

Effectively Leveraging RWD for External Controls: A Systematic Literature Review of Regulatory and HTA Decisions

Oriol Sola-Morales¹ , Lesley H. Curtis², Julien Heidt^{3,*} , Laura Walsh⁴ , Deborah Casso^{5,†} , Susan Oliveria^{6,‡}, Patrick Saunders-Hastings⁷, Yufei Song^{8,§} , Tiffany Mercado³, Robbert Zusterzeel^{3,§}, Vera Mastey⁹, James Harnett⁹ and Ruben G. W. Quck⁹

Real-world data (RWD)-derived external controls can be used to contextualize efficacy findings for investigational therapies evaluated in uncontrolled trials. As the number of submissions to regulatory and health technology assessment (HTA) bodies using external controls rises, and in light of recent regulatory and HTA guidance on the appropriate use of RWD, there is a need to address the operational and methodological challenges impeding the quality of real-world evidence (RWE) generation and the consistency in evaluation of RWE across agencies. This systematic review summarizes publicly available information on the use of external controls to contextualize outcomes from uncontrolled trials for all indications from January 1, 2015, through August 20, 2021, that were submitted to the European Medicines Agency, the US Food and Drug Administration, and/or select major HTA bodies (National Institute for Health and Care Excellence (NICE), Haute Autorité de Santé (HAS), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), and Gemeinsamer Bundesausschuss (G-BA)). By systematically reviewing submissions to regulatory and HTA bodies in the context of recent guidance, this study provides quantitative and qualitative insights into how external control design and analytic choices may be viewed by different agencies in practice. The primary operational and methodological aspects identified for discussion include, but are not limited to, engagement of regulators and HTA bodies, approaches to handling missing data (a component of data quality), and selection of real-world endpoints. Continued collaboration and guidance to address these and other aspects will inform and assist stakeholders attempting to generate evidence using external controls.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Real-world data (RWD)-derived external controls can be used to contextualize efficacy findings for investigational therapies evaluated in uncontrolled trials in the context of oncology and rare diseases with few treatment options, therapeutics seeking accelerated approval, and situations where it is inappropriate to randomize due to ethical considerations.

WHAT QUESTION DID THIS STUDY ADDRESS?

Despite recent guidance from regulators and health technology assessment (HTA) bodies aimed at improving the quality and reliability of real-world evidence (RWE), there is a need to assess the operational and methodological challenges that impede the quality and consistency of RWE generation in the context of external controls.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

In addition to confounding and selection bias, this review identified engagement of regulators and HTA bodies, approaches to handling missing data (a component of data quality), and selection of RW endpoints as key considerations when implementing an external control.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

Continued collaboration to address these aspects through external control-specific guidance will inform and assist stakeholders attempting to generate adequate scientific evidence using RWD-derived external controls.

¹Fundació Health Innovation Technology Transfer and International, University of Catalonia, Barcelona, Spain; ²Duke Clinical Research Institute, Durham, North Carolina, USA; ³IQVIA, Regulatory Science and Strategy, Falls Church, Virginia, USA; ⁴IQVIA, Epidemiology and Drug Safety Practice, Boston, Massachusetts, USA; ⁵IQVIA, Epidemiology and Drug Safety Practice, Seattle, Washington, USA; ⁶IQVIA, Epidemiology and Drug Safety Practice, New York, New York, USA; ⁷Epi Excellence LLC, Garnet Valley, Pennsylvania, USA; ⁸IQVIA, Epidemiology and Drug Safety Practice, Beijing, China; ⁹Regeneron Pharmaceuticals Inc., Health Economics & Outcomes Research, Tarrytown, New York, USA. *Correspondence: Julien Heidt (julien.heidt@iqvia.com)

[†]Contribution to the manuscript was made while employed by IQVIA; these authors are no longer employed by IQVIA.

Received February 13, 2023; accepted April 14, 2023. doi:10.1002/cpt.2914

Real-world data (RWD) analyses have been used for decades to generate real-world evidence (RWE) in support of postmarketing regulatory decision making, including safety signal evaluation, risk management, and product life cycle benefit–risk evaluation.¹ With increased generation of relevant and reliable RWD, the use of RWE in pre-authorization regulatory decision making, and subsequently for reimbursement, has grown.^{2–4} Specifically, RWD-derived external controls are increasingly being used for therapeutics seeking accelerated approval in situations where ethical considerations preclude randomization, and to contextualize efficacy findings for investigational therapies evaluated in uncontrolled trials in oncology and rare diseases with limited treatment options. External controls from RWD may be used as benchmarks or comparators. RWD benchmarks contextualize and characterize the natural history of a disease, whereas RWD comparators are used as a “formal” comparison and, as such, must reflect the patient population, eligibility criteria, design, and analytical features of the uncontrolled trial.⁵ As submissions to regulatory and health technology assessment (HTA) bodies based on data from uncontrolled trials using external controls continues to rise^{6,7} and the agencies’ understanding of RWE evolves, there is a burgeoning need to address the operational and methodological challenges that impede the quality and consistency of evidence generation.⁸

Both the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) have established approval pathways and designations for using RWE in circumstances where the definition of an adequate and well-controlled study deviates from a typical randomized controlled trial (RCT).⁹ In 2016 and 2017, the 21st Century Cures Act and the sixth Prescription Drug User Fee Act (PDUFA VI) in the United States set milestones for the FDA to explore the use of RWE in regulatory decision making. The FDA’s RWE program, released in 2018, outlines core considerations for using RWE to support effectiveness labeling changes for marketed drugs and biologics.^{10,11} The FDA has released several guidance documents aimed at improving the quality and reliability of RWE.^{12–16} In 2023, the FDA released a draft guidance on considerations for the design and conduct of externally controlled trials using patient-level data.¹⁷ The EMA has issued its own set of guidelines and initiatives^{18–22} and works closely with regional and national European HTA bodies who provide recommendations on medicines and other health technologies that can be financed or reimbursed by the healthcare system in a particular European Union Member State or region.²³ Despite limited HTA guidance on RWD-derived external controls,^{24–26} RWE has been used in a multitude of applications to inform healthcare decision making by National Institute for Health and Care Excellence (NICE) and other HTA bodies, particularly in the field of oncology where clinical trials often do not provide all evidence required to inform cost-effectiveness.^{27–29} In June 2022, NICE released its RWE framework, which describes best-practices for planning, conducting, and reporting RWE studies to improve the quality and transparency of evidence.³⁰

Researchers, professional societies, government agencies, and multistakeholder initiatives have generated numerous frameworks and recommendations offering guiding principles for

key RWE study design elements.^{31–40} Whereas creation of these frameworks is invaluable to the acceptance of RWD-derived external controls for regulatory and reimbursement decision making, understanding how sponsors have implemented and operationalized RWE guidance in recent submissions is imperative. Recent publications^{7,9,27,41–49} have reviewed regulatory and HTA submissions using external controls; however, few have been systematic in nature, and none have examined submissions across both regulatory and HTA bodies and all therapeutic areas. Furthermore, there is limited systematic research investigating how RWD-derived external controls have been used in HTA settings.⁵⁰ To address these gaps, this systematic review summarizes publicly available information on the use of RWD-derived external controls to contextualize outcomes from uncontrolled trials submitted to the EMA, the FDA, and/or select major HTA bodies in the United Kingdom (UK), France, and Germany (NICE, Haute Autorité de Santé (HAS), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), and Gemeinsamer Bundesausschuss (G-BA)) for all indications from January 1, 2015, through August 20, 2021. By systematically reviewing submissions to both regulatory and HTA bodies in the context of recent guidance, this study quantitatively and qualitatively provides sponsors, regulators, and HTA bodies with an understanding of how design and analytic choices may be viewed by different agencies in practice and may provide insight into the design of trials using external controls.

METHODS

Systematic review of regulatory and HTA submissions was conducted through the FDA and EMA web portals and databases^{51–53} and IQVIA’s HTA Accelerator (HTAA),⁵⁴ respectively. Each data source is described in Appendix S1. Regulatory and HTA submissions were selected using the following eligibility criteria first by high-level review and then through a second, detailed review (details can be found below). Submissions were eligible for inclusion if an agency decision was made between January 1, 2015, and August 20, 2021, with primary evidence from an uncontrolled trial explicitly utilizing an RWD-derived external control, where RWD is defined as data regarding the health state and/or delivery of health care collected in a noninterventional trial setting.²⁷ Submissions identified as utilizing a “literature-based” external control were confirmed to be derived from either RWD only or a combination of RWD and historical trial data (not trial data only). If it could not be verified using publicly available information whether a literature-based estimate was derived from RWD or historical trial data, the submission was included so as to comprehensively include submissions utilizing RWD.

EMA submissions were only included if designated as conditional approval, exceptional circumstances, accelerated assessment, and/or orphan drug. Similarly, the FDA submissions were only included if designated as accelerated approval, breakthrough therapy, fast-track, priority approval, orphan drug, and/or Real-Time Oncology Review (RTOR) pilot program. Submissions were limited to the regulatory pathways and designations that most commonly involve submissions based on nonrandomized evidence.⁹ To limit the scope of the natural language processing (NLP) search results to submissions relevant to the objectives of this study, a series of HTAA search criteria were applied using fields populated in the HTAA for each submission (Appendix S1). Submissions were excluded from review if RCT data were included in the submission as primary evidence, the product was a generic drug, a biosimilar, a diagnostic technology, or medical device, or if there was no explicit use of RWD analyses.

Regulatory search strategy

Regulatory documents were downloaded directly from the EMA and FDA websites.^{51–53} If a reviewed product had a determination by both the EMA and FDA, submission information for that product was extracted independently for each agency. If a regulatory application for a product was made to one agency but did not appear in the results for the other agency, a search with no date restriction was performed to identify applications made outside the review window (to ensure all comparable submissions had been included). If there were two qualifying submissions for the same product, but two different indications, data were presented for each unique submission. If the initial submission was outside the study period but the resubmission was within the study period, data were also extracted on the initial submission. This systematic review has been registered on the International Prospective Register of Systematic Reviews (PROSPERO; record ID: CRD42021240950).

HTA search strategy: NLP search term development

HTA submissions eligible for high-level review were identified using an NLP search strategy in the HTAA. A list of NLP search terms ([Appendix S1](#)) was iteratively developed based on subject matter knowledge of RWD in the context of external control studies. In this conservative approach, the final list of NLP search terms was inclusive of potential matches and any resultant false positives were manually removed in the subsequent high-level review. The terms were translated into French and German to identify relevant matches in documents from HAS and G-BA/IQWiG, respectively. The list of search terms was then used to identify relevant HTA submissions (referred to as record IDs in HTAA). The output of the NLP search included the phrase containing the qualifying term in the PDF document, count of term occurrence, and HTAA record ID. Submissions for the same product to multiple HTA bodies were assessed independently.

Escalating double review

An escalating double review was conducted in the same manner for high-level and detailed review of regulatory and HTA documents wherein a prespecified number of submissions were subject to 100% double review and the percentage of remaining submissions subject to double review increased incrementally if discrepancies between reviewers were identified during adjudication. Details of the escalating double review are described in [Appendix S1](#).

Regulatory high-level review

High-level review of regulatory submissions involved a piloted, keyword-based review of European Public Assessment Reports (EPARs)/FDA submission documents using a list of search terms based on prespecified eligibility criteria keywords ([Appendix S1](#)). Documents for unique submissions were included/excluded in the review based on the presence of prespecified keywords for inclusion/exclusion and their context. For submissions with no qualifying keywords, the Executive Summary from the FDA submission document and/or the Background section from the EPAR were reviewed for information to determine eligibility. If a submission was excluded, the reason for exclusion was noted.

Following high-level review, the lists of qualifying submissions from the EMA and FDA were compared to confirm that the submissions were made to one agency during the study period but to the other agency outside of the study period.

HTA high-level review

HTA submissions identified in the NLP search were compared against the eligibility criteria in a high-level review within the HTAA. Fields populated in the HTAA record from all available HTA documents were compared against eligibility criteria to judge inclusion. If excluded, the reason for exclusion was noted.

Regulatory detailed review

Detailed review required in-depth review of primary EMA and FDA submission documents to identify reasons for exclusion that were not ascertainable from the initial keyword search and provided further context for the use of RWD. Reason for exclusion is documented in [Figure 1a,b](#).

HTA detailed review

Detailed review of HTA submissions involved screening of submissions against eligibility criteria using only key documents available in the HTAA record: Final Appraisal Determination, Final Guidance (NICE), Final Report (HAS), Summary Decision, Supporting Grounds (G-BA), Full Report, Summary, and Addendum (IQWiG). Key documents from HAS, G-BA, and IQWiG were translated prior to detailed review ([Appendix S1](#)). Reason for exclusion is documented in [Figure 1c](#).

Data extraction

Two reviewers independently extracted data for each submission. Discrepancies were adjudicated by a third, independent reviewer. Data extraction were conducted using a standardized data extraction form and associated data dictionary. Data were extracted on each RWD analysis in each submission. Variables for extraction included, but were not limited to, RWD temporality, endpoints, data source description, study design and methodology, benefit–risk determination, and agency feedback.

The data extraction form for the EMA was populated using the EPAs and, as relevant, additional supporting documentation from the EMA database. The data extraction form for the FDA was populated using both the relevant review(s)/label and additional supporting documentation from the FDA database. The FDA does not publicly provide information on submissions if they were not subsequently approved, which limits insight into FDA feedback on the RWE components. Key HTA submission documents available in the HTAA record for a given product were used to populate the data extraction form for HTA.

During data extraction, reviewers determined whether the RWE analysis was considered by the agency in the final decision (yes or no) based on the “Benefits/Risks” section of the EMA submission documents and the “Benefit–Risk Assessment” of the FDA submission documents (or the “Summary/Conclusions” section if there was no “Benefit–Risk Assessment” included in the documents used for extraction). For HTA, this was ascertained from the “Recommendations” section of the NICE submission documents, “Summary and Discussion” section of the HAS submission documents, “Assessment Summary” section of the G-BA submission documents, and “Summary of Benefit Assessment” section of the IQWiG submission documents. Text from other sections of the submission documents was considered if it clearly indicated that the RWE analysis was considered in the final decision. When RWE was deemed to have influenced the decision for a given submission, text was extracted from the relevant section to verify this determination. Analyses were conducted at the submission level or the analysis level (if multiple RWD analyses were conducted for a given submission).

RESULTS

Included submissions

Following high-level and detailed review, data on 26 EMA and 38 FDA submissions were extracted ([Figure 1a,b](#)). Data on 70 HTA submissions was extracted ($n = 16$ (NICE); $n = 12$ (IQWiG); $n = 25$ (G-BA); and $n = 17$ (HAS); [Figure 1c](#)). A full list of included regulatory and HTA submissions can be found in [Table S1](#).

Submission characteristics

The most common therapeutic areas for regulatory and HTA submissions utilizing an uncontrolled trial with an external

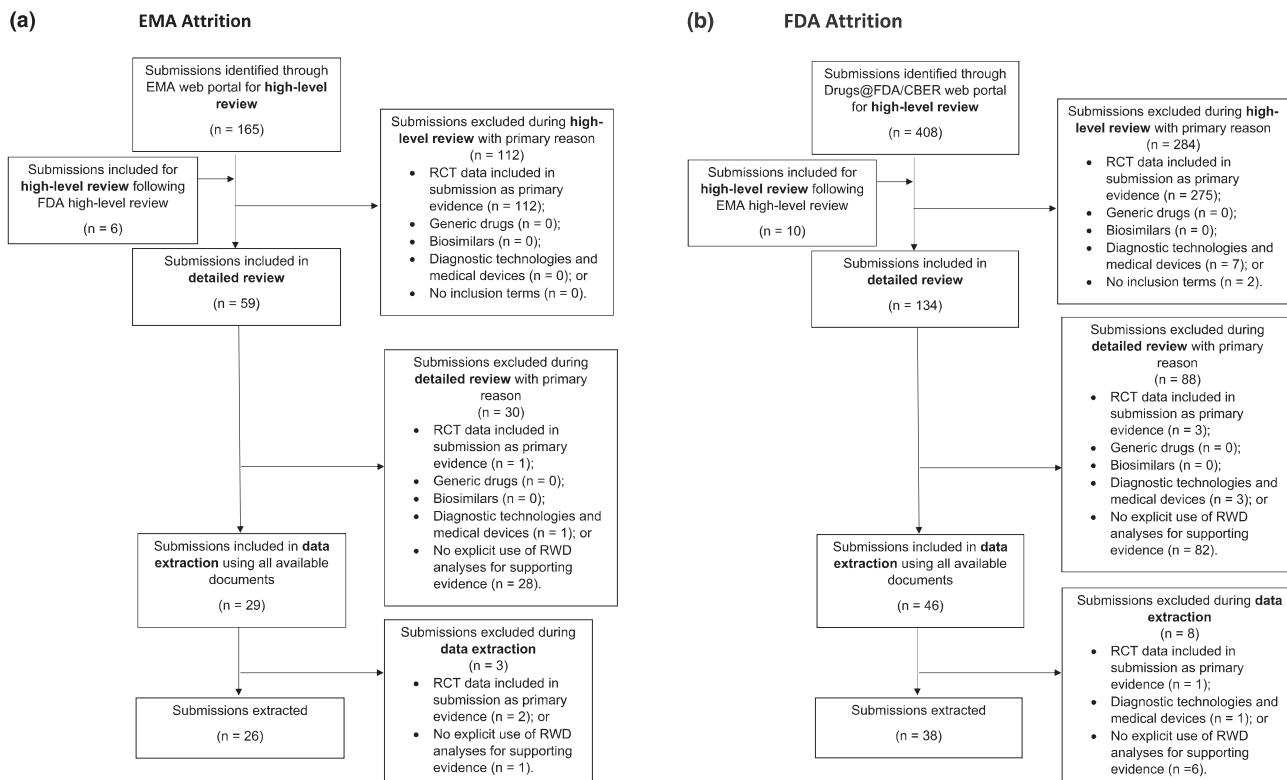


Figure 1 (a) PRISMA flow diagram showing the number of EMA submissions included and excluded during high-level review, detailed review, and data extraction, noting reasons for exclusion. Submissions were identified through the EMA web portal ($n=165$) as eligible for high-level review. Submissions passing high-level review for the EMA were compared with submissions passing high-level review for the FDA to capture submissions made outside of the study period (maximizing the number of submissions for direct comparison between the EMA and FDA). This yielded additional submissions eligible for the EMA ($n=6$) high-level review. During high-level review, the EMA submissions were primarily excluded due to inclusion of RCT data as primary evidence. During detailed review, the EMA submissions were primarily excluded due to no explicit use of RWD analyses for supporting evidence. Of the submissions eligible for data extraction, review of all available submission documents resulted in the additional exclusion of submissions ($n=3$). Ultimately, data on 26 EMA submissions was extracted. (b) PRISMA flow diagram showing the number of the FDA submissions included and excluded during high-level review, detailed review, and data extraction, noting reasons for exclusion. Submissions were identified through Drugs@FDA/Center for Biologics Evaluation and Research (CBER) web portal ($n=408$) as eligible for high-level review. Submissions passing high-level review for the FDA were compared with submissions passing high-level review for the EMA to capture submissions made outside of the study period (maximizing the number of submissions for direct comparison between the EMA and FDA). This yielded additional submissions eligible for FDA ($n=10$) high-level review. During high-level review, the FDA submissions were primarily excluded due to inclusion of RCT data as primary evidence. During detailed review, the FDA submissions were primarily excluded due to no explicit use of RWD analyses for supporting evidence. Of the submissions eligible for data extraction, review of all available submission documents resulted in the additional exclusion of submissions ($n=8$). Ultimately, data on 38 FDA submissions was extracted. (c) PRISMA flow diagram showing the number of HTAA submissions included and excluded during high-level review, detailed review, and data extraction, noting reasons for exclusion. Initially, 772 submissions were identified through the HTAA as eligible for high-level review. In total, 448 submissions were excluded during high-level review using the HTAA record, the majority of which were excluded due to inclusion of RCT data as primary evidence. During detailed review, 231 submissions were excluded, the primary reason for exclusion being no explicit use of RWD analyses ($n=163$). Of the submissions eligible for data extraction, review of the key documents for each HTAA body available in HTAA resulted in the additional exclusion of 17 submissions due to "no explicit use of RWD analyses for supporting evidence" and 6 submissions due to "RCT data included in submission as primary evidence." Ultimately, data on 70 HTAA submissions was extracted ($n=16$ (NICE); $n=12$ (IQWiG); and $n=25$ (G-BA); $n=17$ (HAS)). EMA, European Medicines Agency; FDA, US Food and Drug Administration; G-BA, Gemeinsamer Bundesausschuss; HAS, Haute Autorité de Santé; HTA, health technology assessment; HTAA, health technology assessment accelerator; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; NICE, National Institute for Health and Care Excellence; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized controlled trial; RWD, real-world data; SAT, single-arm trial.

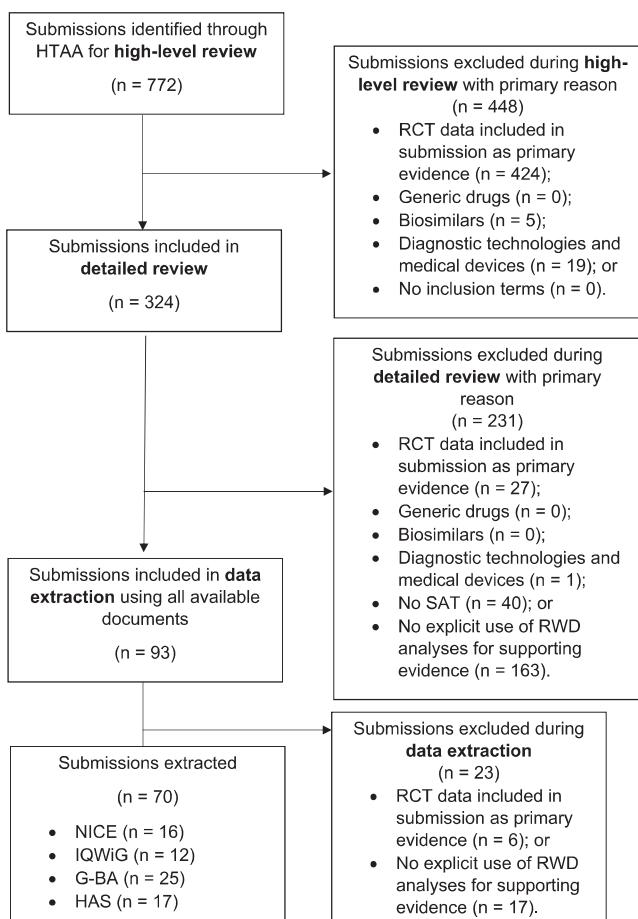
comparator or benchmark were oncology and/or hematology although neurology indications were also common for regulatory submissions (Figure 2a,b).

Among NICE and HAS submissions, the majority received a positive ($n=4/16$ (25%) (NICE); $n=13/17$ (76%) (HAS)) or positive with restrictions ($n=11/16$ (69%) (NICE); $n=3/17$ (18%) (HAS)) recommendation. Among G-BA submissions, the

percentage of submissions receiving a negative recommendation was higher ($n=7/25$ (28%)). Due to the system for reimbursement in Germany, all IQWiG submissions received no recommendation (Figure 2b).

In general, the use of uncontrolled trials with external controls in regulatory submissions increased during the study period (Figure S1A). Among the FDA submissions extracted, the most common

(c) HTA Attrition

**Figure 1** (Continued)

designations were priority review ($n = 40$), orphan drug ($n = 33$), and breakthrough therapy ($n = 25$), whereas the most common designations for the EMA submissions extracted were orphan drug ($n = 23$), conditional approval ($n = 16$), and accelerated assessment ($n = 10$; **Figure S1B**). A given submission could have multiple designations.

Among the five ($n = 1$ (orphan); $n = 4$ (non-orphan)) products that were reviewed by both IQWiG and G-BA, there was full agreement in their benefit ratings ($n = 4$ (no additional benefit); $n = 1$ (non-quantifiable benefit)).

The majority of submissions extracted were original submissions, although label extensions/expansions and resubmissions were also included (**Figure S2A,B**).

The majority of uncontrolled trials included in extracted submissions were phase II (**Figure S3A,B**).

A variety of data source types were leveraged to derive external controls for uncontrolled trials, including administrative claims, electronic medical records (EMRs) or electronic health records (EHRs), published literature, registries, and chart review. A given submission may have included multiple RWD source types. Overall, EMR or EHR data most commonly supported submissions for both the EMA and FDA (**Figure S4A**). The majority of

submissions to the NICE, IQWiG, and HAS used a literature-based external control; the majority of submissions to the G-BA used a registry-based external control (**Figure S4B**). There was a decline in the proportion of submissions using literature-based external controls between 2015 and 2018 ($n = 20/38$ (53%)) and 2019–2021 ($n = 13/32$ (41%)). If it could not be verified using publicly available information whether a literature-based estimate was derived from RWD or historical trial data, the submission was included so as to comprehensively include submissions utilizing RWD ($n = 7/11$ (64%) (FDA); $n = 1/2$ (50%) (EMA); and $n = 2/33$ (6%) (HTA)), where the denominator is the number of submissions with a literature-based analysis.

External controls without any direct comparison (e.g., a reported relative effect estimate (hazard ratio (HR)/ odds ratio (OR)) and corresponding p value) were classified as benchmark whereas those with a direct comparison were classified as comparative. Across all RWD analyses in the qualifying submissions, submissions to the FDA predominantly used RW benchmark data rather than RW comparative data, whereas RW benchmark data was used about as often as RW comparative data when submitted to the EMA (**Figure S5A**). Similar to the FDA, across all RWD analyses in the qualifying submissions, submissions to HTA bodies predominantly used RW benchmark data (**Figure S5B**) although the proportion decreased between 2015 and 2018 ($n = 27/38$ (71%)) and 2019–2021 ($n = 18/32$ (56%)). Further, the majority ($n = 24/45$ (53%)) of benchmark analyses were derived from literature-based data vs. 35.7% ($n = 10/29$) of comparative analyses submitted.

This review identified 12 products (13 submissions) that qualified for both the EMA and FDA extraction (all generic names): idecabtagene vicleucel, avapritinib, belantamab mafodotin-blmf, blinatumomab, cerliponase alfa sodium, fam-trastuzumab deruxtecan-nxki, emapalumab-lzsg, selumetinib, asfotase alfa, axicabtagene ciloleucel, and onasemnogene abeparvovec. In the submissions reviewed by both the EMA and FDA, RWE was considered in the final decision in a comparable percentage of submissions ($n = 8/13$ (62%) (FDA); $n = 7/13$ (54%) (EMA); **Table 1**). There were 18 products (24 submissions) that qualified for extraction from at least 2 HTA bodies (all generic names): avapritinib, avelumab, blinatumomab, cerliponase alfa, pembrolizumab, tisagenlecleucel, cemiplimab, trametinib, metreleptin, adexanet alfa, nivolumab, entrectinib, midostaurin, asfotase alfa, autologous anti-CD19-transduced CD3+ cells, axicabtagene ciloleucel, allogeneic genetically modified T cells, and ceritinib. In the submissions reviewed by at least two HTA bodies, RWE was considered in the final decision most frequently by NICE ($n = 7/24$ (29%)) (NICE); $n = 1/24$ (4%) (IQWiG); $n = 5/24$ (21%) (G-BA); $n = 1/24$ (4%) (HAS)). Data were extracted from three submissions to the EMA that were either refused or withdrawn (**Table 2**). Negative recommendations were given to nine HTA submissions (8 products; $n = 1$ (NICE); $n = 7$ (G-BA); $n = 1$ (HAS); **Table 3**).

Regulator and HTA feedback

Regulators and HTA bodies provided feedback on various operational and methodological aspects of submissions ranging

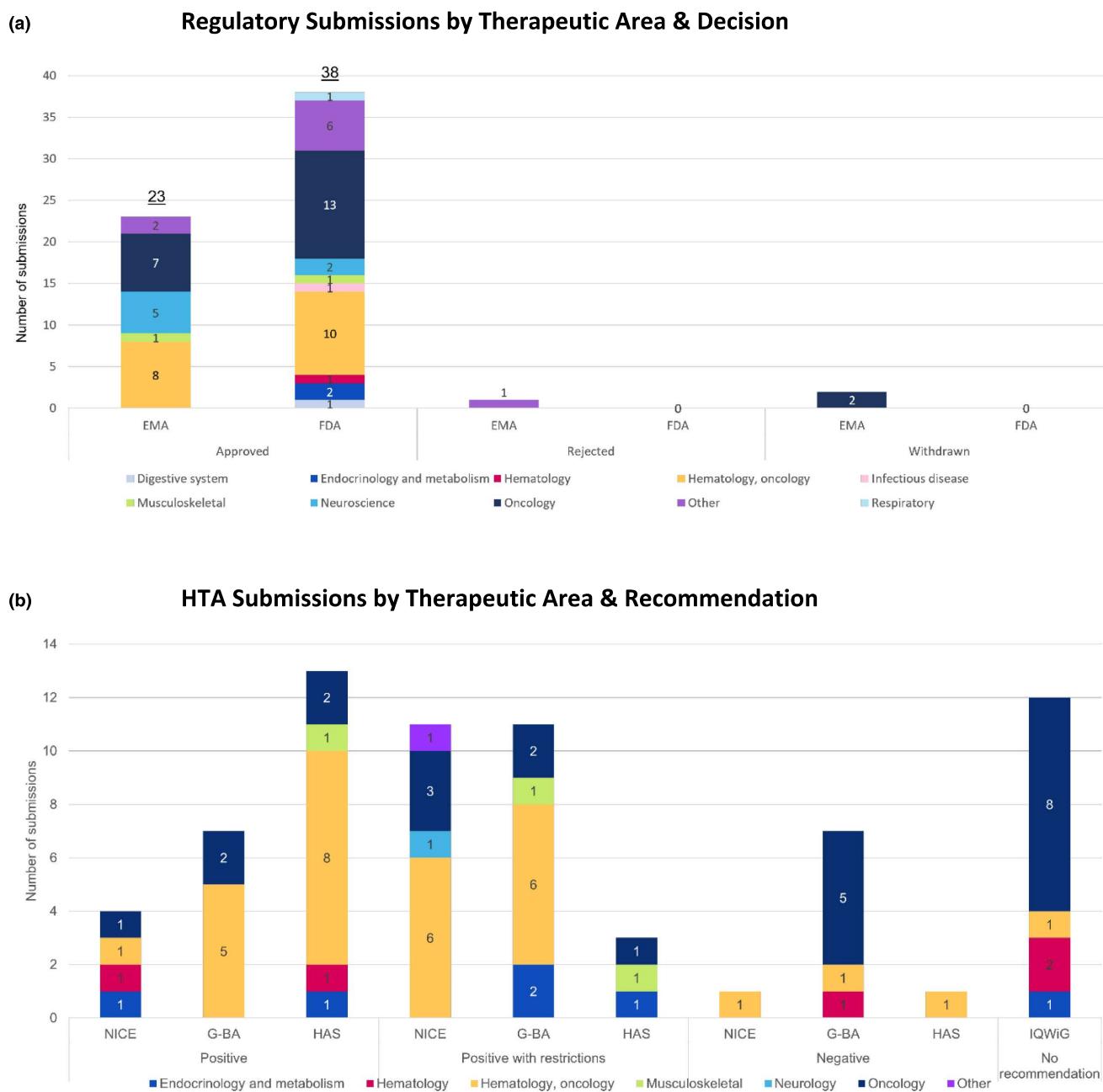


Figure 2 (a) Bar graph showing the number of submissions to EMA and FDA by decision (approved, refused, withdrawn) and colored by therapeutic area. For both EMA and FDA, the most common therapeutic areas for submissions utilizing an uncontrolled trial with an RWD-derived external comparator or benchmark were oncology, hematology, and neurology. (b) Bar graph showing the number of submissions to HTA by recommendation (positive, positive with restrictions, negative, and none) and colored by therapeutic area. Among NICE and HAS submissions, the majority received a positive ($n=4/16$ (25%) (NICE); $n=13/17$ (76%) (HAS)) or positive with restrictions ($n=11/16$ (69%) (NICE); $n=3/17$ (18%) (HAS)) recommendation. Among G-BA submissions, the percentage of submissions receiving a negative recommendation was higher ($n=7/25$ (28%)). All IQWiG submissions received no recommendation. Across all HTA bodies, the most common therapeutic areas for submissions utilizing an uncontrolled trial with an RWD-derived external comparator or benchmark were oncology and/or hematology. Submissions to IQWiG did not receive a recommendation. EMA, European Medicines Agency; FDA, US Food and Drug Administration; G-BA, Gemeinsamer Bundesausschuss; HAS, Haute Autorité de Santé; HTA, health technology assessment; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; NICE, National Institute for Health and Care Excellence; RWD, real-world data.

from bias to cohort comparability, and characterization of standard of care (Table S2). Among these recommendations, feedback related to endpoint definitions and missing data, as well as documentation of early engagement, was particularly

important (Table 4). Both regulatory agencies cautioned against the use of time-to-event endpoints but did not provide guidance on preferred alternatives or methods for more effectively assessing such endpoints, whereas HTA bodies were more

Table 1 Summary of products submitted to both the EMA and FDA

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RW endpoints	Key FDA feedback	RWE consideration in decision		
				FDA	EMA	HTA
• Binatumomab • Angen • Hematology, oncology ^{5,76}	• Historical comparator study (20120310). • Retrospective pooled analysis of historical data available from 1990 to 2014 (US and EU).	• FDA/EMA: Primary: CRR Secondary: OS	• External control conducted in a large number of patients ($n=1,139$). • Accounted for differences in patient characteristics between studies.	• Confirmation from a phase III comparative study was needed in order to better quantify the magnitude of the effect, in particular with respect to time-related endpoints.	• Key HTA feedback	N/A
• Afstafase alfa • Alexion Pharmaceuticals • Musculoskeletal ⁷⁷⁻⁸¹	• Natural history study (ENB-011-10). • Retrospective chart review and retrospective cohort study. • ALX-HPP-502 based on medical records (FDA and HTA)	• Primary: OS • Secondary: Invasive ventilator-free survival time	• Engagement with FDA at various stages of the development program. • The historical comparison resulted in a high risk of selection bias. • Potential for confounding due to baseline imbalances in latent variables. • Conducted in a small number of patients for both patient populations ($n=48$; $n=32$). • Due to the scarcity of the disease population overall, the limitations of the studies, and the post hoc analysis approaches for both patient populations, the determination of clinical effectiveness relied more on clinical judgment than on the statistical rigor usually required for larger RCTs.	• Improvements in supportive care and technologies hampered direct comparison with clinical efficacy data generated by the external control. • The use of external controls was accepted on the grounds that: – Historical patients had received precisely defined standard treatment that is the same for the trial patients. – The method of treatment evaluation was the same. – Historical patient characteristics were sufficiently comparable to the trial patients. – Management was done by same organization and (largely the) same investigators. – No other issues making one expect differing results compared with current trial group were apparent.	• The observed difference for the endpoint (i.e., OS) was too large to be explained by the effect of confounders alone. • The historical comparison resulted in a high risk of bias. • There were considerable limitations in the comparability of data because of different data collection for the endpoint OS. • The studies were conducted at different time periods between which the symptomatic treatment and diagnosis had changed. • Sensitivity analyses for OS were used to confirm whether the result of the primary analysis for this endpoint were robust.	Yes Yes N/A

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWD endpoints	RWE consideration in decision					
			FDA	EMA	HTA	FDA	EMA	HTA
• Defibrotide Sodium • Jazz Pharmaceuticals • Hematology ^{22–34}	• Center for International Blood and Marrow Transplant Research	<p>FDA:</p> <ul style="list-style-type: none"> • Primary: Survival at day 100. • Secondary: Time to death, time from HSCT to end of follow-up, time from VOD to end of follow-up. <p>EMA:</p> <ul style="list-style-type: none"> • Primary: CR • Secondary: Survival rate, time to CR. 	<ul style="list-style-type: none"> • The sponsor's use of the MRC, an independent review group, strengthened their argument that the external control group was a valid comparator. • External control conducted in a small number of patients ($n=32$). • The small sample size in the external control cohort makes the PS stratified analysis difficult to interpret. • The timeframe for the external control cohort differed from that of trial cohort. • Many of the potential prognostic factors that might confound the evaluation of temporal effect were not collected in the external control cohort. • The significance level could not be determined due to many unplanned adaptations, e.g., sample size reduction and planned/unplanned interim analyses. 	<ul style="list-style-type: none"> • The MRC process for selecting subjects into the external control group was not transparent; the rationale and evidence to support the decision was not consistently documented. • The external control group could not be accepted as an objective and reliable control group in the clinical trial setting. • The repeated alteration of the composition of the external control group raised the risk that the finally chosen subset of external control patients had more severe characteristics than the trial group. • Matching was deemed adequate in terms of diagnosis and baseline variables. • There did not appear to be systematic differences between the populations. • Sensitivity analyses were used for handling missing data. • Serious methodological flaws in the study made it extremely difficult to quantify the benefits. 	<ul style="list-style-type: none"> • N/A 	<ul style="list-style-type: none"> • Yes • No • N/A 		

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWD endpoints	Key FDA feedback	Key EMA feedback	RWE consideration in decision		
					FDA	EMA	HTA
Axicabtagene ciloleucel • Kite Pharma • Hematology, oncology ^{85–89}	Global, pooled retrospective study (SCHOLAR-1). Data from 2 RCTs and 2 observational databases.	FDAs: Response and survival EMA: RR, CRR, and OS. HTA: ORR and OS.	Minimal feedback related to use of RWE.	Trial group had a larger proportion of patients with worse disease stages and higher numbers of patients who already received chemotherapy.	Possible selection bias. Temporally different collection of the data of trial compared with external control. Uncertainties regarding the comparability of patient populations.	No	Yes
Cerliponase alfa • Biomarin • Neurology ^{90–93}	Natural history study (190–901). Retrospective analysis of patients from the DEM-CHILD database (US and Europe).	RR	The trial used additional rating guidelines and training for endpoint ascertainment that could not be implemented in the external control due to the mostly retrospective design. The primary statistical reviewer at FDA performed sensitivity analyses. Differences in patient characteristics between the trial and the untreated external control were identified. The schedule of assessments was different in the two studies and thus they did not have the same time points to compare.	The sponsor applied acceptable methods to account for potential bias and provided several sensitivity analyses that support the robustness of the findings. The external control was not concurrent. Different patient information was gathered in each cohort so that only a restricted number of items reflect the data from the pooled cohort.	Lack of comparative data on QoL and side effects. The data available from only two centers of the registry and the selection of the evaluable population may have led to selection effects. Mean age at diagnosis, sex, and genotype distribution was not balanced between the trial and external control groups. It was not possible to compare the study population with regard to concomitant medication or comorbidities because these data were not available for the patients in the external control. An estimate of the frequency of adverse events was limited due to the lack of a concurrent control group. The populations across the studies were generalizable to patients seen in clinical practice in England. ⁹⁴	Yes	Yes

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWD endpoints	Key FDA feedback	Key EMA feedback	Key HTA feedback	RWE consideration in decision		
						FDA	EMA	HTA
• Blinatumomab • Angen • Hematology, oncology ^{94–96}	• Study 20120148 • Retrospective cohort study from Czech Republic, France, Germany, UK, Italy, Poland, Spain, and Russian study group data bases	FDA/EMA: • Primary: RFS • Secondary: • OS FDA: • Secondary: • Mortality HTA: OS and PFS	• There was a limited amount of missing data in the covariate set for two variables contained missing values). • All major efficacy and safety analyses were reproduced by reviewers. Conclusions were limited due to confounding by inappropriate data matching, lack of matching for covariates that would affect the RFS endpoint, inclusion of patients with incomplete hematological recovery, and lack of comparability in the duration of follow-up. • Potentially insufficient power to detect a clinically meaningful difference. Although the number of patients was small, the results were consistent across the subgroups analyzed. • The data were not contemporaneous. • In the absence of the information on disease assessment periods and frequencies, whether or not RFS is adequately assessed between treatment arms was unclear.	• Clinical Endpoints such as RFS and OS were not appropriately assessed. • Results of the endpoint analyses were largely consistent across analysis sets and sensitivity analyses. • Exploratory analyses with alternative PS weightings suggested a more pronounced treatment effect for OS in favor of the treatment. • The use of PS adjustments was successful in creating a balanced population with respect to numerous important baseline covariates.	• Selection of study centers and inclusion/exclusion criteria for patients were unclear. • No data were available on type and duration of prior therapies, standardization of data collection, and extent and management of missing values. • Uncertainties around operationalization of endpoints. • The patient population differed from that of the trial regarding proportion of patients in relapse. • The selection of the adjustment factors taken into account for the analysis was not sufficiently justified.	Yes	No	No
• Enapalumab-lzsg • Swedish Orphan Biologics AB (EMA); Novartime S.A. (FDA) • Immunology ^{97,98}	• Literature based historical control • rate (FDA) • NI-0501-HC (EMA)	FDA: ORR • EMA: OS and post-HSCT survival	• Minimal feedback related to use of RWE.	• Uncertainties pertaining to N/A efficacy. Minimal feedback related to use of RWE.	N/A	No	No	N/A

(Continued)

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWD endpoints	Key FDA feedback	Key EMA feedback	Key HTA feedback	RWE consideration		
						FDA	EMA	HTA
• Onasemnogene abeparvovec • Novartis Gene Therapies EU Limited (EMA), AveXis, Inc (FDA) ^{99–101} • Neurology ^{99–101}	<ul style="list-style-type: none"> Literature based natural history study using data from the PNCR and NeuroNext databases (FDA and EMA) Literature based natural history study using data from the PNCR database (HTA) 	<ul style="list-style-type: none"> Primary: Sitting without support and survival at 14 months of age. Secondary: Independent of ventilatory support 	<ul style="list-style-type: none"> Minimal feedback related to use of RWE. 	<ul style="list-style-type: none"> Patients in the external control cohort had less severe disease. Uncertainties about long-term efficacy. 	<ul style="list-style-type: none"> The external control and trial patients did not seem comparable in regard to age and disease severity. 	Yes	Yes	Yes

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RW endpoints	Key FDA feedback	Key EMA feedback	RWE consideration in decision		
					FDA	EMA	HTA
• Belantamab mafodotin-blmf • Glaxo Smith Kline • Hematology, oncology ^{102–104}	• Literature-based	• FDA/EMA: ORR • HTA: OS	• Biases, including variations in the eligibility criteria, clinical practice, and SoC, in addition to temporal effects, made it difficult to interpret time-to-event endpoints from SATs (e.g., PFS and OS). • Endpoints were unlikely to provide evidence to support efficacy because the measurements in isolation are out of context and comparisons to external controls are prone to bias. • These biases may have been avoided if there was a concurrently sampled and randomly assigned control group.	• Because there was no standard treatment for these patients, the missing control arm was considered acceptable. • SAT settings impair the causality assessment of several key unfavorable effects leading to remaining uncertainties. • The selection of the adjustment and matching factors taken into account for the indirect comparisons was not justified.	• Results for only one patient relevant endpoint were not sufficient. • Lack of specific information on the type and scope of “conventional treatment” in the external control patients. • Due to the lack of information, it was unclear to what extent the included patient populations were comparable between the included studies.	• Results for only one patient relevant endpoint were not sufficient. • Lack of specific information on the type and scope of “conventional treatment” in the external control patients. • Due to the lack of information, it was unclear to what extent the included patient populations were comparable between the included studies.	No No No
• Trastuzumab deruxtecan • Daiichi Sankyo Inc Oncology ^{105–107}	• Database (Unicancer study; French network of 18 private, non-profit hospitals) • HTA: • PFS	• FDA/EMA; • Primary: ORR • Secondary: PFS and OS	• FDA agreed with the sponsor's summary of the study's efficacy endpoints; however, because there was not a comparator arm, time-to-event endpoints (e.g., PFS, time to response) results were uninterpretable.	•保守性敏感性分析 •评估响应的不确定性 •由于图像质量差，外部对照对象可能不匹配。 •外部对照选择基于后基线变量，可能引入选择偏倚。 •有效性评估 •评估依赖于治疗效果的演示，可能因重要 prognostic factors 而异质。	• Included and excluded covariates introduced some uncertainty in the estimated treatment effect. • The primary endpoint of PFS was assessed differently in the main analyses in the studies. • No safety perspective for these treatments was proposed in these indirect comparisons.	•保守性敏感性分析 •评估响应的不确定性 •由于图像质量差，外部对照对象可能不匹配。 •外部对照选择基于后基线变量，可能引入选择偏倚。 •患者在外部对照试验中可能因重要 prognostic factors 而异质。	No No No No No No

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWE endpoints	Key FDA feedback	Key EMA feedback	RWE consideration in decision		
					FDA	EMA	HTA
• Selumetinib • AstraZeneca • Oncology ^{108,109}	• Natural history study (NCI-08-C-0079) from 1 US study site and the placebo arm from Part A of the NCI Study 01-C-0222	• Primary: PN growth rate • Secondary: PFS	• Patient baseline disease characteristics were generally similar. • Heterogeneity in the patient population in measured and unmeasured factors (e.g., disease characteristics, patient characteristics, treatment characteristics).	• Potential subjectivity in the assessment of the imaging data in the external control studies as data had not been assessed by ≥2 independent readers who were blinded for patient exposition and time point, also avoiding sequential presentation of images from the same patient, as was recommended in the Protocol Assistance.	Yes	Yes	N/A
				• Differences between the design of the studies (e.g., eligibility criteria, baseline demographic and disease characteristics, and timing of imaging). • Heterogeneity of the baseline populations (e.g., age, disease progression, tumor volume, comorbidities). • Trial conducted in a small number of patients ($n=50$).			

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RW endpoints	Key FDA feedback	Key EMA feedback	RWE consideration in decision		
					FDA	EMA	HTA
• Idecabtagene vicleucel • Celgene Corporation	• Literature based historical ORR • EMA: Global non-interventional retrospective comparative study (NDS-MM-003)	FDA/EMA: • Primary: ORR EMA: • Secondary: VGPR rate, CRR, TIR, OS and PFS	• The FDA clinical review team re-adjudicated the response assessments.	• Large proportion of missing data for some included covariates and several covariates excluded from the PS model due to >30% missing data.	Yes	Yes	N/A
• Hematology, oncology ^{10,111}			• Results for the secondary endpoints were consistent with the primary endpoint.	• Results compared favorably to those observed in the matched external control as well as those reported in the literature.			
			• Results compared favorably to those observed in the matched external control as well as those reported in the literature.	• The interpretation of time-to-event endpoints is intrinsically limited.			
			• The representativeness of the external control rate for the ORR cutoff defining study success was uncertain.	• The patient population enrolled is highly selected.			
			• Trial conducted in a small number of patients ($n=140$), which limited interpretation of subgroup analyses.	• Long time period (up to 60 days from the index date) allowed for the collection of baseline data.			
			• Overlapping recruitment periods for the EC and the trial at the same study centres. ¹⁰	• Overlapping recruitment periods for the EC and the trial at the same study centres.			
			• Short duration of follow-up for OS.	• Short duration of follow-up for OS.			
			• The true magnitude of the treatment effect could not be reliably ascertained.	• The true magnitude of the treatment effect could not be reliably ascertained.			

(Continued)

Table 1 (Continued)

Product, ^{a,b} sponsor, and therapeutic area	Primary RWD source	RWD endpoints	Key FDA feedback	Key EMA feedback	RWE consideration		
					FDA	EMA	HTA
• Avapritinib • Blueprint Medicines Oncology ^{112–114}	• Retrospective natural history study (BLU-285-1002) using data collected from 22 patients at 3 study centers in the US • Retrospective literature-based studies (EU and Asia)	FDA: • Primary: ORR • Secondary: DOR EMA: • Primary: ORR, DOR, and PFS • Secondary: OS HTA: OS and PFS	• The large tumor burden was comparable between study cohorts. • Minimal feedback related to use of RWE.	• External control conducted in a small number of patients ($n=22$). • Treatment showed an outstanding and durable ORR regardless of prior line of therapy, which was unprecedented in a population subset that traditionally is unresponsive.	• The comparison in the form of a time-to-event analysis was not adequate. • The PS-adjusted indirect comparison between the trial and external control was not considered due to the choice of the starting point of the observation period for the time-to-event analysis in the external control.	No	No
					• There were further limitations regarding the description of the selection of the study population and the endpoints as well as the definition of the inclusion and exclusion criteria.	No	No

CR, complete response; CRR, complete response rate; DCR, duration of complete response; DoR, duration of response; ECOG, Eastern Cooperative Oncology Group; EFS, event free survival; EMA, European Medicines Agency; EU, European Union; FDA, US Food and Drug Administration; HAS, Haute Autorité de Santé; HSCT, hematopoietic stem cell transplantation; HTA, Health Technology Assessment; IPI, International Prognostic Index; MRC, medical research council; NCI, National Cancer Institute; NHS, National Health Service; ORR, objective response rate; OS, overall survival; PFS, progression free survival; PN, plexiform neurofibromas; PNCR, Pediatric Neuromuscular Clinical Research Database; PS, propensity score; QoL, quality of life; RCI, randomized controlled trial; RFS, relapse-free survival; RR, response rate; RWD, real-world data; RWE, real-world evidence; SAI, single-arm trial; SoC, standard of care; TTP, time to progression; TTR, time to response; UK, United Kingdom; US, United States; VOD, veno-occlusive disease; VGPR, very good partial response.
^aThe submission type of all products to both the FDA and EMA is original, with the exception that the submission type of blinatumomab to both the FDA and EMA is label expansion and idecabtagene vicleucel to the FDA is resubmission. ^bProducts are ordered by year of FDA regulatory decision (from earliest [top] to most recent [bottom]).

Table 2 Summary of refused/withdrawn EMA products

Product	Indication	Sponsor	Agency	Decision date	Primary RWD source	RW endpoints	Feedback	Negatives	Key drivers of refusal/withdrawal
emapalumab ⁹⁷	Pediatric patients aged under 18 years with primary HLH	Swedish Orphan Biovitrum AB	EMA	12-Nov-20	Dataset developed in collaboration with the EU and US primary investigators for clinical trial Study NI-0501-04)	• OS • Post-HSCT survival	<ul style="list-style-type: none"> • Lack of strict inclusion criteria. • Transition from exploratory to pivotal trial (resulting in potentially data driven decisions such as the introduction of less strict primary endpoints). • Selection methods for EC not considered appropriate because selection was concurrent with the trial by the same investigators. 		Inappropriate comparator in the clinical study. Uncertain clinical benefit.
allogeneic T cells genetically modified with a retroviral vector encoding for a truncated form of the human low affinity nerve growth factor receptor (Δ LNGFR) and the herpes simplex I virus thymidine kinase (HSV-TK Mut2) ¹¹⁵	Adjuvantive treatment in haploidentical HSCT of adult patients with high-risk hematological malignancies	MolMed	EMA	23-Jun-16	EBMT database	• OS • LFS • NRM • RI	<ul style="list-style-type: none"> • Inherent bias related to the number of patients treated with trial drug during the clinical development. • The impact of potential differences in baseline characteristics not included in the matching strategy (e.g., median year of transplant, stem cell source) is unknown. • Start of calculation of survival time different between treatment arms. • Use of the EMBT database for obtaining control patients and the matching strategy reduced the uncertainty of the requested cross study comparison. • Results from confirmatory Phase III study were expected to address the above uncertainties. 	Inappropriate comparator in the clinical study. Sponsor withdrawal due to Phase III clinical trial results.	

(Continued)

Table 2 (Continued)

Product	Indication	Sponsor	Agency	Decision date	Primary RWD source	RW endpoints	Feedback	Negatives	Key drivers of refusal/withdrawal
ivosidenib ¹¹⁶	Monotherapy for the treatment of relapsed or refractory AML with an IDH1 R132 mutation in adult patients who:	Agios Netherlands	EMA	16-Oct-20	AMLSG & EU RWD datasets	• OS	<ul style="list-style-type: none"> Different starting points used to determine OS; the AMLSG cohort OS was calculated from the time of the most recent determination of relapsed or refractory AML status, while the trial cohort results present OS measured from the time of treatment initiation. SAT amended 6 times, affecting multiple fundamental aspects of the study design (e.g., study population, posology of design, schedule of assessment, statistical methodology, endpoints, sample size, etc.). Time-to-event endpoints (i.e., OS) difficult to interpret because numerous factors involved in survival times cannot be controlled in absence of randomization. Health-related QoL could not be compared due to the SAT design; health-related QoL measures particularly important in a disease such as AML. Differentiation between survival curves early in follow-up indicative of selection bias. Using factors whose values are determined after beginning treatment for matching is problematic because these may be influenced by treatment. 	<ul style="list-style-type: none"> Treatment comparison not well conducted. Lack of QoL data. 	<ul style="list-style-type: none"> Sponsor withdrawal due to insufficient data to conclude a positive benefit-risk balance.

AML, acute myeloid leukemia; AMLSG, AML Study Group Registry; EBMT, European Group for Blood and Marrow Transplantation; EC, external control; EMA, European Medicines Agency; EU, European Union; HH, hemophagocytic lymphohistiocytosis; HSCT, hematopoietic stem cell transplantation; LFS, leukemia-free survival; NRM, non-relapse mortality; OS, overall survival; QoL, quality of life; RI, relapse incidence; RWD, real-world data; SAT, single-arm trial; US, United States.

Table 3 Summary of Negative HTA Recommendations

Product	Indication	Sponsor	HTA body	Decision date	Primary RWD source	RW endpoints	Feedback	Clinical Negatives	Economic Negatives	Key drivers of negative recommendation
trametinib ¹¹⁷	Adult patients with advanced NSCLC with a BRAF V600 mutation (stage IIIb and IV)	Novartis	G-BA	19-Oct-17	Literature based (Cardarella 2013, Ding 2017, NGM Cologne 2017)	• OS	• Data submitted on OS had limited significance due to low patient numbers.	Population • Not providing efficacy data for the target subgroup.	• Not specified.	• Insufficient clinical data.
dabrafenib in combination with trametinib ¹¹⁷	Adult patients with advanced NSCLC with a BRAF V600 mutation	Novartis	G-BA	19-Oct-17	Literature based (Cardarella 2013, Ding 2017, NGM Cologne 2017)	• OS	• Data submitted on OS had limited significance due to low patient numbers. • The identified study populations were insufficiently comparable. Patients in the trial differed from patients in the retrospective studies with regard to ECOG status and a better prognosis.	Population • Not providing efficacy data for the target subgroup. • Comparator • Inappropriate comparator in the clinical study. • Not providing relative data against the appropriate comparator.	• Not specified.	• Insufficient clinical data. • Uncertain clinical benefit.
pembrolizumab ¹¹⁸	Adult patients with relapsed or refractory cHL who have failed ASCT and BV, or who are transplant-ineligible and have failed BV	Merck & Co. / MSD	G-BA	17-Nov-17	GHSG registry data	Unclear based on documents reviewed	• GHSG registry data was not suitable, as the patients were not pretreated with BV. • As no relevant data was submitted, an added benefit for pembrolizumab could not be demonstrated.	Population • Not providing efficacy data for the target subgroup. • Comparator • Not providing relative data against the appropriate comparator.	• Not specified.	• Inappropriate clinical trial design.

(Continued)

Table 3 (Continued)

Product	Indication	Sponsor	HTA body	Decision date	Primary RWD source	RW endpoints	Feedback	Clinical Negatives	Economic Negatives	Key drivers of negative recommendation
allogeneic genetically modified T cells ¹¹⁹	Adjuvantive treatment in haploididentical HSCT of adult patients with high-risk hematological malignancies	Dompé Pharmaceuticals	HAS	19-Jan-19	EBMT registry	Unclear based on documents reviewed	The characteristics of trial patients differed from subjects enrolled in the EBMT group in regard to: <ul style="list-style-type: none">Date of most recent HSCT in the EBMT group vs the trial group (2011 versus 2007 respectively).Mean follow-up times.Proportion of female donor/male user.Origin of HSC.Proportion of in vivo T-cell depletion	Outcomes <ul style="list-style-type: none">Lack of hard outcomes.Lack of QoL data.	<ul style="list-style-type: none">Not specified.	<ul style="list-style-type: none">Uncertain clinical benefit.
idelalisib ¹²⁰	Follicular lymphoma that has not responded to 2 prior lines of treatment in adults	Gilead Sciences	NICE	2-Oct-19	UK HMRN registry	<ul style="list-style-type: none">PFSOS	<ul style="list-style-type: none">The analyses for PFS did not censor patients at transplantation which may have biased the results.The company did not provide a rationale for its choice of PS matching method, nor had it conducted sensitivity analyses using alternative methods.Analyses using PS matching were associated with high levels of uncertainty.The cost-effectiveness analysis had many limitations.It is unclear whether idelalisib was better than individual chemotherapeutic regimens currently offered by the NHS and, if so, by how much.	<ul style="list-style-type: none">Not providing efficacy data for the target subgroup.Comparator<ul style="list-style-type: none">Not providing relative data against the appropriate comparator.	<ul style="list-style-type: none">Model horizon inappropriate.Not cost-effective.	<ul style="list-style-type: none">Insufficient clinical benefit.Non-robust economic analyses.Not cost-effective.

(Continued)

Table 3 (Continued)

Product	Indication	Sponsor	HTA body	Decision date	Primary RWD source	RW endpoints	Feedback	Clinical Negatives	Economic Negatives	Key drivers of negative recommendation
andexanet alfa ¹²¹	Adult patients treated with a direct FXa inhibitor (apixaban or rivaroxaban) when reversal of anti-coagulation is needed due to life-threatening or uncontrolled bleeding	Portola Pharmaceuticals	G-BA	20-Feb-20	RETRACE-II German registry	<ul style="list-style-type: none"> Haematoma enlargement Occurrence of intracranial and extracranial complications during hospitalization Mortality before leaving hospital or after three months Neurological functionality after three months 	<ul style="list-style-type: none"> The patient characteristics of the studies used for the PS adjusted comparison did not show sufficient similarity, especially with regard to the severity of intracerebral bleeding. The subpopulation selected for the comparison did not cover all patients included in the approved andexanet alfa indication. The effect was not large enough to be explained by systematic bias alone. No statistically significant difference between the two groups in the endpoints neurological function and mortality. No comparative data on side effect endpoints are available. Further information on the interventions carried out, e.g., on concomitant medication, local and intensive care measures, was missing, so that no information was available on how these patients were treated. 	<ul style="list-style-type: none"> Lack of direct comparator. Not providing relative data against the appropriate comparator. 	<ul style="list-style-type: none"> Not specified. 	Inappropriate comparator.
cemiplimab ¹²²	Adult patients with metastatic or locally advanced cutaneous squamous cell carcinoma who are not candidates for curative surgery or curative radiation	Sanofi	G-BA	2-Jun-20	Literature based (Hillen 2018)	<ul style="list-style-type: none"> OS Disease status ORR DoR Time to progression 	<ul style="list-style-type: none"> Comparisons for the patient-relevant outcomes, health-related QoL, and adverse events were not presented. The patient populations differed in some characteristics. Due to different observation periods, OS could not be meaningfully compared. The effect estimates for OS were not sufficiently large to be caused by systematic bias alone 	<ul style="list-style-type: none"> Population Not providing efficacy data for the target subgroup. Comparators. Lack of direct comparator. Not providing relative data against the appropriate comparator. Outcomes 	<ul style="list-style-type: none"> Not specified. 	<ul style="list-style-type: none"> Inappropriate clinical trial design. Insufficient clinical data. Lack of hard outcomes. Lack of Ool data. No clinical evidence submitted.

(Continued)

Table 3 (Continued)

Product	Indication	Sponsor	HTA body	Decision date	Primary RWD source	RW endpoints	Feedback	Clinical Negatives	Economic Negatives	Key drivers of negative recommendation
entrectinib ¹²³	ROS1-positive, advanced NSCLC not previously treated with ROS1 inhibitors Adult and pediatric patients aged 12 years and older with solid tumors expressing a NTRK gene fusion, – who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and – who have not received a prior NTRK inhibitor – who have no satisfactory treatment options	Roche	G-BA	18-Feb-21	• US Flatiron Health database • Literature based	• OS • PFS	<ul style="list-style-type: none"> Transferability of the data from the Flatiron Health database (US) to the German health care context was questionable due to structural differences in the health care systems. The results from the indirect comparison were subject to uncertainty due to the lack of randomization. The observed effects were not large, meaning that a systematic bias could not be excluded. The comparative analysis included patients who received a dosage not compliant with the marketing authorization. 	<ul style="list-style-type: none"> Comparator Indirect treatment comparison not well conducted. Lack of direct comparator. Not providing relative data against the appropriate comparator. 	<ul style="list-style-type: none"> Not specified. 	<ul style="list-style-type: none"> Inappropriate clinical trial design. Insufficient clinical data.

AML, acute myeloid leukemia; AMLSG, AML Study Group Registry; ASCT, autologous stem cell transplant; BV, brentuximab vedotin; cHL, classical Hodgkin lymphoma; CRR, complete response rate; DoR, duration of response; EBMT, European Group for Blood and Marrow Transplantation; EC, external control; ECOG, Eastern Cooperative Oncology Group; FXa, factor Xa; G-BA, Gemeinsamer Bundesausschuss (Federal Joint Committee); GHSG, German Hodgkin Study Group; HAS, Haute Autorité de Santé; HMRN, Hematological Malignancy Research Network; HSC, hematopoietic stem cell; HSCT, hematopoietic stem cell transplantation; NHS, National Institute for Clinical Excellence; NSCLC, non-small cell lung cancer; NTRK, neurotrophic tyrosine receptor kinase; ORR, objective response rate; OS, overall survival; PFS, progression free survival; PS, propensity score; QoL, quality of life; RWD, realworld data; US, United States.

Table 4 Spotlight on methodological & operational aspects of interest

Spotlight: Methodological & operational aspects

Early Engagement	<ul style="list-style-type: none"> EMA submissions more frequently documented early engagement (i.e., Scientific Advice/Protocol Assistance) than FDA and HTA submissions Regulatory input during early engagement most often pertained to RW endpoints, RW data source, and acceptability of SAT design Early engagement by sponsors likely differs by HTA body depending on whether the feedback is binding (i.e., G-BA) or non-binding (e.g., NICE, HAS) Consultation from HTA bodies most often pertained to economic modeling, standard of care/appropriate comparator selection, and/or endpoint selection
Real-World Endpoints	<ul style="list-style-type: none"> The analyses of PFS and OS can be difficult to interpret without a concurrent control Time-to-event endpoints can be difficult to interpret if factors related to survival time cannot be controlled The endpoint rwORR may not be comparable to ORR as assessed in a clinical trial, and cross-trial comparisons of time-to-event endpoints may not be valid Large magnitude of effect may outweigh limitations in the context of a very rare disease that is rapidly fatal with no other therapies known to improve survival If observed effects are not large, systematic bias cannot be excluded Operationalization of endpoints should be clearly defined HTA emphasized patient relevant endpoints, comparative data on side effects, and HRQoL
Missing Data	<ul style="list-style-type: none"> Biases in endpoint estimates can result from missing data if more missing data are expected to occur in EC patients than in treated patients Matching may be limited by the large proportion of missing data for some included covariates and several covariates excluded from the PS model due to a large percentage (e.g., >30%) of missing data It is likely that differences in patient populations at baseline cannot be fully assessed if there is incomplete baseline covariate data Incomplete capture of death information in controls Sponsors should summarize differences in baseline characteristics and outcomes between patients for whom a given covariate value is known and patients for whom a given covariate was not captured Imputation was generally conducted as a sensitivity analysis (rather than primary analysis) by sponsors and/or regulators

EC, external control; EMA, European Medicines Agency; FDA, Food and Drug Administration; G-BA, Gemeinsamer Bundesausschuss; HAS, Haute Autorité de Santé; HRQoL, health-related quality of life; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; ORR, objective response rate; OS, overall survival; PFS, progression free survival; PS, propensity score; RW, real-world; rwORR, real-world objective response rate; SAT, single-arm trial.

concerned with the inclusion of patient relevant quality of life (QoL) endpoints. Regulators and HTA bodies also raised concerns regarding missing data, although clear feedback related to the most appropriate method for handling missing data was lacking. Imputation was not documented for the majority of submissions to the FDA ($n = 6/38$ (16%)) or EMA ($n = 3/26$ (12%)) and, when documented, was often used as a sensitivity analysis to address missingness in the trial data (rather than the RWD). No imputation was recorded for the included HTA submissions. Although minimal replication of results or additional analyses by the EMA were explicitly mentioned, the FDA documented conducting their own analyses in 76% of submissions ($n = 29/38$). NICE was the only HTA agency that documented conducting their own analyses ($n = 3/16$ (19%)), the majority of which were by Evidence Review Groups (ERGs). NICE had the most frequently documented early engagement by sponsors ($n = 4/16$ (25%)). Early engagement with regulators was documented more frequently for EMA submissions ($n = 25/26$ (96%)) than FDA submissions ($n = 19/38$ (50%)); however, pre-new drug application (NDA)/pre-biologics license application (BLA) meetings with the FDA were still utilized by sponsors. The extent to which sponsors were able to implement advice received during early engagement remains unclear.

DISCUSSION

This systematic review is the first to compare regulatory and HTA feedback from multiple jurisdictions on recent submissions utilizing external controls across all therapeutic areas. Five themes emerged: (1) the feedback and extent to which RWE was considered in final decisions by the FDA, EMA, and HTA bodies was highly variable even when the agencies considered the same RWE for the same product; (2) HTA decision making is a multidisciplinary process that differs across individual agencies; (3) early engagement with regulators is encouraged by both the EMA and FDA, although early engagement did not ensure that the external control study was ultimately considered in the final decision; (4) criticism about the implementation and interpretation of time-to-event endpoints (e.g., overall survival (OS) and progression-free survival (PFS)) by the EMA and FDA was common; and (5) per our review, the presence of missingness did not in itself result in a negative review; it was important to consider the variable with missing data and its use, as well as the level and type of missingness.

Comparison of submissions using RWE between the EMA and FDA

When RWE was submitted to both regulatory agencies for the same product (Table 1), the feedback and the extent to which

the RWE was considered in the final decision varied. An example of the variation in emphasis on RWE is the chimeric antigen receptor (CAR) T-cell therapy axicabtagene ciloleucel, which was submitted to both the EMA and FDA in 2017. Although the SCHOLAR-1⁵⁵ historical control was included in the FDA submission, there was limited feedback from the agency on the design and conduct of the external control. This is in contrast to the EMA submission, where SCHOLAR-1 was emphasized and prompted more detailed feedback on the choice of external control and methods. Worst case analyses were requested by the EMA to reduce the impact of baseline differences in prognostic factors (e.g., Eastern Cooperative Oncology Group (ECOG) score) and “to avoid overly optimistic findings.”² Additionally, the EMA requested patient-level data from the applicant to allow the assessors to conduct further sensitivity analyses: overall and subgroup specific Kaplan–Meier estimates were derived from the patient-level data, and the concordance of local and central investigators was recalculated to aid in assessment of the differences between tumor assessments.⁵⁶ This difference in emphasis may depend on context, including the agency’s opinion of the strength of evidence provided in the uncontrolled trial and their experience with different therapies (e.g., CAR T-cell therapies; bispecific T-cell engager (BiTE) antibodies).

The FDA more frequently documented conducting additional analyses than the EMA, such as sensitivity analyses, and even suggested raw data be submitted so that analyses could be replicated. The FDA draft guidance states that sponsors “must ensure that they are able to submit patient-level data for any RWD ... which would allow the FDA to replicate the study analysis using the same dataset and analytic approach.”¹² Additional analyses by the agency were most often to replicate and confirm efficacy findings from the RWD analysis. For example, the FDA review of fish oil triglycerides injectable emulsion stated that the review was based primarily on the reviewer’s independent analysis of the datasets provided by the sponsor, and secondarily on the sponsor’s study report. In many cases, the FDA conducted sensitivity analyses, such as imputation of prognostic variables (e.g., worst case imputation) or variations on key assumptions. Additional subgroup analyses by the FDA were also documented.

Comparison of submissions using RWE between HTA bodies

When comparing feedback from submissions to multiple HTA bodies for the same indication, RWD sources often differed depending on the geography of the patient population. The label extension for pembrolizumab was reviewed by NICE, G-BA, and HAS, although the sponsor submitted data from different RWD sources to G-BA (German Hodgkin Study Group (GHSG) registry data) than NICE (Cheah *et al.*, 2016 (US)⁵⁷ and Eyre *et al.*, 2017 (UK)⁵⁸) and HAS (Cheah *et al.*, 2016 (US)⁵⁷). G-BA and HAS did not consider the RWE in the decision whereas NICE did; NICE and HAS gave positive recommendations whereas G-BA gave a negative recommendation, stating that an added benefit for pembrolizumab could not be demonstrated because (1) the GHSG registry data were deemed not suitable due to the treatment history of included patients and (2) submitted results from both trials were viewed as descriptive as they did not calculate any

effect measures to derive an added benefit of pembrolizumab. As such, G-BA reviewers concluded that no relevant data were submitted. In their critique of Cheah *et al.*, 2016,⁵⁷ NICE and HAS both noted an absence of data on OS, leading to uncertainty over the size of the effect and long-term outcomes. The choice by HAS not to consider the RWD in their decision was primarily driven by missing data on important prognostic factors; NICE concluded that although the study⁵⁷ may not fully represent UK clinical practice, it was the best available evidence for standard of care at the time of the sponsor’s submission.

HTA decision making is a multidisciplinary process including many factors that may differ across individual agencies.⁵⁹ This variation may be due to differences in the weighing of evidence, context, processes and procedures, and/or geographic differences in patient populations (e.g., standard of care).⁶⁰ Indeed, findings from this review indicate that HTA bodies (NICE, IQWiG, G-BA, and HAS) have different levels of receptiveness to RWE. NICE most often considered the RWE in the decision compared with the other agencies, consistent with a recent review of RWE used in assessments of cancer drugs by NICE, which found that RWE was accepted in cancer drug submissions to NICE in the majority of cases.²⁷ The NICE Strategy 2021 to 2026 states their ambition to use RWD to resolve gaps in knowledge and drive forward access to innovations for patients, as evidenced by their 2022 RWE framework.³⁰ In contrast, recommendations on use of RWE by IQWiG, G-BA, and HAS have been more limited, which potentially explains their slower acceptance.

As observed in the regulatory context, HTA decisions are influenced by factors, such as unmet need and orphan designation, making it a challenge to correlate the HTA outcome to the strength of the evidence package.⁷ Given that external control arms are generally utilized for rare diseases (which often receive orphan designation), many of the submissions included in this review were designated orphan drugs. Orphan drugs can have different threshold of evidence requirements than non-orphan drugs, an important distinction when comparing results for IQWiG and G-BA. Generally, orphan drugs undergo a simplified assessment procedure in Germany, whereby IQWiG is only contracted to appraise the epidemiology and treatment cost data (not the clinical data, as for non-orphan drugs). Thus, G-BA will conclude on a benefit rating based on a simplified version of the dossier assessment but IQWiG will not. Due to the nature of orphan diseases, submissions for orphan drugs are more likely to include uncontrolled trials, external controls, or some other form of RWD and thus forego standard IQWiG assessment. Notably, these special rules for orphan drugs are only applicable if the annual budget for the drug remains below €50 million. If a drug surpasses this threshold, it will automatically be re-assessed, undergo full IQWiG benefit assessment, and receive a new G-BA rating. Any subsequent assessments (e.g., for label expansions) will undergo standard assessment. As such, our review identified eligible submissions to G-BA for products that did not have eligible submissions to IQWiG. Although this limited the comparison between IQWiG and G-BA, the submissions that were reviewed by both IQWiG and G-BA ($n=5$) demonstrated consistently similar feedback.

Comparison of submissions using RWE between regulators and HTA bodies

HTA feedback differed from regulatory feedback in key ways. Generalizability of the RWD to the country specific population was emphasized, particularly when the RWD was from a different country. NICE was often concerned with how generalizable the results of the trial were to the National Health Service (NHS) or stated that the comparator data may not fully represent UK clinical practice. G-BA noted this in their review of entrectinib, stating that the transferability of the data from the US Flatiron Health database to the German healthcare context is questionable due to structural differences in the healthcare systems. In order to understand if the RWD is representative of clinical practice in a given country, it is important that the standard of care received and the specific therapeutic journey of patients in the external control is well-characterized and comparable across healthcare systems. This includes information on the interventions carried out (e.g., concomitant medications, local and intensive care measures, and precise combinations of chemotherapies given as standard of care). When RWD is combined from multiple sources, there is greater potential for heterogeneity in the definition of standard of care and treatments received. In NICE's review of autologous anti-CD19-transduced CD3+ cells, for example, the committee found the use of McCulloch *et al.* 2020 (UK)⁶¹ alone more appropriate than the sponsor's approach of a blended comparator as the study best represents the patient population in the NHS and avoids issues around the heterogeneity of the identified standard of care studies. This issue of heterogeneity is echoed in the regulatory feedback, as having a well-defined control population comparable to the treated trial population is critical for successful implementation of an external control. Similar to regulatory, HTA feedback emphasized the importance of large effect estimates, comparability of cohorts, index and censoring criteria, appropriate confounding adjustment, and endpoint selection. As expected, regulatory feedback emphasized safety and effectiveness whereas HTA feedback emphasized healthcare utilization and cost-effectiveness.

Regulator and HTA engagement

Regulatory interaction early in the study design process is encouraged by the EMA and FDA. The EMA recommends early discussions take place with involvement of the concerned Rapporteurs or Lead Member States (and concerned EMA Committees) as well as HTA bodies if relevant.²⁰ Sponsors are able to request Scientific Advice from the EMA at any stage of a medicine's development as well as Protocol Assistance, a special form of Scientific Advice available for developers of designated orphan medicines for rare diseases.⁶² The FDA guidance recommends that, for all studies using RWD that will be submitted to the FDA to support a regulatory decision, sponsors should submit protocols and statistical analysis plans (SAPs) before conducting the study and that sponsors seeking FDA input before conducting the study should request comments or a meeting to discuss the study with the relevant FDA review division, potentially with input from the FDA's RWE Subcommittee under the Office of Medical Policy.^{14,63} Early engagement with regulators was documented more frequently for EMA submissions than for FDA submissions, although

pre-NDA/pre-BLA meetings with the FDA were still utilized by sponsors. For both the EMA and FDA, early engagement did not ensure that the external control study was ultimately considered in the final decision. For example, naxitamab-gqqk was approved in November 2020 for the treatment of relapsed/refractory high-risk neuroblastoma. Early on, the FDA suggested the sponsor conduct a retrospective review of the naxitamab compassionate-use program in Spain for additional efficacy and safety data. At a pre-NDA meeting held between the sponsor and the FDA in 2018, the FDA stated that in the context of a single-arm, single center study, a blinded independent review of all imaging and pathologic response data by external expert pediatric radiologists and pathologists would reduce the risk for reader bias and be the most reliable process for the efficacy determination. Although the FDA suggested this retrospective review, the FDA did not consider the RWE study to provide substantive information to support safety and effectiveness because the RWD sample size was too small ($n=36$) to be meaningful and the nature of the study may have led to incomplete data collection, rendering interpretation of the results difficult.

HTA bodies provide their own mechanisms for early engagement. NICE Scientific Advice provides a fee-for-service consultation to advise sponsors on whether these will generate relevant evidence for future submissions to NICE to enable market access. Similarly, the G-BA's Early Advice⁶⁴ and HAS' Early Dialogues⁶⁵ offer pathways for sponsors to gain insight into each agency's view of the study program. Nonetheless, this study found that early engagement with NICE was recorded more frequently than for other HTA bodies. This may be driven by sponsors' reluctance to engage with G-BA due to the binding nature of their recommendations but may also be attributable to bias in how consistently early engagement was recorded for different HTA bodies. The sponsors that did seek some form of consultation most often received feedback on economic modeling, standard of care/choice of comparator, or selection of RW endpoints. In the German context, for example, endpoints that are appropriate for regulatory purposes may not be the same endpoints that are suitable to show a patient the relevant added benefit.⁶⁴

Evaluation of real-world endpoints

The choice of primary and secondary endpoints for an external control study is a critical decision for sponsors involving tradeoffs between the ideal RW endpoint and practical considerations, such as feasibility.⁶⁶ Uncontrolled trials that use surrogate endpoints (e.g., objective response rate (ORR) or duration of response (DoR)) to evaluate clinical benefit have become the basis for accelerated or breakthrough regulatory approval of precision oncology drugs when the study populations are relatively small.⁶⁷ The FDA guidance on appropriate endpoints for single-arm oncology trials states that "the FDA has sometimes accepted ORR and response duration observed in (single-arm trials [SATs]) as substantial evidence supporting accelerated approval" and that SATs "do not adequately characterize time-to-event endpoints such as OS, disease-free survival (DFS), time to progression (TTP), or progression-free survival (PFS). Because of variability in the natural history of many forms of cancer, a randomized study is necessary to evaluate time-to-event endpoints."⁶⁸

This position toward time-to-event endpoints was commonly expressed in regulatory feedback. For tafasitamab-cxix, the FDA considered time-to-event endpoints, such as OS, PFS, event-free survival (EFS), and time to next treatment (TTNT) difficult to interpret in the context of an uncontrolled trial due to inherent differences in data elicitation between clinical trials and clinical practice with respect to what are being measured, how accurately they are being measured, and when they are being measured. Regulators frequently emphasized the challenges with interpretation of time-to-event endpoints when RWD was used to contextualize findings from uncontrolled trials, including inconsistent definitions of time intervals, differences in the study population, differences in the frequency and timing of assessments, and advances in medical care over time.

HTA bodies did not share the same degree of criticism of time-to-event endpoints as regulators, likely because OS (and surrogate measures, such as PFS) are considered patient-relevant endpoints. Nonetheless, both regulators and HTA bodies provided feedback surrounding operationalization of endpoints. For avapritinib, used to treat adult patients with inoperable or metastatic gastrointestinal stromal tumors, the G-BA was critical of OS as an endpoint not because of its time-to-event nature, but because of the starting point of the observation period for the time-to-event analysis in the external control. Specifically, the follow-up of OS was determined for the control population from the start of the first therapy with a tyrosine kinase inhibitor (TKI), meaning that all patients in the control population were in the first line of therapy. The start of the observation period for OS had thus not been chosen in the external control according to the therapy lines of the trial population.

Mortality is generally not captured in administrative claims data and linkage is often required to obtain information on survival; however, certain RWD sources (e.g., patient registries) may more reliably capture mortality data and thus may be more suitable for evaluating time-to-event endpoints (e.g., OS). One example of this is fosdenopterin hydrobromide, approved in 2021 for treatment of combined molybdoflavoprotein enzyme deficiency (MoCD) type A. This comparative, noninterventional natural history study retrospectively and prospectively collected data on untreated patients with MoCD type A using RWD from academic centers in 14 countries. The FDA agreed that OS could be used as a primary endpoint for the RW study because mortality had the most reliable data capture in their selected data source whereas other clinical outcomes had differences in measurement and collection frequency. The FDA also stated that the strengths of the data (e.g., the use of a reliable and objective endpoint of mortality and a large treatment effect size) outweighed the limitations in the context of a rare, fatal disease with no other available therapies to improve survival. This notable exception speaks to the importance of selecting the appropriate data source for the outcomes of interest, as well as the usefulness of achieving a large magnitude of effect in obtaining approval.

Both regulatory and HTA feedback indicated there is a benefit in observing large effect estimates. In their review of cemiplimab, which included an unadjusted literature-based benchmark, the G-BA found that, due to different observation periods and presentation of survival rates at different points in time, OS estimates between arms could not be meaningfully compared. Furthermore,

G-BA stated that added benefit can only be concluded if the effects are sufficiently large (consistent with IQWiG guidance).²⁵ Ultimately, the naïve indirect comparison was deemed unsuitable to demonstrate added benefit and the G-BA concluded that the effect estimates for OS presented for cemiplimab were not sufficiently large to exclude that they could have been caused by systematic bias alone. Similarly, the FDA noted that historical controls can provide convincing evidence of efficacy when the outcomes with currently available treatment options are poor, and the treatment effect is too large to be easily explained by confounding factors. As such, sensitivity analyses and quantitative bias assessment can help ascertain the extent to which the observed effect of treatment may be biased.

Taken together, these findings suggest the need for further guidance and receptivity from regulators and HTA bodies regarding the selection of RW endpoints and evaluation of time-to-event endpoints estimated using RWD.

Missing data

Per our review, the presence of missingness does not in itself result in a negative review. How regulators and HTA bodies view the treatment of missing data is dependent on many factors, including which variable has a high proportion of missingness, how that variable is being used, the underlying mechanism of missingness (i.e., missing at random (MAR), missing not at random (MNAR), and missing completely at random (MCAR)), and the overall effect size. Missingness in outcome assessment is particularly problematic. As an example, the amount of missing data in the EMA submission for tisagenlecleucel was quite large, with OS data being reported in only 81 of the 136 patients (60%) who achieved response to treatment for refractory disease. OS was a secondary endpoint, and it appears the EMA's RWE assessment was based largely on the magnitude of the effect estimate for the primary endpoint (ORR as determined centrally by independent review committee (IRC) assessment based on Lugano Classification criteria), which was quite large (trial ORR: 33.9% vs. RW ORR: 26%) and confirmed by a series of predefined sensitivity analyses.

The FDA draft guidance provides general considerations for handling missing data, such as distinguishing between implicit and explicit missing data and recommends that descriptive analyses be included to characterize missingness.^{13,17,63} Nonetheless, the guidance does not specifically reference analytical tools or strategies deemed acceptable to address bias and confounding due to missingness.⁶⁹ Similarly, the 2022 NICE RWE Framework mentions, but does not provide guidance, around appropriate use of methods for handling missing data, including imputation, inverse probability weighting, and maximum likelihood estimation.³⁰ In the absence of specific guidance, sponsors used different approaches for handling missing data in RWD (**Table S2**). Imputation was not documented for the majority of submissions but, when utilized, was primarily used by sponsors or reviewers as sensitivity analyses to interrogate the robustness of results.

Only one EMA submission explicitly documented imputation to address missingness in the RWD. In the submission for ide-cel, multiple imputation procedures created 30 datasets. As a conservative measure, the analysis allowed for as much as 30% missing

data for highly prognostic covariates in the eligible relapsed/refractory multiple myeloma cohort (covariates with > 30% missing data were excluded from propensity score (PS) balancing). Covariates considered important predictors of outcome by the scientific steering committee (i.e., age, albumin, and number of prior regimens) were forced into the model irrespective of a lack of association with group membership. Despite these efforts, the EMA noted that the robustness of the adjusted indirect treatment comparison based on the RWE was difficult to verify, considering the rather selected study population and the missing data on several important prognostic factors.

Missingness may result in the inability to adjust for important prognostic factors; however, it is the magnitude and direction of the bias that may be due to residual confounding that drives regulator and HTA views on missing data. In particular, HTA bodies emphasized that the exact treatment regimen and patient journey for those in the external control arm must be characterized and described. In the submission of midostaurin (used to treat acute myeloid leukemia) to the G-BA, no adjustment of relevant factors was made between the pooled data of the trials and the historical cohort control. Thus, the G-BA assumed bias in the results in favor of midostaurin and deemed the historical comparison inadequate to evaluate efficacy. Further information on the interventions carried out (e.g., concomitant medication, and local and intensive care measures), is missing so that no information is available on how these patients were treated.

Missingness is closely related to sample size. Even a relatively small amount of missing data in a rare disease population may further reduce the sample size for efficacy analyses. Furthermore, confounding bias and selection bias may be magnified when operating with very small patient counts. As such, maximizing RWD sample size is an important consideration for sponsors. One means of maximizing statistical power is selecting an analytic method that allows leveraging of information from all patients included in the sample.⁶⁷ Using PS weighting, as opposed to matching, includes all patients rather than excluding patients that cannot be matched to a patient in the other treatment arm⁷⁰ (noting that average treatment effect (ATE) and average treatment effect in the treated (ATT) estimate the effect in different treatment populations). This is illustrated through one of the blinatumomab label expansions, approved by the EMA in 2020. The EPAR stated that matching can require large sample sizes if the degree of PS overlap is limited, and many patients need to be dropped from the analysis if a proper match is not identified within the prespecified caliper bounds. Given the limited sample size of the blinatumomab treatment group, matching was not considered viable for PS adjustments and inverse probability weighting was used for PS adjustments instead.

Closely related to missingness is the issue of data quality. Whereas the 2021 draft FDA guidance documents mention data quality in several places,^{12–15} the term itself is not well-defined. The main tenants of data quality as defined by the Duke-Margolis Center for Health Policy are accuracy, completeness, provenance, and transparency of data processing.⁷¹ Regulators did not indicate whether there are data source types (e.g., EHRs, claims, and

registries) that are more appropriate than others in certain circumstances. Feedback centered around cohort comparability and the components of data quality rather than the type of RWD. We also observed that data quality specific terminology appears more commonly in the HTA feedback. Agencies called out when information on data quality was not included and expressed concern regarding the quality of the external control data, particularly as it pertains to accuracy and completeness (i.e., the percentage of data missing and/or how this type of data is managed or replaced). It is these specific tenants of data quality (accuracy and completeness) that are outlined in the NICE RWE framework.³⁰ This review suggests that HTA bodies are more readily integrating a data quality lens and terminology in their review compared with regulators.

Future directions

By cross-referencing the EMA, FDA, and HTA assessments of RWD-derived external controls, we found that the feedback and extent to which RWE was considered in final decisions by the FDA and EMA was highly variable and HTA decision making differed across individual agencies, even when the agencies considered the same RWE for the same product. As such, we recommend increased harmonization between regulatory bodies and between HTA bodies to improve efficiency and lower evidence generation costs.

This review identified methodological and operational aspects of external control design and implementation, including early engagement, selection, and operationalization of RW endpoints, and the treatment of missing data. Importantly, there continues to be a shift away from benchmarking toward comparative external control studies. We encourage sponsors to continue sharing lessons learned from their regulatory and HTA interactions in the public domain to better understand how these and other considerations affect regulatory approval and reimbursement. As a further extension to our study, future systematic literature reviews should focus on summarizing the characteristics of regulatory and HTA reimbursement approvals of drugs evaluated using single-arm trials without accompanying external control studies.

The EMA “Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources” aims to provide recommendations for the use of the European Union metadata catalogue, targeted for release in late 2023, to identify RWD sources suitable for specific research questions and to assess the suitability of data sources proposed to be used in a study protocol or referred to in a study report.²¹ Regulators do not endorse specific data source types; however, this initiative from the EMA will facilitate the discoverability of data sources to generate adequate evidence for regulatory purpose. Ongoing efforts by sponsors, data suppliers, and other stakeholders to inform and improve the acceptability of data for regulatory decision making will be critical. Examples include the draft RWD Audit Readiness Considerations tool from TransCelerate BioPharma Inc.⁷² and recommendations for identifying implications of data source and patient characteristics on RW endpoints published by Friends of Cancer Research and collaborators.⁷³

Strengths and limitations

This study has a number of strengths. First, we systematically reviewed submissions to the EMA, FDA, and select HTA bodies rather than limiting to one agency. This allowed us to compare regulatory and HTA feedback both generally and for specific products, providing a more comprehensive assessment of external control implementation. Furthermore, because external controls are applicable to a wide range of indications, we did not limit to a specific therapeutic area.

To ensure all eligible submissions were appropriately included, the final list of products for data extraction was compared against products included in other reviews (noting that eligibility criteria differed between studies).^{41,42,44,74} Whereas systematic in nature, we limited our scope to products with regulatory pathways (e.g., accelerated approval and breakthrough designation). This may have resulted in the exclusion of some submissions utilizing an external control. Although information on refused and withdrawn EMA submissions was captured, this study was limited by the lack of publicly available information on refused and withdrawn FDA submissions. Similarly, if it could not be verified using publicly available information whether a literature-based estimate was derived from RWD or historical trial data, the submission was included. Due to the nature of this review, we were only able to draw conclusions based on information in the publicly available documents which may have resulted in reporting bias, particularly for our findings on early engagement and agency-conducted analyses. As many of these conclusions are based on qualitative analyses of the regulatory and HTA recommendations, there may be variability in their strength. Based on publicly available data, we were also not able to confirm whether the data was curated (e.g., using chart abstraction) prior to submission.

Additionally, the published appraisals available in HTAA, whereas regularly updated and subject to quality review, may have been variable with regard to the level of detail across the different HTA bodies. The submission documents for IQWiG, G-BA, and particularly HAS may have been subject to variable quality of translation.

Last, extracting multiple RWD analyses for each submission (where applicable) may have resulted in capture of additional literature-based analyses; however, the inclusion of these analyses provides additional context and information on how external controls are included as part of an evidence package.

CONCLUSIONS

This review found that external controls are most frequently utilized in the therapeutic areas of hematology and oncology, as well as neurology, for regulatory submissions. Regulators and HTA bodies have become more familiar with external controls over time, allowing them to offer sponsors more detailed and specific feedback. As such, this review identified several key operational and methodological aspects for which more detailed guidance and alignment within and between regulatory agencies and HTA bodies is necessary, including early, thoughtful engagement of regulators and HTA bodies, methods for addressing missing data, and selection of RW endpoints. We found that when RWE from the same data source was submitted to multiple agencies, the degree of similarities in feedback and the extent to which the RWE

was considered in the final decision varied from submission to submission. Submissions to HTA bodies in different countries for the same indication often utilized different RWD sources. Early engagement with regulators was documented more frequently for EMA submissions than for FDA submissions, and there were differences between the EMA and FDA in regard to how often additional analyses conducted by the agency were documented. Early engagement and agency-conducted analyses were most commonly documented for NICE. When addressing missing data, we found that how regulators and HTA bodies view the treatment of missing data depends on many factors. We also found that the EMA and FDA criticized the implementation and interpretation of time-to-event endpoints in the context of an external control study. Continued collaboration and guidance to address these and other aspects will inform and assist stakeholders attempting to generate adequate scientific evidence using RWD-derived external controls.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

This analysis was funded by Regeneron Pharmaceuticals, Inc. Medical writing and editorial support under the direction of the authors was provided by Julien Heidt of IQVIA and funded by Regeneron Pharmaceuticals, Inc. according to Good Publication Practice guidelines ([Link](#)). Responsibility for all opinions, conclusions, and data interpretation lies with the authors.

FUNDING

This work was funded by Regeneron through a contract with IQVIA.

CONFLICT OF INTEREST

V.M., J.H., and R.G.W.Q. are employees and report ownership interests of Regeneron Pharmaceuticals Inc. JH owns stock in Pfizer Inc. P.S.H. is a consultant who was contracted to support this work. L.H.C. and O.S.M. serve as consultants to Regeneron Pharmaceuticals Inc. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

All authors wrote the manuscript. O.S.M., L.C., J.He., L.W., S.O., D.C., P.S.H., J.Ha., and R.G.W.Q. designed the research. J.He., L.W., and P.S.H. performed the research. J.He., L.W., P.S.H., Y.S., and T.M. analyzed the data.

© 2023 Regeneron Pharmaceuticals Inc and The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. Beaulieu-Jones, B.K. et al. Examining the use of real-world evidence in the regulatory process. *Clin. Pharmacol. Ther.* **107**, 843–852 (2020).
2. European Federation of Pharmaceutical Industries and Associations. Evidence mix (measures, insights, and examples): Evaluating the EU regulatory system <https://www.efpia.eu/media/636564/evidence-mix_final-9-dec-2021.pdf> (2021). Accessed March 31, 2023.
3. Mack, C., Christian, J., Brinkley, E., Warren, E.J., Hall, M. & Dreyer, N. When context is hard to come by: external

- comparators and how to use them. *Ther Innov Regul Sci* (2019). <https://doi.org/10.1177/2168479019878672>.
4. Thorlund, K., Dron, L., Park, J.J.H. & Mills, E.J. Synthetic and external controls in clinical trials – a primer for researchers. *Clin. Epidemiol.* **12**, 457–467 (2020).
 5. Burcu, M. et al. Real-world evidence to support regulatory decision-making for medicines: considerations for external control arms. *Pharmacoepidemiol. Drug Saf.* **29**, 1228–1235 (2020).
 6. Gray, C.M., Grimson, F., Layton, D., Pocock, S. & Kim, J. A framework for methodological choice and evidence assessment for studies using external comparators from real-world data. *Drug Saf.* **43**, 623–633 (2020).
 7. Patel, D. et al. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health* **24**, 1118–1125 (2021).
 8. Cave, A., Kurz, X. & Arlett, P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* **106**, 36–39 (2019).
 9. Goring, S. 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* **9**, e024895 (2019).
 10. US FDA. PDUFA reauthorization performance goals and procedures fiscal years 2018 through 2022 <<https://www.fda.gov/media/99140/download>> (2017). Accessed April 5, 2023.
 11. US FDA. Framework for FDA's real-world evidence program <<https://www.fda.gov/media/120060/download>> (2018). Accessed March 31, 2023.
 12. US FDA. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products <<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug>> (2021). Accessed March 31, 2023.
 13. US FDA. Real-world data: Assessing registries to support regulatory decision-making for drug and biological products guidance for industry <<https://www.fda.gov/media/154449/download>> (2021). Accessed March 31, 2023.
 14. US FDA. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products <<https://www.fda.gov/media/152503/download>> (2021). Accessed March 31, 2023.
 15. US FDA. Data standards for drugs and biological product submissions containing real-world data guidance for industry <<https://www.fda.gov/media/153341/download>> (2021). Accessed April 1, 2023.
 16. US FDA. Submitting documents using real-world data and real-world evidence to FDA for drug and biological products guidance for industry <<https://www.fda.gov/media/124795/download>> (2022). Accessed April 5, 2023.
 17. US FDA. Considerations for the design and conduct of externally controlled trials for drug and biological products guidance for industry <<https://www.fda.gov/media/164960/download>> (2023). Accessed April 1, 2023.
 18. Arlett, P., Kjaer, J., Broich, K. & Cooke, E. Real-world evidence in EU medicines regulation: enabling use and establishing value. *Clin. Pharmacol. Ther.* **111**, 21–23 (2022).
 19. EMA. EMA regulatory science to 2025 strategic reflection <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf> (2020). Accessed April 5, 2023.
 20. EMA. Guideline on registry-based studies <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en-0.pdf> (2021). Accessed March 31, 2023.
 21. EMA. Good practice guide for the use of the metadata catalogue of real-world data sources <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf> (2022). Accessed April 6, 2023.
 22. EMA. ICH guideline E8 (R1) on general considerations for clinical studies <https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-8-general-considerations-clinical-trials-step-5_en.pdf> (2022). Accessed April 6, 2023.
 23. EMA. Health technology assessment bodies <<https://www.ema.europa.eu/en/partners-networks/health-technology-assessment-bodies>> (2023). Accessed April 5, 2023.
 24. HAS. Choices in methods for economic evaluation – HAS <https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_choices_in_methods_for_economic_evaluation.pdf> (2020). Accessed April 5, 2023.
 25. IQWiG. Concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a Social Code Book V <https://www.iqwig.de/download/a19-43_routine-practice-data-for-the-benefit-assessment-of-drugs_extract-of-rapid-report_v1-0.pdf> (2020). Accessed April 5, 2023.
 26. NICE. NICE DSU technical support document 17: The use of observational data to inform estimates of treatment effectiveness in technology appraisal: Methods for comparative individual patient data <<https://www.sheffield.ac.uk/sites/default/files/2022-02/TSD17-DSU-Observational-data-FINAL.pdf>> (2015). Accessed April 5, 2023.
 27. Bullement, A. et al. Real-world evidence use in assessments of cancer drugs by NICE. *Int J Technol Assess Health Care* **36**, 388–394 (2020).
 28. Facey, K., Rannanheimo, P., Batchelor, L., Borchardt, M. & De Cock, J. Real-world evidence to support Payer/HTA decisions about highly innovative technologies in the EU—actions for stakeholders. *Int J Technol Assess Health Care* **36**, 459–468 (2020).
 29. Oortwijn, W., Sampietro-Colom, L. & Trowman, R. How to deal with the inevitable: generating real-world data and using real-world evidence for hta purposes – from theory to action. *Int. J. Technol. Assess. Health Care* **35**, 346–350 (2019).
 30. NICE. NICE real-world evidence framework <<https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837>> (2022). Accessed March 31, 2023.
 31. Berger, M.L. et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. *Pharmacoepidemiol. Drug Saf.* **26**, 1033–1039 (2017).
 32. Burger, H.U. et al. The use of external controls: to what extent can it currently be recommended? *Pharm. Stat.* **20**, 1002–1016 (2021).
 33. Franklin, J.M., Glynn, R.J., Martin, D. & Schneeweiss, S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin. Pharmacol. Ther.* **105**, 867–877 (2019).
 34. Gatto, N.M., Reynolds, R.F. & Campbell, U.B. A structured preapproval and postapproval comparative study design framework to generate valid and transparent real-world evidence for regulatory decisions. *Clin. Pharmacol. Ther.* **106**, 103–115 (2019).
 35. Hampson, G., Towse, A., Dreitlein, W.B., Henshall, C. & Pearson, S.D. Real-world evidence for coverage decisions: opportunities and challenges. *J Comp Eff Res* **7**, 1133–1143 (2018).
 36. Jaksa, A., Wu, J., Jonsson, P., Eichler, H.G., Vititoe, S. & Gatto, N.M. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. *J Comp Eff Res* **10**, 711–731 (2021).
 37. Mahendaratnam, N., Mercon, K., Gill, M., Benzing, L. & McClellan, M.B. Understanding use of real-world data and real-world evidence to support regulatory decisions on medical product effectiveness. *Clin. Pharmacol. Ther.* **111**, 150–154 (2022).
 38. Mishra-Kalyani, P.S. et al. External control arms in oncology: current use and future directions. *Ann. Oncol.* **33**, 376–383 (2022).
 39. Orsini, L.S. et al. Improving transparency to build trust in real-world secondary data studies for hypothesis testing—why, what, and how: recommendations and a road map from the real-world evidence transparency initiative. *Value Health* **23**, 1128–1136 (2020).

40. Wang, S.V. et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies v1.0. *Pharmacoepidemiol. Drug Saf.* **26**, 1018–1032 (2017).
41. Aetion. Ebook: The role of real world evidence in FDA approvals <<https://aetion.com/evidence-hub/ebook-the-role-of-real-world-evidence-in-fda-approvals/>> (2020). Accessed April 5, 2023.
42. Arondekar, B. et al. Real-world evidence in support of oncology product registration: a systematic review of new drug application and biologics license application approvals from 2015 to 2020. *Clin. Cancer Res.* **28**, 27–35 (2022).
43. Bakker, E., Plueschke, K., Jonker, C.J., Kurz, X., Starokozhko, V. & Mol, P.G.M. Contribution of real-world evidence in European Medicines Agency's regulatory decision making. *Clin. Pharmacol. Ther.* **113**, 135–151 (2023).
44. Bolislis, W.R., Fay, M. & Kuhler, T.C. Use of real-world data for new drug applications and line extensions. *Clin. Ther.* **42**, 926–938 (2020).
45. Flynn, R. et al. Marketing authorization applications made to the European medicines agency in 2018–2019: what was the contribution of real-world evidence? *Clin. Pharmacol. Ther.* **111**, 90–97 (2022).
46. Hatswell, A.J., Baio, G., Berlin, J.A., Irs, A. & Freemantle, N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open* **6**, e011666 (2016).
47. Jahanshahi, M. et al. The use of external controls in FDA regulatory decision making. *Ther. Innov. Regul. Sci.* **55**, 1019–1035 (2021).
48. Makady, A. et al. Policies for use of real-world data in health technology assessment (HTA): a comparative study of six HTA agencies. *Value Health* **20**, 520–532 (2017).
49. Purpura, C.A., Garry, E.M., Honig, N., Case, A. & Rassen, J.A. The role of real-world evidence in FDA-approved new drug and biologics license applications. *Clin. Pharmacol. Ther.* **111**, 135–144 (2022).
50. Kang, J. & Cairns, J. Protocol for data extraction: how real-world data have been used in the National Institute for Health and Care Excellence appraisals of cancer therapy. *BMJ Open* **12**, e055985 (2022).
51. EMA. Medicines <<https://www.ema.europa.eu/en/medicines>> (2023). Accessed April 5, 2023.
52. US FDA. Drugs@FDA: FDA-approved drugs <<https://www.accessdata.fda.gov/scripts/cder/daf/>> (2023). Accessed April 5, 2023.
53. US FDA. Biological approvals by year <<https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/biological-approvals-year>> (2023). Accessed April 5, 2023.
54. IQVIA. Leveraging HTA accelerator: A flexible approach to finding the answers you need <<https://www.iqvia.com/landing/hta-accelerator>> (2023). Accessed April 5, 2023.
55. Crump, M. et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood* **130**, 1800–1808 (2017).
56. Papadoulis, I. et al. EMA review of axicabtagène ciloleucel (Yescarta) for the treatment of diffuse large B-cell lymphoma. *Oncologist* **25**, 894–902 (2020).
57. Cheah, C.Y. et al. Patients with classical Hodgkin lymphoma experiencing disease progression after treatment with brentuximab vedotin have poor outcomes. *Ann. Oncol.* **27**, 1317–1323 (2016).
58. Eyre, T.A. et al. Results of a multicentre UK-wide retrospective study evaluating the efficacy of brentuximab vedotin in relapsed, refractory classical Hodgkin lymphoma in the transplant naïve setting. *Br. J. Haematol.* **179**, 471–479 (2017).
59. Pinto, A., Naci, H., Neez, E. & Mossialos, E. Association between the use of surrogate measures in pivotal trials and health technology assessment decisions: a retrospective analysis of NICE and CADTH reviews of cancer drugs. *Value Health* **23**, 319–327 (2020).
60. Vreman, R.A., Mantel-Teeuwisse, A.K., Hovels, A.M., Leufkens, H.G.M. & Goettsch, W.G. Differences in health technology assessment recommendations among European jurisdictions: the role of practice variations. *Value Health* **23**, 10–16 (2020).
61. McCulloch, R. et al. Efficacy of R-BAC in relapsed, refractory mantle cell lymphoma post BTK inhibitor therapy. *Br. J. Haematol.* **189**, 684–688 (2020).
62. EMA. Scientific advice and protocol assistance <<https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance#:~:text=EMA%20gives%20scientific%20advice%20by,development%20of%20a%20particular%20medicine.&text=Scientific%20advice%20from%20EMA%20is,applications%20for%20the%20medicine%20concerned>> (2023). Accessed April 1, 2023.
63. Aetion. Comment on "Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products; Draft guidance for industry" <<https://www.regulations.gov/comment/FDA-2020-D-2307-0062>> (2022). Accessed April 5, 2023.
64. Wiebe, K. G-BA early advice: How do we prepare for the requirements that the early benefit assessment in Germany poses for clinical trials? <<https://www.clinicaltrialsarena.com/news/g-ba-early-advice-how-do-we-prepare-for-the-requirements-that-the-early-benefit-assessment-in-germany-poses-for-clinical-trials-5761431-2/>> (2017). Accessed April 6, 2023.
65. HAS. Guidance for national early dialogues on medicinal products <https://www.has-sante.fr/jcms/c_2623726/en/guidance-for-national-early-dialogues-on-medicinal-products> (2016). Accessed April 6, 2023.
66. Duke-Margolis Center for Health Policy. A roadmap for developing study endpoints in real-world settings <<https://healthpolicy.duke.edu/sites/default/files/2020-08/Real-World%20End%20points.pdf>> (2020). Accessed April 6, 2023.
67. Yap, T.A., Jacobs, I., Baumfeld Andre, E., Lee, L.J., Beaupre, D. & Azoulay, L. Application of real-world data to external control groups in oncology clinical trial drug development. *Front. Oncol.* **11**, 695936 (2021).
68. US FDA. Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry <<https://www.fda.gov/media/71195/download>> (2018). Accessed April 6, 2023.
69. Duke-Margolis Center for Health Policy. Comment letter in response to FDA's "Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products; draft guidance for industry." <<https://www.regulations.gov/comment/FDA-2020-D-2307-0061>> (2022). Accessed April 1, 2023.
70. Aetion. Understanding propensity score weighting methods <<https://aetion.com/evidence-hub/understanding-propensity-score-weighting-methods-rwe/>> (2022). Accessed 2023.
71. Duke-Margolis Center for Health Policy. Characterizing RWD quality and relevancy for regulatory purposes <https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf> (2018). Accessed April 1, 2023.
72. TransCelerate. Real-world data audit readiness considerations <<https://www.transceleratebiopharmainc.com/wp-content/uploads/2022/12/RWD-Audit-Readiness-Considerations-Document.pdf>> (2022). Accessed April 6, 2023.
73. Friends of Cancer Research. Considerations for the use of real-world evidence in oncology: Lessons learned from Friends of Cancer Research collaborations <https://www.focr.org/sites/default/files/Use_of_Real-World_Evidence_in_Oncology-2020.pdf> (2020). Accessed April 6, 2023.
74. Jaksa, A. et al. A comparison of seven oncology external control arm case studies: critiques from regulatory and health technology assessment agencies. *Value Health* **25**, 1967–1976 (2022).
75. EMA. Blincyto (blinatumomab) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/blincyto-epar-public-assessment-report_en.pdf> (2015). Accessed April 6, 2023.
76. US FDA. Blincyto (blinatumomab) statistical review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000StatR.pdf> (2014). Accessed April 6, 2023.
77. Gemeinsamer Bundesausschuss. Asfotase alfa (reassessment of an orphan drug after the € 50 million turnover limit was

- exceeded: Hypophosphatasia). Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XXI – benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-6470/2020-04-02_AM-RL-XII_Asfotase-alfa_D-494_TrG_EN.pdf> (2020). Accessed April 6, 2023.
78. EMA. Strensiq (asfotase alfa) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/strensiq-epar-public-assessment-report_en.pdf> (2015). Accessed April 6, 2023.
 79. IQWiG. Asfotase alfa (hypophosphatasia) benefit assessment according to §35a social code book V <https://www.iqwig.de/download/a19-89_asfotase-alfa_extract-of-dossier-assessment_v1-0.pdf> (2020). Accessed April 6, 2023.
 80. US FDA. Strensiq™ (asfotase alfa) statistical review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/1255130rig1s000StatR.pdf> (2015). Accessed April 6, 2023.
 81. US FDA. Strensiq (asfotase alfa) summary review for regulatory action <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/1255130rig1s000SumR.pdf> (2015). Accessed April 6, 2023.
 82. EMA. Defitelio (defibrotide) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/defitelio-epar-public-assessment-report_en.pdf> (2012). Accessed April 6, 2023.
 83. US FDA. Defitelio (defibrotide sodium) summary review for regulatory action <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2016/2081140rig1s000SumR.pdf> (2016). Accessed April 6, 2023.
 84. US FDA. Defitelio (defibrotide) addendum to statistical review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2016/2081140rig1s000StatR.pdf> (2016). Accessed April 6, 2023.
 85. Gemeinsamer Bendesausschuss. Axicabtagene ciloleucel. Justification to the resolution of the federal joint committee (G-BA) on an amendment to the pharmaceuticals directive (am-rl): Annex XII – resolutions on the benefit assessment of medical products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-5742/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG_EN.pdf> (2019). Accessed April 6, 2023.
 86. EMA. Yescarta (axicabtagene ciloleucel) assessment report <https://www.ema.europa.eu/en/documents/variation-report/yescarta-h-c-004480-ii-0042-epar-assessment-report-variation_en.pdf> (2018). Accessed April 1, 2023.
 87. HAS. Yescarta (axicabtagene ciloleucel) $0.4\text{--}2 \times 10^8$ cellules, dispersion pour perfusion. Direction de l'évaluation médicale, économique et de santé publique <https://www.has-sante.fr/upload/docs/application/pdf/2018-12/yescarta_pic_ins_avis3_ct17214.pdf> (2018). Accessed April 6, 2023.
 88. NICE. Axicabtagène ciloleucel for treating diffuse large B-cell lymphoma and primary mediastinal large B-cell lymphoma after 2 or more systemic therapies <<https://www.nice.org.uk/guidance/ta559/resources/axicabtagene-ciloleucel-for-treating-diffuse-large-bcell-lymphoma-and-primary-mediastinal-large-bcell-lymphoma-after-2-or-more-systemic-therapies-pdf-82607030270917>> (2019). Accessed April 6, 2023.
 89. US FDA. Yescarta™ (axicabtagene ciloleucel) summary basis for regulatory action <<https://www.fda.gov/media/108788/download>> (2017). Accessed April 6, 2023.
 90. EMA. Brineura (cerliponase alfa) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/brineura-epar-public-assessment-report_en.pdf> (2017). Accessed April 6, 2023.
 91. Gemeinsamer Bendesausschuss. Cerliponase alfa. Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – resolutions on the benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-4721/2017-12-21_AM-RL-XII_Cerliponase-alfa_D-298_TrG_EN.pdf> (2017). Accessed April 6, 2023.
 92. NICE. Cerliponase alfa for treating neuronal ceroid lipofuscinosis type 2 <<https://www.nice.org.uk/guidance/hst12/resources/cerliponase-alfa-for-treating-neuronal-ceroid-lipofuscinosis-type-2-pdf-50216255489221>> (2019). Accessed April 6, 2023.
 93. US FDA. Brineura (cerliponase alfa) summary review for regulatory action <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/7610520rig1s000SumR.pdf> (2016). Accessed April 6, 2023.
 94. EMA. Blincyto (blinatumomab) assessment report <https://www.ema.europa.eu/en/documents/variation-report/blincyto-h-c-3731-ii-0011-epar-assessment-report-variation_en.pdf> (2019). Accessed April 6, 2023.
 95. Gemeinsamer Bendesausschuss. Blinatumomab (new therapeutic indication: Acute lymphatic leukaemia, mrd-positive patients). Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XXI – with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-5933/2019-08-15_AM-RL-XII_Blinatumomab_D-429_TrG_EN.pdf> (2019). Accessed April 6, 2023.
 96. US FDA. Blincyto (blinatumomab) accelerated approval letter <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/1255570rig1s013.pdf> (2018). Accessed April 6, 2023.
 97. EMA. Gamifant (emapalumab) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/gamifant-epar-refusal-public-assessment-report_en.pdf> (2020). Accessed April 6, 2023.
 98. US FDA. Gamifant (emapalumab) BLA multi-disciplinary review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/7611070rig1s000MultidisciplineR.pdf> (2018). Accessed April 6, 2023.
 99. EMA. Zolgensma (onasemnogene abeparvovec) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/zolgensma-epar-public-assessment-report_en.pdf> (2020). Accessed April 6, 2023.
 100. HAS. Zolgensma (onasemnogene abeparvovec) 2×10^{13} génomes du vecteur/ml, solution pour perfusion. Première évaluation <https://www.has-sante.fr/upload/docs/application/pdf/2020-12/zolgensma_16122020_synthese_ct18743.pdf> (2020). Accessed April 6, 2023.
 101. US FDA. Zolgensma (onasemnogene abeparvovec-xioi) summary basis for regulatory action <<https://www.fda.gov/media/127961/download>> (2019). Accessed April 6, 2023.
 102. EMA. Blenrep (belantamab mafodotin) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/blenrep-epar-public-assessment-report_en-0.pdf> (2020). Accessed April 6, 2023.
 103. Gemeinsamer Bendesausschuss. Belantamab mafodotin (multiple myeloma, at least 4 prior therapies, monotherapy). Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – resolutions on the benefit assessment of medical products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-7356/2021-03-04_AM-RL-XII_Belantramab-Mafodotin_D-582_TrG_EN.pdf> (2021). Accessed April 6, 2023.
 104. US FDA. Blenrep (belantamab mafodotin-blmf) BLA multi-disciplinary review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/7611580rig1s000MultidisciplineR.pdf> (2019). Accessed April 6, 2023.
 105. EMA. Enhertu (trastuzumab deruxtecan) CHMP assessment report <https://www.ema.europa.eu/en/documents/assessment-report/enhertu-epar-public-assessment-report_en.pdf> (2020). Accessed April 6, 2023.
 106. HAS. Enhertu (trastuzumab deruxtecan) 100 mg, poudre pour solution à diluer pour perfusion. Première évaluation <https://www.has-sante.fr/upload/docs/application/pdf/2021-06/enhertu_16062021_synthese_ct19098.pdf> (2021). Accessed April 6, 2023.

107. US FDA. Enhertu (fam-trastuzumab deruxtecan-nxki) NDA/BLA multi-disciplinary review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/761139Orig1s000MultidisciplineR.pdf> (2019). Accessed April 6, 2023.
108. EMA. Koselugo (selumetinib) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/koselugo-epar-public-assessment-report_en.pdf> (2021). Accessed April 6, 2023.
109. US FDA. Koselugo (selumetinib) NDA multi-disciplinary review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/213756Orig1s000MultidisciplineR.pdf> (2019). Accessed April 6, 2023.
110. EMA. Abecma (idecabtagene vicleucel) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/abecma-epar-public-assessment-report_en.pdf> (2021). Accessed April 6, 2023.
111. US FDA. Abecma (idecabtagene vicleucel) <<https://www.fda.gov/vaccines-blood-biologics/abecma-idecabtagene-vicleucel>> (2021). Accessed April 6, 2023.
112. EMA. Ayvakyt (avapritinib) assessment report <https://www.ema.europa.eu/en/documents/assessment-report/ayvakyt-epar-public-assessment-report_en.pdf> (2020). Accessed April 6, 2023.
113. Gemeinsamer Bendesausschuss. Avapritinib (gastrointestinal stromal tumors). Justification of the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-7446/2021-04-15_AM-RL-XII_Avapritinib_D-583_TrG_EN.pdf> (2021). Accessed April 6, 2023.
114. US FDA. Ayvakit (avapritinib) NDA/BLA multi-disciplinary review and evaluation <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/212608Orig1s000MultidisciplineR.pdf> (2019). Accessed April 6, 2023.
115. EMA. Zalmoxis assessment report <https://www.ema.europa.eu/en/documents/assessment-report/zalmoxis-epar-public-assessment-report_en.pdf> (2016). Accessed April 6, 2023.
116. EMA. Tibsovo (ivosidenib) assessment report <https://www.ema.europa.eu/en/documents/withdrawal-report/withdrawal-assessment-report-tibsovo_en.pdf> (2020). Accessed April 6, 2023.
117. Gemeinsamer Bendesausschuss. Mekinist (trametinib) supporting grounds. Available through IQVIA HTAA <<https://www.iqvia.com/landing/hta-accelerator>> (2017).
118. Gemeinsamer Bendesausschuss. Keytruda (pembrolizumab) supporting grounds. Available through IQVIA HTAA <<https://www.iqvia.com/landing/hta-accelerator>> (2017).
119. HAS. Zalmoxis final report. Available through IQVIA HTAA <<https://www.iqvia.com/landing/hta-accelerator>> (2019).
120. NICE. Idelalisib for treating refractory follicular lymphoma <<https://www.nice.org.uk/guidance/ta604/resources/idelalisib-for-treating-refractory-follicular-lymphoma-pdf-82608896324293>> (2019). Accessed April 6, 2023.
121. Gemeinsamer Bendesausschuss. Andexanet alfa. Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-6367/2020-02-20_AM-RL-XII_Andexanet-alfa_D-487_TrG_EN.pdf> (2020). Accessed April 6, 2023.
122. Gemeinsamer Bendesausschuss. Cemiplimab. Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-6323/2020-02-06_AM-RL-XII_Cemiplimab_D-466_TrG_EN.pdf> (2020). Accessed April 6, 2023.
123. Gemeinsamer Bendesausschuss. Entrectinib (ros1-positive, advanced non-small cell lung cancer). Justification to the resolution of the federal joint committee (G-BA) on an amendment of the pharmaceuticals directive (am-rl): Annex XII – benefit assessment of medicinal products with new active ingredients according to section 35a SGB V <https://www.g-ba.de/downloads/40-1465-7323/2021-02-18_AM-RL-XII_Entrectinib_D-558_TrG_EN.pdf> (2021). Accessed April 6, 2023.