

Regulatory and HTA Considerations for Development of Real-World Data Derived External Controls

Lesley H. Curtis¹, Oriol Sola-Morales² , Julien Heidt^{3,*} , Patrick Saunders-Hastings⁴, Laura Walsh⁵, Deborah Casso^{6,†} , Susan Oliveria^{7,†}, Tiffany Mercado³, Robbert Zusterzeel^{3,†}, Rachel E. Sobel⁸ , Jessica J. Jalbert⁹, Vera Mastey⁹, James Harnett⁹ and Ruben G. W. Quek⁹

Regulators and Health Technology Assessment (HTA) bodies are increasingly familiar with, and publishing guidance on, external controls derived from real-world data (RWD) to generate real-world evidence (RWE). We recently conducted a systematic literature review (SLR) evaluating publicly available information on the use of RWD-derived external controls to contextualize outcomes from uncontrolled trials submitted to the European Medicines Agency (EMA), the US Food and Drug Administration (FDA), and/or select HTA bodies. The 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. This paper builds on the SLR findings by delineating a set of key takeaways for the responsible generation of fit-for-purpose RWE. Practical methodological and operational guidelines for designing, conducting, and reporting RWD-derived external control studies are explored and discussed. These considerations include: (i) early engagement with regulators and HTA bodies during the study planning phase; (ii) consideration of the appropriateness and comparability of external controls across multiple dimensions, including eligibility criteria, temporality, population representation, and clinical evaluation; (iii) ensuring adequate sample sizes, including hypothesis testing considerations; (iv) implementation of a clear and transparent strategy for assessing and addressing data quality, including data missingness across trials and RWD; (v) selection of comparable and meaningful endpoints that are operationalized and analyzed using appropriate analytic methods; and (vi) conduct of sensitivity analyses to assess the robustness of findings in the context of uncertainty and sources of potential bias.

Regulators and health technology assessment (HTA) bodies are increasingly receiving submissions that use external controls derived from real-world data (RWD) to generate real-world evidence (RWE).^{1–3} The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) define RWD as routinely collected data relating to patient health status and/or the delivery of health care from a variety of sources other than traditional clinical trials.^{4,5} RWE, generated from analysis of RWD, is increasingly leveraged to support evaluation of product effectiveness. This adoption has accelerated since the emergence of the coronavirus disease 2019 (COVID-19) pandemic, which both made traditional trials more difficult and motivated new efforts and collaborations to use RWE to inform decision making.⁶ Despite recent guidance on the appropriate use of RWE to support regulatory and HTA approvals,^{7–17} best operational and methodological practices for executing an RWD-derived external control study are still being defined.^{18–20}

The aim of this review is to delineate a set of key takeaways to support the responsible generation of RWD-derived external controls for stakeholders submitting evidence to regulatory and HTA decision-making bodies to support evidence of product effectiveness and safety. This paper puts forth practical methodological considerations for designing, conducting, and reporting RWD-derived external control studies for regulatory and HTA submissions. These key takeaways are based on (i) guidance and frameworks put forward in recent thought leadership papers,^{2,21–24} (ii) guidance on the use of RWE published by regulatory and HTA bodies,^{7–17} and (iii) a recent systematic literature review (SLR) we conducted and reported separately that examined regulatory and HTA submissions using RWD-derived external controls to identify feedback related to operational and methodological choices.²⁵ In the next section, we will explore each of these in sequence before moving to a discussion of key takeaways and case studies.

¹Duke Department of Population Health Sciences and Duke Clinical Research Institute, Durham, North Carolina, USA; ²Fundació HiTT and Universitat Internacional de Catalunya (UIC), Barcelona, Spain; ³IQVIA, Regulatory Science and Strategy, Falls Church, Virginia, USA; ⁴Epi Excellence LLC, Garnet Valley, Pennsylvania, USA; ⁵IQVIA, Epidemiology and Drug Safety Practice, Boston, Massachusetts, USA; ⁶IQVIA, Epidemiology and Drug Safety Practice, Seattle, Washington, USA; ⁷IQVIA, Epidemiology and Drug Safety Practice, New York, New York, USA; ⁸Regeneron Pharmaceuticals Inc., Pharmacoepidemiology, Tarrytown, New York, USA; 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; authors are no longer employed by IQVIA.

Received February 13, 2023; accepted April 12, 2023. doi:10.1002/cpt.2913

EVOLVING GUIDANCE ON EXTERNAL CONTROLS

With the increasing use of RWE to support healthcare decision making, recent publications have reviewed and summarized examples of external controls and relevant guidance documents. Researchers have specifically reviewed RWD-derived external controls submitted in support of the FDA^{23,26} and EMA²⁷ decision making, noting the need for clearer guidance on best practices. Burns and colleagues²² recently published a review of RWE in the context of regulatory decision making, assessing adoption across the dimensions of a regulatory framework, data quality and standards guidance, and study methods guidance.²² In this review, the authors identified a need for additional collaboration across countries and groups, as well as continued guidance on the execution of innovative trials combining traditional methods with RWE designs. Other recent reviews have highlighted cohort comparability, data completeness, and data accuracy as essential elements to support the usability of RWE^{28,29} and noted selection bias and confounding as key limitations.³⁰ Although these reviews provide useful frameworks for conceptualizing design and implementation of external controls, there is still a need for more practical, readily operationalized guidance for stakeholders seeking to conduct an RWD-derived external control study of their own.

Globally, regulatory agencies have released guidance on best practices for incorporating RWE in regulatory submissions.²² Whereas specific recommendations vary across regulators, the guidances are designed to help recognize, address, and minimize the limitations inherent to RWE. For example, the FDA 2021 draft guidance documents^{13–16} outlined a set of considerations for using RWE to inform regulatory decision making, including the need for early engagement with regulators, transparency in the data collection and analysis, access to patient-level source data, and study monitoring for data integrity; nonetheless, further elucidation of the best practices on concepts such as validation, data quality, and data standards is needed. In 2023, the FDA released a draft guidance on considerations for the design and conduct of externally controlled trials using patient-level data (external controls using summary-level data are not addressed).³¹ Guidance from both the EMA and the FDA show openness to RWE and trial design modernization to support regulatory decision making, and both agencies continue to stress the importance of early consultation and engagement with regulators to inform RWE study design strategies as needed.^{8,13,32} In general, the FDA guidance is more prescriptive whereas the EMA tends to provide general guidance on RWE use, leaving the sponsor more flexibility to determine the most appropriate approach. Both the FDA and the EMA encourage conducting sensitivity analyses to assess the impact of different demographic factors on the study outcomes to ensure that the study results are robust and can be extrapolated to the broader patient population. In regard to submission of individual patient-level data (IPLD), the EMA does not require access to this data whereas the FDA requires access to IPLD for external comparator studies derived from IPLD but not for benchmarking studies using summary-level data.

There are a number of stakeholders with multidisciplinary expertise directly and indirectly involved in providing feedback to regulators and/or influencing RWE policy. Some of these stakeholders

include trade associations, such as the Biotechnology Innovation Organization (BIO), Pharmaceutical Research and Manufacturers of America (PhRMA), and Transcelerate BioPharma Inc., scientific organizations, such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), coalitions such as the RWE Alliance and Friends of Cancer Research as well as sponsors themselves. The June 2022 International Coalition of Medicines Regulatory Authorities (ICMRA) workshop on RWE identified 4 areas of opportunity for regulatory collaboration which could help address common challenges and further enable the integration of RWE into regulatory decision making including (i) harmonization of RWD and RWE terminologies, (ii) convergence on RWD and RWE guidance and best practice, (iii) readiness, and (iv) transparency.³³ The workshop proceedings concluded that, going forward, it will be important for regulators to strike a balance between clear, prescriptive guidance that also offers sufficient space for further innovation and creativity without compromising the quality of submissions.³⁴

The HTA bodies are also pursuing expanded use of RWE. For example, in June 2021, the French National Authority for Health (Haute Autorité de Santé (HAS)) published a guide on real-world study methods following their 2020 action plan for the assessment of innovative medicines.⁹ In June 2022, the UK National Institute for Health and Care Excellence (NICE), published a framework for RWE, with considerations for incorporating best practices across study design, execution, and reporting phases; these include items related to study planning and protocol development, use of sensitivity analyses to test robustness, and transparency in the reporting process.¹² The guidance from the German Independent Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)) and the Federal Joint Committee Gemeinsamer Bundesausschuss (G-BA) has been more limited, highlighting the importance of confounder selection, prespecified sensitivity analyses, and the use of IPLD.¹¹ As such, further guidance on RWD-derived external controls in the HTA is needed to accelerate their use and acceptability for reimbursement.

We recently conducted an SLR evaluating 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 HTA agencies for all indications.²⁵ The review summarized publicly available information, identifying 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 engagement, methods for addressing missing data, and selection of real-world endpoints. Our review found that submission of the same RWE from the same data source often received different feedback when submitted to multiple agencies, suggesting that further international harmonization and standardization of RWE best practices is of value.

This landscape of knowledge has informed the considerations we elucidate below. We also present four recent regulatory and HTA submissions as case studies, each illustrating the ways in which the operational and methodological considerations work together to influence regulatory/HTA outcomes. The following case studies

were selected as they are representative of different regulatory and HTA bodies, illustrative of both positive and negative decisions, and vary in their feedback and key messages. Taken together, we hope that these considerations and case studies will be helpful to industry sponsors and clinical researchers considering how best to design and implement their own RWD-derived external control study.

KEY TAKEAWAYS

Early engagement with regulators and HTA bodies during study planning phase is recommended, but not a guarantee of acceptance

The 2021 FDA draft guidances recommend that sponsors engage with the FDA in the early stages of designing a noninterventional study intended to support a marketing application to discuss Agency expectations for the design and conduct of their studies.^{13–16} In our recent SLR, early engagement with regulators and HTA bodies was noted as good practice to support the proper use of RWD-derived external controls. For example, the FDA noted that the external control study for tazemetostat, indicated for metastatic or locally advanced epithelioid sarcoma, was performed without regulatory input on the protocol or approach. Although the sponsor sought to demonstrate efficacy through a single-arm trial and RWD-derived external comparator, the FDA “did not consider the design of the study adequate to provide direct or relevant evidence of any aspect of efficacy reviewed.”³⁵ Similarly, the FDA’s review of selinexor, indicated for relapsed/refractory multiple myeloma, stated that “without having reviewed and consented to a protocol and statistical analysis plan [SAP],” they could not be certain that the protocol and SAP for the external control study were prespecified and unchanged during the data selection and analyses.³⁶

The timing of engagement is critical, and sponsors should consider engaging with regulators and HTA bodies while there is still flexibility in the choice of RWD source and study design (i.e., during the study and analysis planning stage) in order to receive feedback that can be readily considered and implemented. In this context, sponsors should critically assess (i) when to engage, (ii) what type of evidence is required for early engagement, and (iii) what questions they should be asking during the engagement process.

The EMA, the FDA, and HTA bodies each offer their own mechanisms for early engagement. For the EMA, this may take the form of non-binding Scientific Advice (available at any stage) or Protocol Assistance (for orphan medications).³⁷ The FDA recommends that protocols and SAPs be submitted for review and comment in advance of study commencement,¹⁵ and that study protocols be posted on a publicly available website, such as ClinicalTrials.gov.³⁰ For HTA applications, Joint Scientific Consultations (JSC) are recommended after feasibility assessment and proof of concept are complete but before study kick-off in order to provide nonbinding advice to the sponsor.³⁸ Additionally, the 2021 European Union HTA procedure will require new medicines to undergo joint European assessment.³⁹

When designing an RWD-derived external control study, sponsors may request a type B (including end-of-phase II), type C, or type D pre-new drug application (NDA) and biologics

license application (BLA) meeting with the FDA. The FDA guidance provides a detailed outline of the information to be provided in a meeting request and meeting package.⁴⁰ Knowledge of the division conducting the review is important. When a sponsor submits a request for a meeting with the FDA to discuss the external control study, sponsors should consider asking for an RWE subcommittee consult so that the study can receive input from those with the most experience in this domain as unique expertise is needed to evaluate RWE as evidence of drug effectiveness. More broadly, the ability to adequately assess RWE in support of effectiveness (instead of or in addition to safety) continues to mature across regulatory and HTA bodies as education on and experience with such studies increases.

The initial meeting request will typically include questions related to both the trial and the RWD-derived external control study design. Although specific questions will vary based on the particulars of the application in question, potential question topics include the proposed RWD source(s), eligibility criteria, acceptability of endpoints and their operationalization in RWD, representativeness of the standard of care treatment in the external control group, and methods for addressing confounding.

It is also at this stage that the sponsor will determine whether the RWD-derived external control arm will serve as a formal comparator or a benchmark. Briefly, external controls without any direct comparison (e.g., a reported relative effect estimate of hazard ratio/odds ratio and corresponding *p*-value) are generally classified as benchmark whereas those with a direct comparison are classified as a formal comparator. In general, both benchmark and comparator cohorts should reflect a similar patient population and eligibility criteria as the reference trial, although benchmark studies generally do not attempt to control for confounding through analytic features (e.g., matching, inverse probability of treatment weighting (IPTW)) to draw causal inferences.⁴¹ If possible, a prespecified, adequately powered formal comparison is a stronger design; however, this decision will depend on factors such as the overall regulatory strategy, RWD source selected, availability and quality of IPLD, and the resources required to conduct an external control arm for formal comparison, which are often more operationally complex and may require supplemental data approaches. Such supplemental data approaches include data linkage or curation (e.g., chart review/abstraction).

If the meeting request is accepted, sponsors should aim to optimize the efficiency and value of the feedback by providing regulators with detailed qualitative and quantitative information on the proposed data source(s), including data relevance/reliability, and analysis plan prior to the interaction. Of note, sponsors generally submit a study synopsis for the trial and a separate study synopsis for the RWD-derived external control study. The synopses can reference each other for analytic descriptions, but it is important to consider how regulator feedback on the RWD-derived external control might affect the clinical trial protocol. According to the FDA’s 2023 draft guidance on externally controlled trials, sponsors should finalize a study protocol before initiating the externally controlled trial, including selection of the external control cohort and analytic approach, rather than selecting an external control cohort after the completion of a single-arm trial.³¹

Despite the majority of submissions in our prior review having sought advice, we found that early engagement (e.g., Scientific Advice and Protocol Assistance) did not guarantee that the RWE was considered in the final authorization application. There are a number of potential explanations for such a discrepancy. One major reason is the RWD sources often have limitations, including data missingness or nuances related to the nature of data collection in clinical practice, that are often not possible to determine until data abstraction or analysis is complete. Another reason may be related to insufficiencies in the early engagement process in which sponsors may benefit from being more thoughtful and strategic in the questions they pose, the information they provide, and the way they implement feedback. Further, once sponsors receive the agency's (nonbinding) recommendations, they may incorporate that feedback at their own discretion but recognize that adjustments may generate other issues. In 2018, Gilead Sciences submitted an extension of indication to NICE for treatment of follicular lymphoma (FL) with idelalisib (**Box 1**).⁴² Based on early discussion with NICE, a propensity score (PS) analysis and an additional indirect comparison were incorporated into the protocol; however, the submission ultimately received a negative recommendation due to various other methodological concerns (e.g., differential variable definitions and lack of sensitivity analyses). This case

study demonstrates that early engagement does not in itself ensure a positive final decision. In general, it is difficult to quantify using publicly available information how many companies have gone through the avenues available for early engagement and, as such, it is difficult to gauge whether industry and regulatory agencies/HTA bodies are maximizing the opportunities to share knowledge and collectively further regulatory science. Furthermore, this discussion does not capture information on RWE studies that were not publicly disclosed as they were evaluated internally and/or with the FDA/EMA and, based on feedback, not further pursued by sponsors.

Consideration of the appropriateness and comparability of external controls across multiple dimensions, including eligibility criteria, temporality, population representation (e.g., geography, race/ethnicity, and sex), and clinical evaluation

A benefit of randomized controlled trials (RCTs) is that, by assigning study arms randomly from individuals within the same cohort, it is possible to optimize comparability between treatment and control groups. Such comparability is not as readily attained with RWD-derived external controls. RWD-derived cohorts may differ from trial cohorts across a variety of dimensions, leading to

Box 1 Case Study 1: In 2018, Gilead Sciences submitted to NICE an extension of indication for Zydelig (idelalisib). Previously approved for chronic lymphocytic leukemia, the submission requested that the indication be extended to include FL. In addition to the DELTA study (a phase II single-arm study), the submission included RWD from the UK HMRN, a population-based registry. HMRN was compared with DELTA using PS analysis and indirect comparisons. The final NICE recommendation was negative.

Guiding principle	Experience
Early engagement	The sponsor-initiated discussions with NICE prior to formal submission. A PS analysis and an additional indirect comparison were incorporated into the protocol based on the early discussions.
Appropriateness and comparability of external controls	The sponsor selected a range of therapeutic comparators including best supportive care. However, NICE expressed concerns about elements of comparability between the SAT and RWD including differences in baseline characteristics and differences in prior lines of therapy that might impact patient outcomes. NICE concluded that the association between treatment and survival may have been confounded.
Adequate sample size	Limited sample size ($n=34$) of HMRN cohort was noted by NICE.
Strategy for assessing and addressing data quality	HMRN was considered to be reflective of the NHS for time to progression, and the only source of comparative data for the UK. However, there was no information included regarding data quality or missingness.
Meaningful endpoints and appropriate analytic methods	Longer-term survival data for the trial and RW cohorts were provided to reduce uncertainty in effectiveness. However, patients were not censored at transplantation, which can improve outcomes, thereby invalidating effectiveness findings. Differential scan frequency between the trial and RWD impacted the measurement of endpoints, for example, PFS. There was no systematic review of risk factors presented for progression or death and there was no rationale provided for the selected matching method. Multiple matching adjusted indirect comparison methods were used; however, missing data and confounding impacted reliability.
Sensitivity analyses	No sensitivity analyses using different PS matching methods were conducted.
Consideration of RWE in final decision:	"There are several complex indirect comparisons of idelalisib, but these are based on sparse data and have other problems. So, it is unclear whether idelalisib is better than individual chemotherapeutic regimens currently offered by the NHS and, if so, by how much."
Key message:	Early engagement may not ensure a positive final decision in the absence of other operational and methodological considerations.
FL, follicular lymphoma; HMRN, Hematological Malignancy Resource Network; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; PFS, progression-free survival; PS, propensity score; RW, real world; RWD, real-world data; RWE, real-world evidence; SAT, single-arm trial; UK, United Kingdom.	

unmeasured confounding; this is especially true given that RWD is generally collected in the absence of a formal study protocol⁴³ and its primary purpose is not generally research. A recent systematic review of oncological NDAs and BLAs noted an FDA emphasis on the comparability of RWD-derived external control and trial populations, recommending that applicants attempt to match trial eligibility criteria and adjust baseline characteristic imbalances.⁴⁴ Indeed, comparability between cohorts must extend across multiple dimensions, including sources of data, eligibility criteria, temporality, population representation, and clinical evaluation.

Sources of data. The process of establishing comparability between the trial and RWD-derived external control cohorts begins with specifying an appropriate research question, which then drives RWD source selection. Generally, this will require a data landscaping exercise to identify putative data sources based on high-level criteria such as data source type (e.g., electronic medical/health record [EMR/EHR] data, chart review, and registry), geographic regions (e.g., the United States and the European Union), desired population, key endpoints, and necessary sample size. After preliminary data sources have been identified, sponsors should conduct quantitative and qualitative feasibility assessments. Depending on the RWD sources, this may involve sending a feasibility questionnaire to sites to obtain specific details on variable capture, potential for missingness, follow-up duration, and preliminary patient counts. Even in the absence of direct access to the data, the outputs of this initial feasibility query will inform the fit-for-purpose data assessment, in which dimensions of data relevancy and reliability are evaluated for a potential regulatory decision within the context of a specific disease state or therapeutic area.⁴⁵ Although this assessment is critical for initial feasibility, this does not guarantee that the data are fit-for-purpose until final data abstraction or analyses are complete. Importantly, some RWD sources may require expert curation of unstructured data from EHRs or “omic” (e.g., genomic and proteomic) data to identify key variables or events of interest.⁴⁶

Eligibility criteria. Once the data source(s) has been selected, eligibility criteria from the trial cohort should be applied to the RWD-derived external control to the extent possible, as this will support comparability (although additional adjustments may be needed to balance confounders following application of eligibility criteria).⁴⁷ It is important that the definition of patient populations be informed by the specific research question to facilitate drawing external control patients from the same source population as the reference trial. This approach may enable better balance in the baseline characteristics between the two populations; however, there are many challenges to implementing strict inclusion/exclusion criteria in practice. These challenges will be specific to the RWD source selected but are often attributable to incomplete variable capture for important prognostic or laboratory values (e.g., Eastern Cooperative Oncology Group (ECOG) score, expectation that patients will live ≥ 6 months). In some cases, additional curation of certain variables via chart review may be needed, such as human epidermal growth factor receptor 2

(HER2) status. Proxy variables may be used to approximate the trial eligibility criteria where possible (e.g., requiring no records of ECOG performance status > 2 or Karnofsky < 60 as documented by the treating physician as a proxy for ECOG performance status ≤ 2 in the single-arm trial), and the protocol should justify choosing different eligibility criteria when applicable. Sponsors should consider providing justification for discrepancies in eligibility criteria and rationale for why the resulting populations may be assumed to be similar. The submission for idecabtagene vicleucel to the EMA, approved for the treatment of relapsed/refractory multiple myeloma in 2021 (**Box 2**), demonstrates the effect strict inclusion criteria in the trial population can have on the external validity of the trial and interpretation of the contextualizing RWD-derived external control.⁴⁸ If the patient population in the trial is highly selected, it may no longer be considered representative of a real-world patient population. Depending on the RWD source, particularly for a benchmarking study relying on published literature, it may not be feasible to match the eligibility criteria between the trial and external control cohorts.

Tazemetostat, approved by the FDA for the treatment of epithelioid sarcoma in 2020, is another recent submission in which the lack of comparability between the trial and RWD-derived external control as defined by the eligibility criteria was a major driver of the FDA’s negative opinion on the use of RWD. Specifically, the differences in age inclusion criteria, years during which patients received treatment, and integrase interactor 1 status called into question the validity of the outcomes in the RWE study. In their feedback, the FDA highlighted that the protocol for the historical study should clarify which, if any, prior cancer therapies should be discontinued before selection into the study.³⁵

With regard to index date, a similar index event should be selected for the trial and external control to support cohort comparability. In their review of selinexor, the FDA identified that systematic differences in how the index date was defined between the trial and external control cohorts may have introduced immortal time bias, which is defined as a span of cohort follow-up during which, because of exposure definition, the outcome under study could not occur.⁴⁹ As such, the FDA made specific and detailed analytic suggestions to address this bias, illustrating the importance of careful decisions and specialized expertise in designing and implementing the real-world analytic plans.³⁶

Population representation. When selecting RWD-derived external controls, stakeholders often grapple with identifying an appropriate and comparable patient population with respect to geographic and demographic representation. This includes diversity in the countries, study centers, races/ethnicities, and sexes represented.

Representation of groups that have been historically underrepresented in clinical trials, including racial and ethnic minorities and rural populations, is a critical consideration when selecting an RWD source.⁵⁰ Although essential efforts are underway to improve equitable access to and participation in clinical trials, it is important for sponsors to consider that the representation and inclusion of such groups may vary between RWD and clinical data sources.⁵⁰

Box 2 Case Study 2: In 2020, Celgene Europe BV submitted to the EMA an original submission for Abecma (idecabtagene vicleucel) for the indication of multiple myeloma in patients who have received at least 3 prior therapies, including an immunomodulatory agent, a proteasome inhibitor and an anti CD38 antibody. In addition to study MM-001 (KarMMa; a phase II single-arm study), the submission included RWD from study NDS-MM-003 (data sources including clinical sites, registries, and research databases collated in a single data model). Study NDS-MM-003 was compared with KarMMa using IPTW. The final decision was Conditional Approval.

Guiding principle	Experience
Early engagement	The sponsor received protocol input regarding study design and study population for the proposed phase II registration study, MM-001.
Appropriateness and comparability of external controls	Due to the strict inclusion criteria of the pivotal trial, the general representativeness of the historical control was unclear. Further, the representativeness of the historical control for the study population included in MM-001 was compromised by the extended baseline data collection period (i.e., 60 days) and the overlapping recruitment periods for the historical control and trial at the same centers.
Adequate sample size	Limited trial sample size ($n=128$) was noted by the EMA.
Strategy for assessing and addressing data quality	For highly prognostic covariates the analysis permitted up to 30% missing data. Missing data for important prognostic variables and therefore exclusion from the PS model was cited as a concern.
Meaningful endpoints and appropriate analytic methods	The short duration of follow-up in the trial was noted by the EMA as impacting the ability to ascertain long-term benefit in OS.
Sensitivity analyses	No sensitivity analyses were conducted.
Consideration of RWE in final decision: "Contextualization of the effect is challenging. The robustness of the adjusted indirect treatment comparison based on the RWD is difficult to verify, considering the rather selected study population, and the missing data of several important prognostic factors."	
Key message: Sponsors should consider how an overly selected trial population may limit representativeness of RWD.	

EMA, European Medicines Agency; ORR, objective response rate; OS, overall survival; PS, propensity score; RWD, real-world data; RWE, real-world evidence; IPTW, inverse probability of treatment weighting.

With respect to geography, there is a call to move away from single country trials and toward multicountry trials.⁵¹ In general, if a sponsor is planning to submit to both the EMA and FDA, then the RWD source(s) might include patients from multiple countries so as to match the trial population as closely as possible. At a more granular level, sponsors should aim to avoid differential selection of comparison groups (e.g., if trial patients were enrolled from academic medical centers and controls were enrolled from community oncology centers, or substantial differences in standard of care treatments across regions). Although these considerations are important from a regulatory perspective, they become more pronounced from the HTA perspective. If the RWD is not from the country of submission, the generalizability of the results to clinical practice may be called into question as data from the RWD-derived external control may not be transferable to the country-specific health care context due to structural differences in the healthcare systems or differences in approved drugs and/or standard of care treatments. Because of this emphasis, sponsors may need to utilize different RWD sources when submitting to HTA bodies in different countries and/or consider stratifying results by country.

Selecting an appropriate control condition means defining the standard of care, which is characterized by treatment received (either an individual treatment or a group of treatments). Submissions to regulators and HTA bodies often report that patients in the RWE cohort received "standard of care" without defining what this means in the patient population of interest. In our SLR, we observed that regulatory and HTA bodies reported

positively on submissions that characterized standard of care in detail and negatively on those where this detail was either insufficient or not provided. HTA bodies in particular were more likely to accept data they viewed as being representative of national clinical practice⁵² and were prone to dismiss data that they deemed unrepresentative or not generalizable. Regulatory and HTA reviewers have noted a lack of comparability in applying standard of care data from one national health system to another or combining data across multiple countries.⁵³ A well-defined, comparable, RWD-derived external control cohort should be provided and characterized across multiple dimensions, such as medications received, dose, interval, procedures, and healthcare utilization. Importantly, whereas external controls are often used to contextualize findings from uncontrolled or single-arm trials, external controls may also be used in controlled trials for which there are different local standards of care and thus the control condition of interest may differ for regulatory and reimbursement considerations.

Temporality. A key component of defining standard of care is the temporality of the RWD-derived external control cohort (historical vs. contemporaneous). This can create challenges in assessing the impact of trial treatments against care and treatments received in the real world.⁵⁴ Time trends and geographic differences between cohorts can have a major impact on the treatment(s) received and, ultimately, the findings of a trial using an RWD-derived external control to contextualize results.^{54,55} As available, sponsors are encouraged to consider RWD sources that

would support analyses within the same or similar time periods as the trial (i.e., not distant past) and provide data derived via similar methods. If historical data are being used, it is important to understand if and how the standard of care has changed over that time period. If the standard of care has improved over the course of retrospective data collection, it may be prudent, for example, to use only the portion of the RWD that is contemporaneous with the trial.

Clinical evaluation. Ideally, the RWD-derived external control cohort should have the same medical workup and clinical evaluations as the reference trial. For example, the recent autologous CD34+ cell encoding arylsulfatase A (ARSA) gene submission (approved by the EMA in 2020 for the treatment of metachromatic leukodystrophy), was commended for reducing the variability in clinical assessments across the treated and natural history patients by using the same staff and methodologies; however, this is not always possible.⁵⁶ For example, in the context of oncology, information on timing and follow-up scans with information on tumor status at each assessment may not be available. Determining the timing of a diagnosis from clinical/EHR data alone can be challenging because not every diagnosis is recorded at every visit, the first diagnosis in a database is not necessarily an incident case of disease, and RWD may not capture all therapies received by individual patients. Thus, defining the number of lines of therapy that patients have previously received may be problematic. It is in the best interest of stakeholders to anticipate these issues and convey them to regulators early on along with potential mitigation strategies. Such mitigation strategies may include validation of mortality data at the site level, implementing standardized line of therapy rules to be applied across sites, and linkage to other data sources. From the HTA perspective, information on the number, type, and/or duration of prior therapies for each patient should be provided if possible. This includes providing detail on precise combinations of chemotherapies/chemotherapy regimens or best supportive care (which may change over time) and data on the frequency of scanning in the RWD-derived external control cohort compared with the trial cohort. These details can allow sponsors to anticipate potential bias and the direction that bias might occur in the eventual interpretation of results.

Ensuring adequate sample sizes, including hypothesis testing considerations

As mentioned above, RWD-derived external controls can be particularly useful in the context of rare disease. Investigators that may not be able to identify and recruit a sufficiently large population to conduct an RCT may instead turn to an RWD-derived external control to contextualize the findings of a reference trial; however, these external controls may also be subject to small sample sizes. Given that RWE is inherently subject to higher uncertainty, this can lead to difficulty in establishing the magnitude and significance of impacts observed between groups. In our SLR, we observed that regulators and HTA bodies often did not consider external controls because the sample size was too small (i.e., $n \leq 50$) to allow comparison to the trial group, conduct analyses in meaningful subgroups, and include the data variability and covariates necessary to support adjustments.³ Small sample sizes

may be conducive to benchmarking rather than formal comparison, although sample size of the reference trial is still a factor in regulatory decision making even if no hypothesis testing is planned. Given these concerns, sponsors face ambiguity in determining (i) what constitutes a sufficiently large sample and (ii) how to maximize the chances that the minimum sample size requirements are met.

With regard to the former, the FDA guidance clarifies that formal sample size and power calculations remain relevant in RWE-based studies.³⁰ Venz and colleagues⁵⁷ outline a useful approach for estimating sample size for a targeted power; however, there is no formal rule of thumb, and dimensions such as data missingness can impact the sample size required to reliably establish a meaningful impact.^{3,58} Sponsors should conduct *a priori* sample size and power calculations, presenting these to the relevant regulatory body for comment before commencement of their study. Of note, sponsors may choose to power the trial as a single-arm trial and then separately power for comparative outcomes in the RWD-derived external control protocol. This way, if regulators do not agree with the external control approach, the sponsor will at least have a target sample size for the trial.

Methods to optimize sample size can also vary. Specifically, statistical power can be maximized in comparative studies by selecting an analytic method (e.g., IPTW) that allows researchers to leverage information for all patients included in the sample. A recent publication by Rippin and colleagues (2022)⁵⁹ outlines key considerations for determining the appropriate external control sample size. Of particular interest to sponsors is the recommendation to have an external control arm which is substantially larger in terms of patient number than the reference trial (e.g., in the range of 1.2–1.5) depending on feasibility considerations for rare diseases.

Sponsors may include multiple RWD sources in an external control to increase the sample size (for adequate power, subgroup analyses, etc.). There are, however, some tradeoffs to having a heterogeneous external control population. There could be heterogeneity in terms of the baseline characteristics of the populations themselves, which variables are collected, or how those variables are measured. There is also the potential for heterogeneity in outcome assessment (e.g., response rate). For example, the external control for axicabtagene ciloleucel (SCHOLAR-1), submitted to the EMA in 2017 for treatment of patients with relapsed/refractory diffuse large B-cell lymphoma (DLBCL), primary mediastinal B-cell lymphoma, and transformed FL, was composed of data from 2 randomized phase III studies and 2 observational databases (MD Anderson Cancer Center and Mayo Clinic/University of Iowa Specialized Program of Research Excellence (SPORE)). As such, evidence of heterogeneity in response rate among the institutions comprising the external control was assessed to determine whether data from the four institutions could be combined.^{60,61} In general, sponsors should assess how data elements are collected across different data sources. If there are major differences in how study variables are collected and defined, it may be preferable to analyze the data sources separately. Understanding the implications of analytic methods, data pooling, and population representation on sample size is important for stakeholders conducting their own external control studies.

Implementation of a clear and transparent strategy for assessing and addressing data quality, including data missingness across trials and RWD

Data quality includes accuracy, completeness, traceability, and transparency.⁴⁵ The extent to which a sponsor can characterize data quality for a given data source depends on access to IPLD. Sponsors will often need to rely on assessments of data quality conducted by others (e.g., data owners). If a data source is not accurate, complete, and/or traceable, sponsors should ensure they are transparent in their reporting of the gaps and deficiencies in the data source to regulators and HTA bodies. Because RCT data goes through source verification and monitoring, but RWD generally do not, it is important to characterize how the RWD were collected, entered, derived, and manipulated. Approval of an external data source by a data governance committee beforehand can also support transparent usage of RWD, encourage the use of fit-for-purpose data, and ensure only data of appropriate quality are curated. Furthermore, sponsors and/or data suppliers should be prepared to provide all datasets in required formats, programming code, and accompanying documentation to regulators and others.

As opposed to the separate but related concept of validity, which is often viewed dichotomously (i.e., data are valid or invalid), data quality can be viewed as a continuum that is modifiable (i.e., there are certain actions that can be taken to improve data quality). A key dimension of data quality that sponsors must address in RWD is missingness. Missing data can bias estimates of effect between cohorts, especially when one cohort is more likely to have missing data.

Sponsors may consider developing and reporting a strategy *a priori* for transparently assessing data quality (including missingness) between trial and RWD-derived external control cohorts, as well as a plan for imputation of missing data. However, sponsors may face ambiguity in (i) understanding whether to proceed with imputation, and (ii) the most robust imputation methods.

Missing data should be considered from the outset and built into the study planning process. For example, RWD selection should consider the amount of missing data for high-priority data elements, and methods for reporting and imputing missing data (both for the reference trial and RWD-derived external control cohort) should be outlined explicitly in the SAP. It is also good practice to report a descriptive characterization of data missingness across cohorts, promoting transparency in data missingness.^{14,53} Importantly, missingness considerations differ depending on whether the missingness is for a baseline characteristic/covariate or an outcome variable. A useful rule of thumb is to set thresholds for missing data, which may vary based on the prognostic importance of particular variables, as well as how that variable is being used (e.g., eligibility criteria and PS-matching). For example, variables for matching with missingness under 5–10% can likely be included in most situations, whereas those with ≥ 30% missingness should likely be excluded from any matching analyses⁶² and the utility of the data source may need to be reconsidered. Missingness in the outcome variable (e.g., missing assessment of response) may be less well-tolerated as it cannot be properly estimated or adjusted for.⁶³

Imputation is one strategy that has been used by stakeholders as a means of testing different missingness assumptions; however, there is no clear guidance from regulators or HTA bodies on when to use various imputation methodologies. In our recent SLR, imputation to address missing data was infrequently documented in regulatory and HTA submissions. If imputation was conducted, it was usually done as a sensitivity analysis to address missingness in trial data with minimal consideration of RWD missingness. Although reviewers noted this as a limitation, clear guidance related to good practice for dealing with missing data in RWD-derived external controls is lacking.⁶⁴

Dataset creation and curation is a distinct concept from data quality. Some RWD sources may contain key data elements in an unstructured format that require further curation. During initial feasibility and early engagement with regulatory or HTA bodies, the sponsor may choose to: (i) provide an aggregate output of the structured data and work with the data source to describe how unstructured data will be validated and extracted; (ii) curate a limited random sample of the identified eligible patients; or (iii) curate the entire eligible population and provide a complete summary of data feasibility. Relying on the structured data alone, for example, may result in inflated population counts during quantitative feasibility due to limited application of eligibility criteria. However, curation often requires additional monetary and time investment, particularly for rare conditions or for indications in which patient identification requires deep curation on a high volume of patients.⁴⁶ As such, the sponsor and/or data supplier will need to assess the challenges involved with each approach if curation is needed.

Selection of comparable, meaningful endpoints that are operationalized and analyzed using appropriate analytic methods

A critical decision for sponsors is the selection of real-world endpoints that are relevant to the patient population and can be readily operationalized and interpreted for a given RWD source. Once qualitative and quantitative feasibility confirms a given endpoint can be reliably assessed, the sponsor must propose an analysis plan that will adequately control for confounding and mitigate biases that preclude meaningful interpretation of the proposed endpoints. Sponsors will need to determine (i) how to justify, operationalize, and validate real-world endpoints as well as (ii) appropriately balance key prognostic factors in the analysis stage.

Endpoints. Sponsors aim to assess outcomes in RWD similar to the key endpoints in trials but may face challenges in determining how real-world endpoints should be operationalized and assessed. As expected, these challenges vary depending on the nature of the endpoint. For example, because real-world tumor response is assessed and described in a variety of ways in clinical practice, the data usually carry large variability.⁶⁵ More commonly used oncology endpoints, such as real-world objective response rate (ORR), present their own set of considerations. For example, sponsors will need to decide if centralized adjudication of endpoint(s) will be conducted (instead of or in addition to treating physician evaluation), what proportion of patients should be selected for validation, and how. According to the

FDA, if imaging data in the RWD-derived external control study have not been assessed by ≥ 2 independent readers blinded for patient exposure and timepoint then there is potential for subjectivity in the assessment, as noted in the selumetinib submission (**Box 3**; approved in 2021 by the EMA for the treatment of neurofibromatosis type 1 and symptomatic, inoperable plexiform neurofibromas (PN)).⁶⁵ One approach is to assess ORR per local investigator assessment and obtain independent central review (ICR) for the subset of patients whose local scans demonstrated a prespecified decrease (e.g., 20%).⁶⁶ Alternatively, all images may be assessed by ICR. This was the approach taken for trastuzumab deruxtecan, approved by the EMA in 2020 for the treatment of HER2-positive breast cancer: 2 trained radiologists read each study subject's images independently, according to the ICR Charter. In case of discordance, adjudication was performed by a third radiologist. The adjudication variables were best confirmed response and date of progression. Indeed, if ICR and investigator assessments are used, concordance should be assessed and reasons for discordance provided, if feasible.⁶⁷ In the FDA's review of moxatumomab pasudotox-tdfk, approved in 2018 for the treatment of hairy cell leukemia, discordance between blinded ICR (BICR) and investigator's assessment in durable complete response (CR) and best ORR was deemed a major statistical issue. The FDA requested that the sponsor explain the discordance: the investigators determined CR earlier than BICR and thus, by the data cutoff date, some patients did not reach the durable CR per BICR, and some patients did not achieve a CR per BICR due to nonevaluable bone marrow. The main reason for discordance of best overall response was missing disease response assessment by BICR.⁶⁸

Our SLR found that both the FDA and the EMA raised concerns about the implementation and interpretation of time-to-event endpoints, citing 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. Common

time-to-event endpoints in real-world oncology studies include real-world overall survival (OS) and real-world progression-free survival (rwPFS). Although less frequently used, real-world time to next treatment (rwTTNT) can serve as a proxy for rwPFS that may be more easily estimated and interpreted using RWD (e.g., EMR/EHR data).^{66,69} Specifically, manual chart abstraction is typically required to capture rwPFS, whereas rwTTNT can be calculated based on structured EHR data alone, using the dates of various procedures and diagnoses.^{69,70} Importantly, rwTTNT may overestimate progression events if patients initiate another treatment due to tolerability or may be systematically longer than rwPFS as new treatments will be initiated after progression has been identified.⁶⁹

An example of the FDA's criticism of time-to-event endpoints comes from tafasitamab-cxix, approved for relapsed/refractory DLBCL by the FDA in 2020 (**Box 4**).⁷¹ The FDA considers information from time-to-event endpoints difficult to interpret due to inherent differences between clinical trials and clinical practice in the elements, timing, and accuracy of measurements.⁷¹ Importantly, although OS may be viewed by regulators and HTA bodies as more objective than other time-to-event endpoints, the maturity of trial data may limit its utility. Although endpoint selection is primarily driven by endpoint hierarchy in the reference trial, when proposing time-to-event endpoints for regulatory/HTA input, stakeholders should consider justifying the appropriateness of the endpoints clinically (in the context of the indication) and in terms of the available data (from feasibility assessments). The more information that can be provided earlier on endpoint operationalization and RWD data quality, the more informed and useful the feedback from the agencies will be.

Analytic methods. As there will not be perfect balance between the baseline characteristics of the trial and RWD-derived external control cohorts, sponsors often need to implement analytic

Box 3 Case Study 3: In 2020, AstraZeneca submitted to the EMA an original submission for Koselugo (selumetinib), for the indication of NF1 and symptomatic, inoperable PN in pediatric patients aged 3 years and above. In addition to study D1532C00057 (SPRINT Phase II Stratum 1; a phase II single-arm study), the submission included RWD from study NCI-08-C-0079, an observational, prospective, natural history study. The final decision was Approval.

Guiding principle	Experience
Early engagement	The sponsor received protocol assistance including on the single pivotal study strategy and use of external historical control data.
Appropriateness and comparability of external controls	The comparison was limited due to differences in study design (e.g., eligibility criteria [natural history study included a broader population], demographic and disease characteristics, and assessment of images).
Adequate sample size	The natural history study included a significant number of patients with sufficient follow-up (plans to enroll up to 250 patients with follow-up of up to 10 years).
Strategy for assessing and addressing data quality	No specific strategies for assessing or addressing data quality were noted for the RWD.
Meaningful endpoints and appropriate analytic methods	Unlike the single-arm trial, endpoints were not assessed by ≥ 2 independent readers who were blinded for patient exposure and timepoint.
Sensitivity analyses	No sensitivity analyses were noted for the RWD.
Consideration of RWE in final decision:	"The comparison to the external controls provided to contextualize the data was merely of a descriptive nature and hampered by differences between the design of the studies."
Key message:	If feasible, sponsors should consider utilizing an ICR if assessing imaging data.

EMA, European Medicines Agency; ICR, independent central review; NF1, neurofibromatosis type 1; PN, plexiform neurofibroma; RWD, real-world data; RWE, real-world evidence.

Box 4 Case Study 4: In 2019, Morphosys US Inc. submitted to the FDA an original submission for Monjuvi (tafasitamab-cxix) in combination with LEN, for the indication of relapsed or refractory DLBCL not otherwise specified, in patients ineligible for allogenic stem cell transplant. In addition to study MOR208C203 (L-MIND; a phase II single-arm study), the submission included RWD from study MOR208C206 (RE-MIND), an observational, retrospective, cohort study using EHR data. RE-MIND was conducted to generate a matched control cohort for L-MIND. The final decision was Approval.

Guiding principle	Experience
Early engagement	The sponsor had a type C meeting with the FDA on November 9, 2018, to discuss RWD to provide evidence for LEN monotherapy cohort, including using a matching approach. On April 15, 2019, the FDA provided response and comments on RWD proposal for observational retrospective cohort of LEN monotherapy.
Appropriateness and comparability of external controls	The FDA did not believe the studied cohorts to be representative of relapsed/refractory patients in the target population. The distribution (marginal and joint) of the observable covariates were not well-balanced.
Adequate sample size	Limited sample size ($n=76$) was noted by the FDA.
Strategy for assessing and addressing data quality	Imputation of missing baseline covariates was conducted as a sensitivity analysis.
Meaningful endpoints and appropriate analytic methods	The FDA did not find the time-to-event endpoints to be interpretable. The FDA also expressed concern regarding the relevance of the selected covariates and whether they sufficiently addressed confounding. Potential outcome misclassification was possible as not all patients had response assessed by ICR.
Sensitivity analyses	Additional sensitivity analyses were conducted. These included analysis of PFS and EFS with modified censoring rules, imputed missing baseline covariates, and revised analysis sets including patients not meeting the 6-month follow-up rule.
Consideration of RWE in final decision: "However, there were several limitations to the observational data and formal comparisons to the data from RE-MIND were not made."	
Key message: The inclusion of sensitivity analyses is critical to assess the robustness of findings.	

DLBCL, diffuse large B-cell lymphoma; EFS, event-free survival; EHR, electronic health record; FDA, US Food and Drug Administration; ICR, independent central review; LEN, lenalidomide; PFS, progression-free survival; RWD, real-world data; RWE, real-world evidence.

methods, such as PS matching or IPTW. One method is not superior to the other; both PS matching and IPTW require assumptions (i.e., adequate sets of prognostic factors, sufficient sample size, and strict positivity). In particular, regulatory and HTA feedback emphasized the importance of selecting prognostic factors for inclusion in PS models. Stakeholders should consider undertaking a systematic review to identify known prognostic factors of relevance and targeting these for PS methods to control for confounding. Sponsors may also want to consider the inclusion of time-varying covariates. For example, hematopoietic stem cell transplantation (HSCT) may be considered an important time-varying covariate to include in the PS model due to the increased use of HSCT procedures in practice between the time period of the historical control data and the more recent RCT data. Regulatory agencies in particular may conduct their own sensitivity analyses wherein they change the covariates included in the PS model. Assessing covariate balance following PS matching or IPTW often involves checking whether standardized mean differences are < 0.1 ,⁵⁹ calculated as the absolute value in the difference in means of a covariate across the treatment groups, divided by the standard deviation in the treated group.⁷² Larger standardized biases indicate that groups are too different from one another for reliable comparison. It is also recommended that investigators remain blinded to the outcome data when assessing balance.

Whereas the specifics of PS matching and IPTW are beyond the scope of this review, recent studies outline useful methods for

PS analyses in the context of external controls for contextualizing single-arm trials.^{73,74}

Conduct of sensitivity analyses to assess the robustness of findings in the context of uncertainty and sources of potential bias

Despite its potential benefits, RWE carries more uncertainty than RCT data in its internal validity (although not in external validity). In our recent SLR, regulators sometimes requested additional data/analyses or conducted their own sensitivity analyses, including imputation of prognostic variables (e.g., worst case imputation) and variations of key assumptions. There is a recognized need to assess the robustness of findings in the face of missing data and unmeasured or residual confounding⁷⁵; however, sponsors may face ambiguity in determining how best to plan, conduct, and report sensitivity analyses.

The potential for unmeasured or residual confounding is always present in RWE; nonetheless, there is a lack of consensus guidance on good practice in confounding control and sensitivity analysis.^{76,77} Sensitivity analyses should be planned and defined *a priori* in the SAP, designed to address key assumptions and definitions.^{11,78} For example, sponsors may present data for the naïve comparison, matching with/without replacement, different matching ratios (e.g., 1:1 and 1:2), and/or using IPTW vs. stabilized IPTW (sIPTW). If IPTW is used and there are subjects with outlying very large weights due to low PS, the sponsor may consider

performing sensitivity analyses that handle the outlying subjects in different ways (e.g., trim or cap the weights). If a single method is chosen, adequate justification for that method over other methods must be provided.

Sensitivity analyses testing various outcome definitions should be strongly considered where outcome validation is not feasible.⁷⁶ For example, sponsors may analyze time-to-event endpoints (e.g., EFS and PFS) with different censoring rules. The planned sensitivity analyses should be presented to regulators for review and comment during the study planning phase, along with details for the selection of variables and missing data handling strategy, in order to identify any additional analyses that may be of value and feasible in the given datasets. Furthermore, *post hoc* analyses to test emergent assumptions may be utilized by sponsors as clear guidance on variables for inclusion in *a priori* sensitivity analyses is lacking.^{19,77}

During the reporting phase, sponsors should clearly define the baseline characteristics of the sensitivity analysis populations in order to demonstrate that the cohort(s) are comparable to the original cohort of patients (if the patient population changed). Further, applicants should assess the likelihood of their various assumptions in the context of the findings of their sensitivity analyses. In this way, the credibility of their findings, and RWE more generally, can be improved substantially.

CONCLUSION

RWD-derived external controls hold value in supporting assessments of effectiveness and safety, especially in situations where well-controlled clinical trials may be infeasible, impractical, or unethical. This includes oncology indications but also rare diseases where external controls (e.g., using natural history studies or disease registries) present tremendous opportunity. Recognition of this value is driving increased interest in the use of RWE to support applications for regulatory and HTA approvals. Based on feedback provided in past regulatory/reimbursement decisions, this review discussed key considerations for the planning, execution, and reporting of RWD-derived external control studies. These considerations are robust but not exhaustive; for example, evolving considerations related to data sharing and privacy are not discussed here. Nonetheless, this discussion provides a reference for many of the practical challenges and ambiguities sponsors may face.

Increased harmonization within and between regulatory and HTA bodies will support the advancement of these considerations. For example, EMA-FDA Parallel Scientific Advice aims to provide a mechanism for EMA assessors and FDA reviewers to concurrently exchange their views on scientific issues during medicinal product development. Such interactions are expected to increase dialogue between the two regulatory agencies and sponsors from the beginning of the lifecycle of a new product, provide a deeper understanding of the bases of regulatory decisions, and optimize product development.⁷⁹ Similarly, as of 2022, the EMA is offering parallel joint scientific consultation with regulators and HTA bodies which allows sponsors to obtain feedback on their evidence-generation plans to support decision making on marketing authorization and reimbursement of new medicines at the same time.⁸⁰ To increase harmonization between HTA bodies and prevent duplication of work, the new European Union HTA procedure will require

certain new products to undergo joint European assessment.³⁸ If HTA bodies are able to apply the joint clinical reports rather than carry out their own clinical assessment of the same evidence, this may mitigate issues that interfere with market access and reduce business predictability.⁸¹

Last, stakeholders across industry, regulatory agencies, and HTA bodies should be encouraged to share both successes and challenges from their experiences to help further learnings and optimize how and when RWD-derived external control studies are conducted. Such knowledge sharing would ideally facilitate the optimal use of RWD-derived control approaches to get medicines to patients more efficiently and improve their care.

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., R.G.W.Q., J.J.J., and R.E.S. are employees and report ownership interests of Regeneron Pharmaceuticals Inc. J.H. and R.E.S. own 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.

© 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. European Federation of Pharmaceutical Industries and Associations. Evidence mix (measures, insights, and examples): Evaluating the EU regulatory system <https://www.ewpia.eu/media/636564/evidence-mix_final-9-dec-2021.pdf> (2021). Accessed March 31 2023.
2. 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>
3. 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).
4. 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).
5. US FDA. Framework for FDA's real-world evidence program <<https://www.fda.gov/media/120060/download>> (2018). Accessed March 31 2023.
6. Rivera, D.R. et al. Clinical evidence generation during a pandemic: lessons learned for sustaining progress. *Cancer J.* **28**, 151–156 (2022).
7. 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).
8. 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.

9. HAS. Real-world studies for the assessment of medicinal products and medical devices <https://www.has-sante.fr/upload/docs/application/pdf/2021-06/real-world_studies_for_the_assessment_of_medicinal_products_and_medical_devices.pdf> (2021). Accessed March 31 2023.
10. Health Canada. Health Canada's evolving approach to leveraging real world evidence (RWE) for drug regulatory decisions <<https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html>> (2019). Accessed March 31 2023.
11. 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 (SGB V) <https://www.iqwig.de/download/a19-43_routine-practice-data-for-the-benefit-assessment-of-drugs_rapid-report_v1-0.pdf> (2020). Accessed April 1 2023.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. US FDA. Real-world evidence <<https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>> (2023). Accessed March 31 2023.
18. Gatto, N.M. et al. The structured process to identify fit-for-purpose data: a data feasibility assessment framework. *Clin. Pharmacol. Ther.* **111**, 122–134 (2022).
19. 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).
20. Wang, S.V. et al. Harmonized protocol template to enhance reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: a good practices report of a joint ISPE/ISPOR task force. *Value Health* **25**, 1663–1672 (2022).
21. Beyer, J., Abedtash, H., Hornbuckle, K. & Murray, J.F. A review of stakeholder recommendations for defining fit-for-purpose real-world evidence algorithms. *J. Comp. Eff. Res.* **11**, 499–511 (2022).
22. Burns, L. et al. Real-world evidence for regulatory decision-making: guidance from around the world. *Clin. Ther.* **44**, 420–437 (2022).
23. Jahanshahi, M. et al. The use of external controls in FDA regulatory decision making. *Ther. Innov. Regul. Sci.* **55**, 1019–1035 (2021).
24. Levenson, M.S. Regulatory-grade clinical trial design using real-world data. *Clin. Trials* **17**, 377–382 (2020).
25. Sola-Morales, O. et al. Effectively leveraging RWD for external controls: a systematic literature review of regulatory & HTA decisions. *Clin Pharmacol Ther.* (Accepted Author Manuscript). <https://doi.org/10.1002/cpt.2914>
26. Mahendraratnam, 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).
27. 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).
28. Carrigan, C. et al. External comparator groups derived from real-world data used in support of regulatory decision making: use cases and challenges. *Curr. Epidemiol. Rep.* **9**, 326–337 (2022).
29. Izem, R., Buenconsejo, J., Davi, R., Luan, J.J., Tracy, L. & Gamalo, M. Real-world data as external controls: practical experience from notable marketing applications of new therapies. *Ther. Innov. Regul. Sci.* **56**, 704–716 (2022).
30. 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).
31. 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.
32. EMA. European Medicines Agency guidance for applicants seeking scientific advice and protocol assistance <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/european-medicines-agency-guidance-applicants-seeking-scientific-advice-protocol-assistance_en.pdf> (2022). Accessed April 1 2023.
33. ICMLRA. ICMLRA statement on international collaboration to enable real-world evidence (RWE) for regulatory decision-making <https://icmlra.info/drupal/sites/default/files/2022-07/icmlra_statement_on_rwe.pdf> (2022). Accessed April 1 2023.
34. Concato, J. & Corrigan-Curay, J. Real-world evidence - where are we now? *N. Engl. J. Med.* **386**, 1680–1682 (2022).
35. US FDA. NDA 211723 NME - multi-disciplinary review and evaluation tazemetostat (tazemetostat) <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/211723orig1s000MultidisciplineR.pdf> (2018). Accessed April 1 2023.
36. US FDA. NDA/BLA multi-disciplinary review and evaluation – NDA 212306: Xpovio (selinexor) <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212306orig1s000MultidisciplineR.pdf> (2020). Accessed April 1 2023.
37. 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.
38. European Parliament and the Council of the European Union. Regulation (EU) 2021/2282 of the European parliament and of the council of 15 December 2021 on health technology assessment and amending directive 2011/24/EU <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32021R2282>> (2021). Accessed April 1 2023.
39. Unethita. Joint scientific consultation (JSC) <<https://eur-lex.europa.eu/eli/reg/2021/2282/oj>> (2021). Accessed April 1 2023.
40. US FDA. Guidance for industry formal meetings between the FDA and sponsors or applicants <<https://www.fda.gov/media/72253/download>> (2009). Accessed April 1 2023.
41. 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).
42. NICE. Zydelic: Technology appraisal guidance <<https://www.nice.org.uk/guidance/ta604/resources/idelalisib-for-treating-refractory-follicular-lymphoma-pdf-82608896324293>> (2019). Accessed April 1 2023.
43. 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).
44. 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 2015–2020. *Clin. Cancer Res.* **28**, 27–35 (2022).

45. 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.
46. Dagenais, S., Russo, L., Madsen, A., Webster, J. & Beclen, L. Use of real-world evidence to drive drug development strategy and inform clinical trial design. *Clin. Pharmacol. Ther.* **111**, 77–89 (2022).
47. Oksen, D. et al. Treatment effectiveness in a rare oncology indication: lessons from an external control cohort study. *Clin. Transl. Sci.* **15**, 1990–1998 (2022).
48. EMA. Assessment report. Abecma. International non-proprietary name: Idecabtagene vicleucel <https://www.ema.europa.eu/en/documents/assessment-report/abecma-epar-public-assessment-report_en.pdf> (2021). Accessed April 1 2023.
49. Suissa, S. Immortal time bias in pharmaco-epidemiology. *Am. J. Epidemiol.* **167**, 492–499 (2008).
50. US FDA. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials; draft guidance for industry; availability <<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations>> (2022). Accessed April 1 2023.
51. Singh, H. & Pazdur, R. Importing oncology trials from China: a bridge over troubled waters? *Lancet Oncol.* **23**, 323–325 (2022).
52. 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).
53. 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).
54. Beaulieu-Jones, B.K. et al. Examining the use of real-world evidence in the regulatory process. *Clin. Pharmacol. Ther.* **107**, 843–852 (2020).
55. Korn, E.L. & Freidlin, B. Adaptive clinical trials: advantages and disadvantages of various adaptive design elements. *J. Natl. Cancer Inst.* **109**, djx013 (2017). <https://doi.org/10.1093/jnci/djx013>
56. EMA. Chmp assessment report. Libmeldy. Usual common name: Autologous CD34+ cell encoding ARSA gene <https://www.ema.europa.eu/en/documents/assessment-report/libmeldy-epar-public-assessment-report_en.pdf> (2020). Accessed April 1 2023.
57. Ventz, S., Lai, A., Cloughesy, T.F., Wen, P.Y., Trippa, L. & Alexander, B.M. Design and evaluation of an external control arm using prior clinical trials and real-world data. *Clin. Cancer Res.* **25**, 4993–5001 (2019).
58. Moons, K.G. et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): explanation and elaboration. *Ann. Intern. Med.* **162**, W1–W73 (2015).
59. Rippin, G., Ballarini, N., Sanz, H., Largent, J., Quinten, C. & Pignatti, F. A review of causal inference for external comparator arm studies. *Drug Saf.* **45**, 815–837 (2022).
60. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
61. EMA. Assessment report. Yescarta. International non-proprietary name: Axicabtagene ciloleucel <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.
62. Dong, Y. & Peng, C.Y. Principled missing data methods for researchers. *Springerplus* **2**, 222 (2013).
63. US FDA. NDA/BLA multi-disciplinary review and evaluation (NDA [NME] 212018) balversatm (erdafitinib) <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212018orig1s000_MultidisciplineR.pdf> (2019). Accessed April 1 2023.
64. 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.
65. EMA. Assessment report. Koselugo. International non-proprietary name: Selumetinib <https://www.ema.europa.eu/en/documents/assessment-report/koselugo-epar-public-assessment-report_en.pdf> (2021). Accessed April 1 2023.
66. US FDA. NDA/BLA multi-disciplinary review and evaluation BLA 761115 trodelvy, sacituzumab govitecan-hziy <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/761115orig1s000_MultidisciplineR.pdf> (2018). Accessed April 1 2023.
67. EMA. Assessment report. Enhertu. International non-proprietary name: Trastuzumab deruxtecan <https://www.ema.europa.eu/en/documents/variation-report/enhertu-h-c-005124-ii-0014-epar-assessment-report-variation_en.pdf> (2022). Accessed April 1 2023.
68. US FDA. NDA/BLA multi-disciplinary review and evaluation BLA 761104, lumoxiti, moxetumomab pasudotox. <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/761104orig1s000_MultidisciplineR.pdf> (2018). Accessed April 1 2023.
69. Walker, B. et al. Comparisons of real-world time-to-event endpoints in oncology research. *JCO Clin. Cancer Inform.* **5**, 45–46 (2021).
70. Li, Q. et al. Validation of real-world data-based endpoint measures of cancer treatment outcomes. *AMIA Annu. Symp. Proc.* **2021**, 716–725 (2021).
71. US FDA. BLA multi-disciplinary review and evaluation BLA 761163 monjuvi (tafasitamab-cxix) <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/761163orig1s000MultidisciplineR.pdf> (2020). Accessed April 1 2023.
72. Rubin, D.B. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv. Outcomes Res. Methodol.* **2**, 169–188 (2001).
73. Lim, J. et al. Minimizing patient burden through the use of historical subject-level data in innovative confirmatory clinical trials: review of methods and opportunities. *Ther. Innov. Regul. Sci.* **52**, 546–559 (2018).
74. Spin, P., Siddiqui, M., Hutton, B. & Cameron, C. PNS9 propensity score matching and reweighting to assess comparative efficacy of single-arm trials: a general framework and simulation study. *Value Health* **22**, S289 (2019).
75. Zhang, X., Stamey, J.D. & Mathur, M.B. Assessing the impact of unmeasured confounders for credible and reliable real-world evidence. *Pharmacoepidemiol. Drug Saf.* **29**, 1219–1227 (2020).
76. 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).
77. Jaksa, A., Wu, J., Jonsson, P., Eichler, H.G., Vittitoe, 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).
78. Faria, R., Alava, M.H., Manca, A. & Wailoo, A.J. 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/nice-dsu/tsds/observational-data>> (2015). Accessed April 1 2023.
79. US FDA & EMA. General principles, EMA-FDA parallel scientific advice (human medicinal products) <https://www.fda.gov/media/105211/download?utm_medium=email&utm_source=govdelivery> (2021). Accessed April 1 2023.
80. EMA. Parallel joint scientific consultation with regulators and health technology assessment bodies <<https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/parallel-joint-scientific-consultation-regulators-health-technology-assessment-bodies>> (2022). Accessed April 1 2023.
81. AISBL, E.C.o.P.E. Pharmaboardroom opinion: New EU HTA procedure – will it reduce the complexity and burdens for manufacturers? <<https://www.euope.org/pharmaboardroom-opinion-new-eu-hta-procedure-will-it-reduce-the-complexity-and-burdens-for-manufacturers>> (2022). Accessed April 1 2023.