently in the clinical trial vs RWD.
Data loss/insufficiency	Due to matching, the power to detect effect was reduced or substantial missing data impacted results.

ECA indicates external control arm; RWD, real-world data; SoC, standard of care.

Blinatumomab Ph− ALL

Blinatumomab is approved for "treatment of adults with Ph− ALL." The FDA approved blinatumomab in 2014, followed by EMA and HC approvals in 2015. Regulatory and HTA submissions included a phase II single-arm trial with an ECA as supplemental evidence. The goal of the ECA, built from a historical database from EU and US study sites, was to provide a relative measurement of complete remission (CR) and overall survival between blinatumomab in the experimental arm and standard of care (SoC) in the external arm. A literature review determined that the percentage of patients on SoC that achieved CR ranged from 18% to 45%; the ECA was used to demonstrate that the efficacy cut point of 30% CR used in the experimental arm was reasonable.

Table 2. Critiques by case study and agency.

Case	Type	Agency	Influence	SoC inconsistent over time
Blinatumomab Ph− ALL	Regulator	FDA	High	
		EMA	High	
		HC	Low	✓
	HTA	NICE	Low	
		G-BA	Low	✓
		HAS	Low	
		pCODR	Low	✓
		PBAC	High	✓
Blinatumomab MRD+ ALL	Regulator	FDA	Low	
		EMA	Low	
		HC	Low	
	HTA	NICE	Low	
		G-BA	Low	
		pCODR	Low	✓
		PBAC	High	✓
Avelumab	Regulator	FDA	Low	
		EMA	Low	
		HC	Low	
	HTA	NICE	High	
		G-BA	Low	
		HAS	Low	
		pCODR	Low	
		PBAC	High	
Erdafitinib	Regulator	FDA	Low	
Entrectinib	Regulator	FDA	Low	
		EMA	Low	
	HTA	G-BA	Low	
		pCODR	Low	
Fam-trastuzumab deruxtecan-nxki	Regulator	FDA	Low	
		EMA	Low	✓
	HTA	HAS	Low	
Idecabtagene vicleucel	Regulator	FDA	Low	✓
		EMA	High	
	HTA	HAS	Low	
Sum of critiques across cases				8

ALL indicates acute lymphoblastic leukemia; ECA, external control arm; EMA, European Medicines Agency; FDA, Food and Drug Administration; G-BA, Federal Joint Committee; HAS, Haute Autorité de Santé; HC, Health Canada; HTA, health technology assessment; LoT, line of therapy; MRD, minimal residual disease; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; pCODR, pan-Canadian Oncology Drug Review; Ph−, Philadelphia chromosome-negative; SoC, standard of care.
*Large percentages of patients in ECA had comparable efficacy endpoints.
†FDA noted that key differences (eg, age, LoT) were accounted for; HAS and pCODR had criticisms.
‡NICE mentioned that arms are balanced.
§NICE mentioned that weighting methods are appropriate.

The ECA was a main component of the clinical evaluation for the efficacy of blinatumomab for the FDA, EMA, and PBAC and appeared to highly influence these agencies' decisions (Table 2). Although critiques varied across agencies (Table 2), the most prevalent critique related to using a historical control arm spanning multiple years when the SoC varied. Blinatumomab was granted regulatory approval and reimbursement from each agency, although in multiple cases the reimbursement was restricted (Table 3).

Blinatumomab MRD+

The FDA, EMA, and HC approved expansion of blinatumomab's label to include MRD+ ALL. The FDA and HC approved the therapy for both children and adults, whereas the EMA included adults only. The regulatory and HTA submissions included a phase II single-arm study and an ECA to provide supplemental evidence of relapse free survival and overall

Table 2. Continued

ECA nongeneralizable to clinical practice	Unmeasured confounding	Unjustified confounders	Selection bias	Incorrect adjusting	Inconsistent outcomes definitions	Data loss/ insufficiency
*	†					
			✓		✓	
	‡					
				✓		
	✓				✓	
	✓		✓			
		✓		✓		
	✓		✓	✓		
✓	§					
			✓			
	✓					
✓	✓		✓			✓
			✓			
✓						
			✓			
			✓			
					✓	
						✓
✓	✓		✓		✓	✓
✓			✓		✓	
✓	✓					
	✓		✓			
	✓		✓		✓	✓
	✓					
	✓		✓		✓	✓
	✓		✓			✓
	✓					✓
6	13	1	14	3	7	7

survival in patients treated with the SoC. The RWD came from a registry outside the United States.

Blinatumomab's ECA appeared to have a low impact on all agency decision making except for PBAC (Table 3). All agencies except PBAC appeared to use only the single-arm study evidence as proof of efficacy and safety. Although PBAC's critiques of the ECA were like the other agencies (Table 2), it did conclude that there was evidence of superior comparative effectiveness based on the ECA results, although the magnitude was uncertain.

Avelumab

Avelumab was approved by the FDA, EMA, and HC in 2017 for the treatment of patients with mMCC. The FDA approval included adult and pediatric patients at the age of 12 years or older and was

Table 3. External control arm's level of influence on the decision, by agency and drug.

			Regulator	
Drug	**Indication**		**FDA**	**EMA**
Blinatumomab (Blincyto)	(Ph-) R/R BCP ALL	Decision	Accelerated approval	Conditional approval
		Influence	High	High
Avelumab (Bavencio)	mMCC	Decision	Accelerated approval for pts 12+ years old	Conditional approval for adults
		Influence	Low	Low
Blinatumomab (Blincyto)	(MRD+) R/R BCP ALL	Decision	Accelerated approval	Conditional approval
		Influence	Low	Low
Erdafitinib (Balversa)	FGFR2/3+ mUC	Decision	Accelerated approval	-
		Influence	Low	-
Entrectinib (Rozlytrek)	ROS1+ mNSCLC	Decision	Accelerated approval	Conditional approval for pts not previously treated with a ROS1
		Influence	Low	Low
Fam-trastuzumab deruxtecan-nxki (Enhertu)	unresectable or metastatic HER2+ breast cancer (2+L)	Decision	Accelerated approval	Conditional approval
		Influence	Low	Low
Idecabtagene vicleucel (Abecma)	Multiple myeloma (4+L)	Decision	Approved	Conditional Approval
		Influence	Low	High

1L indicates first-line; 2L, second-line; ASMR, drug's improvement in medical benefit measured on a scale from I (major benefit) to IV (no improvement); CDF, Cancer Drugs Fund; CR, complete remission; ECA, external control arm; EMA, European Medicines Agency; FDA, Food and Drug Administration; FGFR2/3+ mUC, metastatic urothelial carcinoma that has susceptible FGFR3 or FGFR2 genetic alterations; G-BA, Federal Joint Committee; HAS, Haute Autorité de Santé; HC, Health Canada; HTA, health technology assessment; mMCC, metastatic Merkel Cell carcinoma; (MRD+) R/R BCP ALL, B-cell precursor acute lymphoblastic leukemia (ALL) with minimal residual disease (MRD) greater than or equal to 0.1%; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; pCODR, pan-Canadian Oncology Drug Review; Ph− R/R ALL, Philadelphia chromosome-negative relapsed or refractory B-cell precursor acute lymphoblastic leukemia; Pt, patient; PS, performance status; SMR, drug's medical benefit, which ranges from insufficient to important.

indicated for chemotherapy naive patients, whereas the EMA and HC's approved indications focused on adults only. HC further restricted the indicated population to previously treated adults. Regulatory and HTA submissions for avelumab included a single-arm study and an ECA, based on a retrospective chart review and registry data to characterize the natural history of mMCC with respect to treatment outcomes and contextualize the risk/benefit profile of avelumab. The manufacturer made no formal statistical comparison between treatment arms.

The major critiques from most agencies were the lack of a formal statistical comparison between the ECA and the clinical trial and the potential for selection bias, due to missing data on the ECA patients including Eastern Cooperative Oncology Group (ECOG) performance score, life expectancy, and previous therapies. Even when agencies agreed on critiques of avelumab's ECA, its level of influence seemed to differ among them (Table 2). For most, the ECA's level of influence appeared low; thus, agencies relied on the single-arm study results alone in the evaluation of efficacy and did not consider comparative efficacy.

Erdafitinib

Erdafitinib was approved by the FDA and HC in 2019 and 2020, respectively, for the treatment of adult patients with FGFR2/3+ mUC and who progressed during or after at least 1 line of previous platinum-containing chemotherapy including within 12 months

Table 3. Continued

Regulator	HTA				
HC	NICE	G-BA	HAS	pCODR	PBAC
Conditional approval	Recommended with restrictions (only if discount provided)	Nonquantifiable additional benefit	Recommended for 2L: ASMR III, SMR Substantial	Recommended with restrictions for # of cycles	Recommended with restrictions (# cycles; recommended only after resubmissions)
Low	Low	Low	Low	Low	High
Conditional approval (2L)	Recommended with restrictions (2L)	Nonquantifiable additional benefit	Recommended with restrictions: 2L+ ASMR: IV, SMR: Substantial	Recommended with restrictions (good PS, # cycles)	Recommended with restrictions for 1L (# cycles)
Low	High	Low	Low	Low	High
Conditional approval	Recommended with restrictions (first remission)	Nonquantifiable additional benefit	Do not recommend; HAS issued notice that manufacturer did not submit for reimbursement	Recommend with restrictions (good PS)	Recommended with restrictions (# cycles; recommended only after resubmissions)
Low	Low	Low	N/A	Low	High
Conditional approval	-	-	-	-	-
N/A - did not review ECA	-	-	-	-	-
Conditional approval	Recommended for pts not previously treated with a ROS1	No proof of added benefit for pts not previously treated with a ROS1	Do not recommend; ASMR N/A, SMR Insufficient	Recommended with restrictions (good PS)	Recommended
N/A - did not review ECA	N/A - did not review ECA	Low	N/A - did not review ECA	Low	N/A - did not review ECA
Conditional approval	Recommended with restrictions (CDF)	-	Recommended with restrictions (reassessment in 18 mo); ASMR V, SMR Substantial	-	-
N/A - did not review ECA	N/A - did not review ECA	-	Low	-	-
Conditional Approval	-	-	Recommend with restrictions (registry; qualified centers only); ASMR V, SMR Important	-	-
N/A - did not review ECA	-	-	Low	-	-

of neoadjuvant or adjuvant platinum-containing chemotherapy. The FDA regulatory submission included a single-arm study and an ECA generated from US EHRs and German registry data. The goal of the ECA was to "demonstrate that patients with FGFR alterations have lower rates of response to the SoC."[11] HC's approval documentation did not include a description or discussion of the ECA. Either the ECA was not submitted, or HC did not believe it was relevant evidence for decision making. Only the FDA approval documentation was available to analyze.

Due to methodological concerns, the ECA seemed to have little influence on the FDA decision (Table 2). The FDA approved erdafitinib based on the acceptable safety profile, response rate, and duration of response from single-arm study alone.

Entrectinib

Entrectinib was approved by the FDA and HC in 2019 and 2020, respectively, for ROS1+ NSCLC and was approved by the EMA in 2020 for ROS1+ NSCLC patients not previously treated with an ROS1 inhibitor (Table 3). The clinical evidence included an integrated analysis of 3 phase II trials and an ECA, generated from EHR data, to compare time-to-treatment discontinuation and progression-free survival with patients on crizotinib.

The ECA was evaluated by the FDA, EMA, pCODR, and G-BA. HAS, NICE, and PBAC relied on indirect comparison to chemotherapy from RCTs instead of the ECA, likely due to differences in standards of care (eg, crizotinib was not considered the appropriate

comparator in the NICE assessment). The ECA had a low impact on decision making due to perceived methodological issues (Table 2). For example, G-BA stated that selection bias and unmeasured confounding could explain the small, estimated effect, and thus, the ECA was not suitable to demonstrating a comparative benefit.

Trastuzumab Deruxtecan

Trastuzumab deruxtecan was approved by the FDA in 2019 and the EMA and HC in 2021 for treatment of unresectable or metastatic HER2-positive breast cancer in patients who have failed previous therapy; HC specified that patients must have received previous therapy with trastuzumab emtansine. The clinical evidence included multiple phase II studies and an ECA built from French hospital EHR data. The goal of the ECA was to provide a baseline for progression-free survival and ORR of other therapies with a comparable patient population.

The ECA appeared to have low impact on regulatory and HTA decision making (Table 2). Although HC and NICE assessed the drug, they did not discuss the ECA. Both the FDA and HAS described the ECA, but only relied on and reported the single-arm study results. HAS and EMA noted that the ECA did not control for all relevant confounders and the EMA took issue with several other aspects (Table 2).

Idecabtagene Vicleucel

Idecabtagene vicleucel was approved for fourth-line relapse or refractory multiple myeloma by the FDA, EMA, and HC in 2021. The clinical evidence included a single-arm study and an ECA generated from RWD across multiple sources including clinical sites, registries, and a research database. The goal of the ECA was to provide an estimate of ORR in patients receiving at least 3 previous myeloma regimens.

The FDA, EMA, and HAS evaluated the ECA and agreed that, due to missing data for baseline prognostic factors, important covariates were not included in the model, and thus, results were likely biased (Table 2). Both the FDA and HAS considered these methodological issues to be substantial enough to not consider the ECA in their decision making. Nevertheless, despite the limitations, the EMA "found that the ECA contextualized the findings" from the single-arm study and "thus supported approval."

Quantifying Regulatory and HTA Critiques

Within the 7 case studies and across the 8 regulatory and HTA agencies, 59 critiques were recorded. Selection bias and unmeasured confounding were the most cited critiques (Table 2). On average, there were 2.4 critiques per agency for each ECA. PBAC had the highest average number of critiques (3.3) followed by the FDA (2.4). There was a similar number of critiques from regulatory and HTA agencies (2.1 each), on average.

Comparing Critiques

When 8 regulatory and HTA agencies reviewed the same clinical evidence, reviewers raised different critiques (Table 2) and, in several cases, had opposing conclusions on the clinical evidence base and subsequent impact the ECA had on decisions. In blinatumomab's Ph− ALL ECA, the FDA noted that the propensity score matching analysis resulted in key potential confounders becoming balanced between the ECA and single-arm study. Nevertheless, HAS and pCODR did not agree; HAS noted that there was a mismatch of patients receiving allograft treatment post propensity score matching, and pCODR noted differences in previous lines of therapy and a lack of information on the performance status of patients in the ECA. Opposing views on both the impact of confounding and the quality of the ECA likely led to differences in how blinatumomab's ECA potentially affected agencies' decision making. Even when agencies agreed on critiques, the impact of the ECA varied. The FDA and EMA agreed that the ECA for idecabtagene vicleucel suffered from data loss, unmeasured confounding, and selection bias. Nevertheless, the EMA considered the ECA supportive in decision making (high impact) whereas the FDA did not (low impact).

Differences in regulator and HTA remits might explain differences in critiques. Nevertheless, we observed differences in critiques among regulators and among HTAs. Regulatory reviewers from the FDA and HC critiqued the same evidence differently for blinatumomab's Ph− ALL's ECA. HC cited the potential presence of selection bias due to differences in enrollment criteria regarding blast counts and inconsistencies in primary outcome definitions between the single-arm study and the ECA. HC based its reimbursement decision solely on the single-arm study due to these methodological flaws, yet the FDA did not raise similar concerns. NICE and G-BA cited selection bias as a major issue for avalumab's ECA whereas HAS, pCODR, and PBAC did not critique it. For NICE and PBAC, avelumab's ECA was used to provide generalizable estimates of efficacy versus SoC, whereas HAS and G-BA relied only on the single-arm study and dismissed the ECA for methodological flaws.

Discussion

ECAs generated from RWD supplementing single-arm oncology trials for regulatory approval and reimbursement decisions have become an increasingly popular strategy to provide decision makers with contextual information on the experience of patients not given the study therapy.[15] The 7 cases analyzed were all approved by regulators, mostly through accelerated pathways, and in most cases were granted positive reimbursement decisions (although numerous decisions needed resubmissions to negotiate terms for pricing and came with additional restrictions on use; Table 3). HTA resubmissions were not driven by the need to clarify the ECA clinical evidence, potentially indicating that in these cases the ECA evidence was sufficient for submission or that the ECA was not necessary for decision making. The ECA evidence appears to have limited impact on regulatory and HTA decision making; the ECA appeared to have high impact on the decision in only 7 of 34 total agency assessments (1 high impact for FDA; 2 for EMA; 3 PBAC; 1 NICE). Given these use cases were identified first via FDA approvals, we expected the ECAs—especially in FDA reviews—to be more impactful. Thus, the limited impact of the ECAs is an interesting finding. It could be due to the inherent challenges with generating robust ECAs within RWD (eg, selection bias or confounding), regulatory and HTA's limited experience with interpreting ECAs, or it could be related to special circumstances within oncology. The impact of ECAs on regulatory and HTA decision making should be explored over time and in disease conditions outside oncology.

When reviewing the same clinical evidence, there was disagreement between and among regulatory and HTA agencies on ECA critiques and the ECA's impact on decision making within the 7 case studies. Differences in clinical evidence interpretation are not a new phenomenon in agency decision making.[16] Differences in remits, approaches to risk, and processes can affect evidence interpretation. Nevertheless, we did not observe distinctive patterns in critiques or ECA's influence between agency types and among agencies in the 7 case studies. In addition, the number of critiques did not appear to be related to the impact of the ECA. For example, the EMA discussed 3 critiques (ie, unmeasured confounding, selection bias, and data loss/insufficiently) for idecabtagene vicleucel and, despite these limitations, found the ECA to contextualize the findings from the single-arm study and thus support approval.

Consistent with challenges in RWE studies in general, the common critiques across agencies were selection bias and

confounding. Selection bias was the most cited critique of ECAs for multiple agencies across case studies (eg, G-BA blinatumomab's MRD+ ALL; FDA and G-BA avelumab; Table 2). Both selection bias and confounding can be mitigated through selecting high-quality fit-for-purpose data (ie, data relevant to the research question) and good epidemiological practices that focus on a priori study design.[16] Although using appropriate methods is essential, methodological choices cannot compensate for poor quality data. For ECAs to be successful components of evidence packages, high-quality RWE must exist. There may be unique challenges in RWD collection (eg, lack of ECOG performance status in RWD sources) that increase the frequency of selection bias and confounding issues; further studies should explore frequency of ECA methodological critiques in other therapeutic areas.

For manufacturers to ensure they are presenting the best available evidence to regulatory and reimbursement decision makers, they must understand how evidence will be interpreted and used in specific jurisdictions. How agencies weigh and interpret clinical evidence, like an ECA, warrants guidance from decision makers on what constitutes the appropriate use of an ECA and a description on how the evidence will be evaluated. Although there are numerous recommendations on RWE, a recent article[16] highlighted the current gaps in existing RWE guidance and noted the need for additional guidance on ECA study design and execution. RWD-based ECA guidance from the FDA is expected in 2022.[17]

This article has some limitations. From the systematic search of ECAs in FDA approvals, we selected 7 case studies that the agencies were likely to have an in-depth evaluation of the ECA evidence (ie, mostly focused on ECAs with formal statistical comparison submitted as efficacy evidence). We also excluded cases due to the lack of assessments from other agencies because our objective was to compare agency decision making when the same ECA evidence was submitted. There is the potential this introduced selection bias in our sample. Our analysis was based on publicly available decision documentation published by regulatory and HTA bodies. It is possible that not all criticism or acceptance of ECAs' evidence was included in agency reviews. Nevertheless, these documents are commonly used in the literature to detail why a specific regulatory and reimbursement decision was made. We focused on the presence of an ECA in the FDA documentation as an anchor. There may be instances for which therapies have been approved with ECA evidence in ex-US jurisdictions but were not approved by the FDA. The FDA does not publish decision documentation for therapies that are not approved, which could skew the selection of cases to more positive assessments of ECAs. For some agencies, the single-arm study evidence was enough to justify regulatory or reimbursement approval, which could indicate that the severity of the cancer and lack of alternative treatments are more influential to the decision than the need for comparative effectiveness evidence. Finally, agencies make decisions on the totality of evidence and do not publicly discuss which evidence ranked higher in their decision making, which makes it difficult to determine the impact of the ECA on the final decision. We subjectively rated the level of influence based on the agencies' summary of the evidence base and discussion of the ECA.

Although limited to 7 cases, to the best of our knowledge, this was the first qualitative assessment to compare regulatory and HTA agencies' evaluations of ECAs generated from RWD for oncology therapies. With increased interest in RWE to supplement RCT data and ECAs to provide contextual information to single-arm studies, it is important for researchers to understand how ECAs have historically been evaluated. With the absence of regulatory and HTA agency guidance on design and use of ECAs, researchers must focus on recent examples to help guide impactful evidence generation. Although agencies did not consistently agree on the quality and impact of the ECA, 2 main critiques emerged: selection bias and confounding. These issues can be mitigated through appropriate data selection and study design. Future research should focus on the applicability of oncology RWD for ECA use cases (eg, the presence key study design elements like ECOG status) and expand the evaluation of ECAs outside of oncology.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.05.016.

Article and Author Information

Accepted for Publication: May 10, 2022

Published Online: June 24, 2022

doi: https://doi.org/10.1016/j.jval.2022.05.016

Author Affiliations: Scientific Research, Aetion Inc, New York, NY, USA (Jaksa, Louder, Maksymiuk, Gatto); Daiichi-Sankyo Europe GmbH, Munich, Germany (Vondeling, Martin, Richards, Yver, Rosenlund); Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden (Rosenlund).

Correspondence: Ashley Jaksa, MPH, Scientific Research, Aetion Inc, 5 Penn Plaza, 7th Floor, New York, NY 10001, USA. Email: ashley.jaksa@aetion.com

Author Contributions: *Concept and design:* Jaksa, Louder, Vondeling, Martin, Yver, Rosenlund
Acquisition of data: Jaksa, Louder, Maksymiuk
Analysis and interpretation of data: Jaksa, Louder, Maksymiuk, Vondeling, Martin, Gatto, Richards, Yver, Rosenlund
Drafting of the manuscript: Jaksa, Louder, Maksymiuk
Critical revision of the paper for important intellectual content: Jaksa, Louder, Vondeling, Martin, Gatto, Richards, Yver, Rosenlund
Statistical analysis: Jaksa
Obtaining funding: Rosenlund
Administrative, technical, or logistic support: Martin
Supervision: Vondeling, Martin, Yver, Gatto, Richards, Rosenlund

Conflict of Interest Disclosures: Mr Jaksa and Drs Louder, Maksymiuk, and Gatto are employed by and reported stock ownership in Aetion Inc and reported receiving funding from Daiichi-Sankyo Europe GmbH. Dr Gatto reported stock ownership in Pfizer Inc. Drs Vondeling, Martin, Richards, Yver, and Rosenlund were employed by Daiichi-Sanko. Drs Richards, Yver, and Rosenlund reported stock ownership in Daiichi-Sanko. No other disclosures were reported.

Funding/Support: This work was funded by Daiichi-Sanko.

Role of the Funder/Sponsor: The funder was involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Beaver JA, Howie LJ, Pelosof L, et al. A 25-year experience of US Food and Drug Administration accelerated approval of malignant hematology and oncology drugs and biologics: a review. *JAMA Oncol*. 2018;4(6):849–856.
2. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000-2016 found variation in establishing comparative clinical effectiveness. *J Clin Epidemiol*. 2019;105:50–59.
3. Framework for FDA's real-world evidence program. FDA, US Food and Drug Administration. https://www.fda.gov/media/120060/download. Accessed October 5, 2020.
4. Characterizing the use of external controls for augmenting randomized control arms and confirming benefit. Friends of Cancer Research. https://www.focr.org/

sites/default/files/Panel-1_External_Control_Arms2019AM.pdf. Accessed July 15, 2020.
5. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol*. 2020;12:457–467.
6. Seeger JD, Davis KJ, Iannacone MR, et al. Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. *Pharmacoepidemiol Drug Saf*. 2020;29(11):1382–1392.
7. Burcu M, Dreyer NA, Franklin JM, et al. Real-world evidence to support regulatory decision-making for medicines: considerations for external control arms. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1228–1235.
8. Mack C, Christian J, Brinkley E, Warren EJ, Hall M, Dreyer N. When context is hard to come by: external comparators and how to use. https://journals.sagepub.com/doi/abs/10.1177/2168479019878672. Accessed March 8, 2021.
9. Raphael MJ, Gyawali B, Booth CM. Real-world evidence and regulatory drug approval. *Nat Rev Clin Oncol*. 2020;17(5):271–272.
10. Choice of control group in clinical trials. European Medicines Agency. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf. Accessed July 15, 2020.
11. Demonstrating substantial evidence of effectiveness for human drug and biological products. US Food and Drug Administration. https://www.fda.gov/media/133660/download. Accessed July 15, 2020.
12. Goring S, Taylor A, Müller K, et al. Characteristics of non-randomised studies using comparisons with external controls submitted for regulatory approval in the USA and Europe: a systematic review. *BMJ Open*. 2019;9(2):e024895.
13. Bolislis WR, Fay M, Kühler TC. Use of real-world data for new drug applications and line extensions. *Clin Ther*. 2020;42(5):926–938.
14. External controls in research: the "what, why and how." UBC. https://ubc.com/external-controls-in-research/. Accessed October 14, 2021.
15. Patel D, Grimson F, Mihaylova E, et al. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health*. 2021;24(8):1118–1125.
16. Jaksa A, Demb J, Ho YS, Daniel K. Agency agreement in health technology assessment reimbursement decisions. *Value Health*. 2013;16(3):A3.
17. US Food and Drug Administration. CDER 2021 guidance agenda. https://www.fda.gov/media/134778/download. Accessed July 13, 2021.