

# Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance

Journal of **Comparative Effectiveness Research**

Ashley Jaksa<sup>\*,1</sup> , James Wu<sup>2</sup> , Páll Jónsson<sup>3</sup> , Hans-Georg Eichler<sup>4</sup> , Sarah Vititoe<sup>1</sup>  & Nicolle M Gatto<sup>1</sup> 

<sup>1</sup>Scientific Research, Aetion Inc., 5 Penn Plaza, 7th Fl., New York, NY 10001, USA

<sup>2</sup>Amgen Inc, 1 Amgen Center Drive, Thousand Oaks, CA 91320-1779, USA

<sup>3</sup>National Institute for Health & Care Excellence, Level 1A, City Tower, Piccadilly Plaza Manchester, M1 4BT, UK

<sup>4</sup>European Medicines Agency, Domenico Scarlattilaan 6, Amsterdam 1083 HS, Netherlands

\*Author for correspondence: Tel.: +1-844-823-8466; [ashley.jaksa@aetion.com](mailto:ashley.jaksa@aetion.com)

Decision-makers have become increasingly interested in incorporating real-world evidence (RWE) into their decision-making process. Due to concerns regarding the reliability and quality of RWE, stakeholders have issued numerous recommendation documents to assist in setting RWE standards. The fragmented nature of these documents poses a challenge to researchers and decision-makers looking for guidance on what is 'high-quality' RWE and how it can be used in decision-making. We offer researchers and decision-makers a structure to organize the landscape of RWE recommendations and identify consensus and gaps in the current recommendations. To provide researchers with a much needed pathway for generating RWE, we discuss how decision-makers can move from fragmented recommendations to comprehensive guidance.

First draft submitted: 21 October 2020; Accepted for publication: 18 March 2021; Published online: 30 April 2021

**Keywords:** decision-making • health technology assessment • methodology • observational research • real-world evidence • regulatory approval

## Background

Health technology assessment (HTA) agencies and regulators, including the National Institute for Health and Care Excellence (NICE), the US FDA and the European Medicines Agency (EMA) have recently committed to evaluating opportunities to increase the use of real-world evidence (RWE) in their decision-making processes [1–3]. With the Coronavirus disease 2019 (COVID-19) pandemic, regulators and HTA agencies have an increased sense of urgency to use RWE, alongside randomized control trials (RCTs), to evaluate the effectiveness of treatment and vaccines under compressed timelines [4].

Across the healthcare ecosystem, however, there are concerns over wider adoption of RWE in regulatory and reimbursement decision-making. Critics are concerned that researchers will be disincentivized from conducting RCTs and healthcare decision-makers could be forced to rely on 'inferior' evidence [5]. Several high-profile 'disasters,' including recent retractions of a COVID-19 RWE study from major journals [6,7], have solidified the concern that RWE could lead to inaccurate results and poor patient outcomes [8]. Critics also fear that, if allowed to do so, industry will prefer RWE instead of RCTs because RWE is cheaper. Critics thus propose continued adherence to the current paradigm of traditional evidence hierarchies, which display RCTs at the pinnacle and non-randomized studies as inferior [9,10].

The lack of a gold standard in defining and creating decision-quality RWE further contributes to variability in RWE study quality [11], which in turn casts doubt on the validity of RWE and fuels skepticism. We define decision-quality RWE as RWE generated from methods that follow epidemiologic and scientific best practices that enables decision-makers to draw causal conclusions and inform critical decisions. Not all healthcare decisions will require

causal RWE studies; however, a clear roadmap defining scientific thresholds needed to draw causal conclusions from RWD will set the target and help define standards for decisions that require lower evidence bars. Manufacturers and researchers may be reluctant to generate RWE for decision-makers without a clear understanding of the decision-maker's standards for the design, conduct and reporting of RWE, and whether the RWE will have a role in decision-making, even if the standards are met. These concerns are at the forefront of the COVID-19 pandemic. Although the need for RWE standards is irrespective of the pandemic, we think that the pandemic increases the urgency and impetus for creating standards especially with the anticipated reliance on RWE to study vaccine safety and long-term effectiveness. While stakeholders agree that RWE scientific rigor is essential, researchers require a clear path to achieving high-quality RWE that will be accepted by global decision-makers and calm the fears of critics.

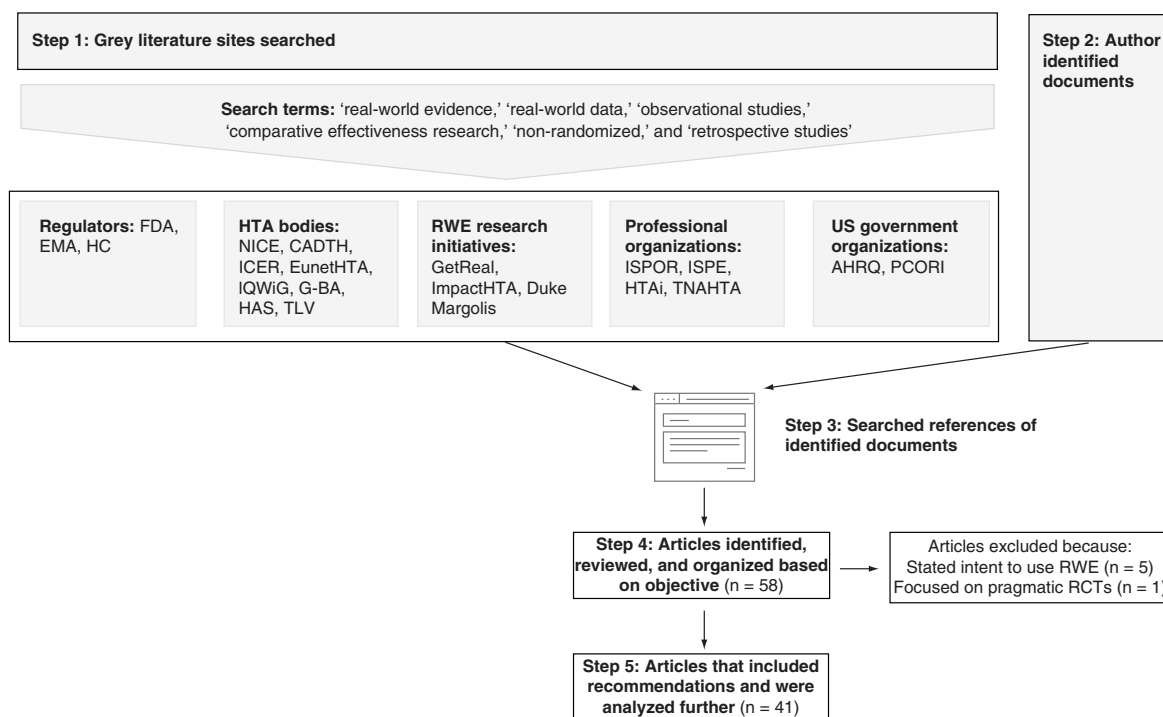
Aiming to improve the quality and reliability of RWE, researchers, professional societies, government agencies and multi-stakeholder initiatives have issued numerous recommendation documents, white papers, peer-review publications and position papers to set standards for generating high-quality RWE (i.e., RWE based on principled, causally interpretable epidemiology) in order to inform regulatory (e.g., approval, indication expansion, confirmation of safety and effectiveness under routine care) and HTA (e.g., reimbursement, performance-based contract) decisions. Some of these recommendations are broad and overlap; some are more narrowly focused. While many of these documents were developed to address specific aspects of RWE generation, it is notable that no stakeholder has released a fully comprehensive document that synthesizes their body of guidance. The fragmented nature of these documents poses a challenge to both researchers looking for guidance on what is required by each decision-maker and a decision-making organization aiming to survey and synthesize the available advice to provide concise recommendations to manufacturers and researchers on generating high-quality RWE. Here, we organize and summarize the current landscape of RWE recommendations for both researchers looking for RWE study design standards to follow and decision-makers evaluating the completeness of their own RWE guidance. We discuss the current gaps in recommendations and outline a way to accelerate the move from fragmented recommendations to comprehensive guidance. Such guidance can provide researchers and manufacturers with a much-needed pathway to generate high-quality RWE, and a clear understanding of how it will be evaluated; however, we do not expect it to be an exact recipe for generating RWE. Rather, the guidance should offer guardrails that focus on key RWE study design elements like concise criteria for determining data quality and fitness-for-purpose, best practices in observational study design that avoid well-known biases, validated and acceptable analytical methods that mitigate confounding, and study reporting and transparency that facilitate trust and reproducibility. We postulate that comprehensive guidance will improve the quality and reliability of RWE and consequently increase trust in the evidence generated.

## Current landscape: fragmented RWE recommendations

### Identifying RWE recommendation documents, white papers, peer-reviewed publications & position papers

Our targeted literature review searched documents from North American and European stakeholders, especially decision-making bodies, that offered recommendations on best practices for RWE generation or use. Typically, decision-making bodies, like the FDA, publish recommendation documents directly to their websites and not through the peer review process. Therefore, we completed a targeted gray literature search of key stakeholder websites as opposed to a systematic review of the peer reviewed literature.

In April 2020, we searched the websites of key regulatory agencies (FDA, EMA, Health Canada), HTA bodies (NICE, Canadian Agency for Drugs and Technologies in Health (CADTH), Institute for Clinical and Economic Review (ICER), European Network for Health Technology Assessment (EUnetHTA), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Gemeinsamer Bundesausschuss (G-BA), Haute Autorité de Santé (HAS), Tandvårds-och läkemedelsförmånsverket (TLV), RWE research initiatives (GetReal, ImpactHTA, Duke-Margolis Health Policy Center), professional organizations (The International Society for Pharmacoeconomics and Outcomes Research [ISPOR]), International Society for Pharmacoepidemiology (ISPE), Health Technology Assessment International (HTAi), International Network of Agencies for Health Technology Assessment (INAHTA) and US government organizations (Agency for Health Research and Quality [AHRQ], Patient Centered Outcomes Research Institute [PCORI]) using the terms 'real-world evidence,' 'real-world data,' 'observational studies,' 'comparative effectiveness research,' 'non-randomized' and 'retrospective studies'. Documents published in English that included recommendations or commentary on how RWE should be generated or used in decision-making were included. In addition, documents and websites that described research projects with the main goal



**Figure 1. Search process for RWE recommendation documents.**

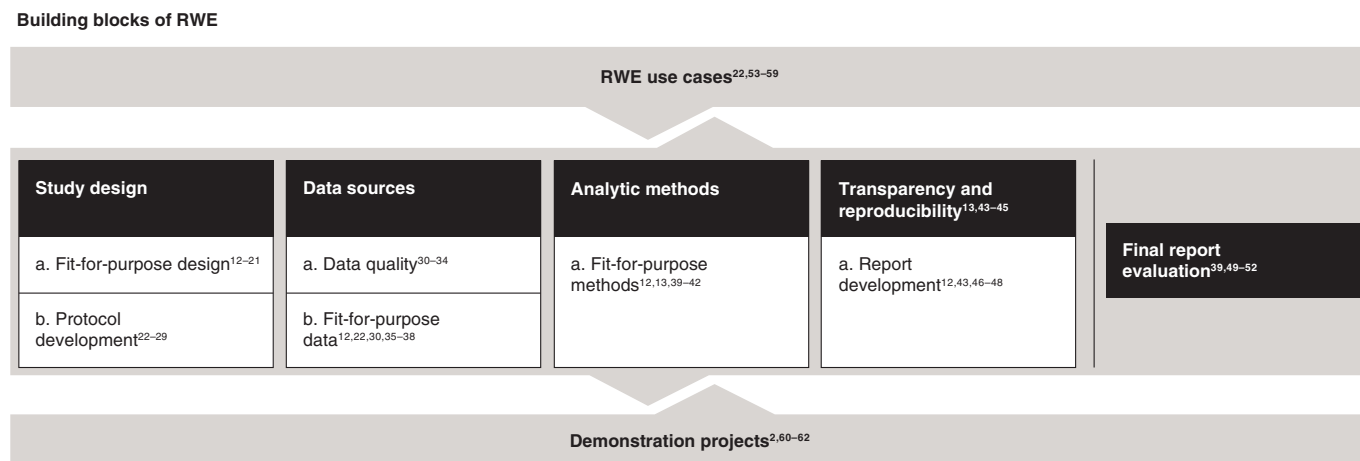
AHRQ: Agency for Health Research and Quality; CADTH: Canadian Agency for Drugs and Technologies in Health; EMA: European Medicines Agency; EunethHTA: European Network for Health Technology Assessment; G-BA: Gemeinsamer Bundesausschuss; HAS: Haute Autorité de Santé; HTAi: Health Technology Assessment International; IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; ICER: Institute for Clinical and Economic Review; INAHTA: International Network of Agencies for Health Technology Assessment; ISPE: International Society for Pharmacoepidemiology; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; NICE: National Institute for Health and Care Excellence; RCT: Randomised control trial; RWE: Real-world evidence; TLV: Tandvårds-och läkemedelsförmånsverket.

of informing RWE standards were included. The references of these documents were mined to grow the list of relevant stakeholders and RWE recommendations. We also identified RWE recommendation documents through the co-authors' expert collective knowledge (Figure 1).

### Developing an organized structure: the building blocks of RWE

One author (AJ) read each document identified through the targeted gray literature search ( $n = 58$ ) to extract the document's objective and identify any RWE recommendations. A document was excluded if it only stated an intent to use RWE or did not address RWD or RWE ( $n = 6$ ). A qualitative analysis of the document objectives was done by two authors (AJ and SV) for the included 52 documents and common themes were identified. The themes resembled the typical scientific process from hypothesis generation through study design and execution. In addition to themes related to the typical scientific process, documents also focused on demonstration projects and how decision-makers should evaluate the quality of RWE studies. We recognize that the processes of developing recommendations can vary widely (e.g., may involve numerous stakeholders or include a public comment period); however, we treated each recommendation document equally when extracting key themes. We arranged the themes into what we consider seven 'building blocks.' Collectively, these 'building blocks' create a systematic arrangement of the RWE recommendations (i.e., an 'organized structure') representing what is needed to build high-quality RWE (Figure 2).

The building blocks center on the theme that high-quality RWE is based on a principled scientific process including: 'fit-for-purpose study design' [12–21] and transparent 'protocol development' [22–29] ('study design' building block), determining 'data quality' (i.e., general quality of RWD sources [30–34]) and identifying 'fit-for-purpose' RWD sources (i.e., data is relevant and reliable for the research question [12,22,30,35–38]; 'data source' building block), selecting appropriate 'analytical methods' [12,13,39–42] ('analytic methods' building block), and enduring



**Figure 2. Building blocks of RWE.**  
RWE: Real-world evidence.

‘transparency and reproducibility,’ [13,43-45] especially in ‘study report development’ [12,43,46-48] (‘transparency and reproducibility’ building block). The principled scientific process concludes with a clear understanding of how decision-makers will evaluate the quality of the RWE study [39,49-52] (‘final report evaluation’ building block).

These core building blocks are supported by both the ‘RWE use cases’ [22,53-59]) and ‘demonstration projects’ building blocks [2,60-62]. The ‘RWE use cases’ building block outlines which type of hypotheses are relevant for RWE studies (e.g., post-marketing safety or indication expansion). When principled scientific processes are followed, the trust in the validity of the studies can expand the number of RWE use cases. Demonstration projects are often multi-organization efforts that attempt to address gaps or uncertainties in one or more of the building blocks (e.g., “what elements are essential in a study report to facilitate replication of the study?”). Once completed, the demonstration project can inform and shape recommendations within building blocks. Each key published or publicly available position piece was sorted, based on the objective and recommendations, into the corresponding building block(s).

### Summarizing recommendations & identifying gaps

Documents that included recommendations (n = 41) were analyzed further. We focused on recommendations in the ‘study design,’ ‘data sources,’ ‘analytic methods,’ ‘transparency and reproducibility’ and ‘final report evaluation’ building blocks. Documents in ‘RWE use cases’ outline potential uses of RWE but did not offer recommendations on when RWE should or should not be used within a use case (e.g., coverage decision). By definition, documents in ‘demonstration projects’ do not offer recommendations, instead the documents describe projects that will help inform and shape recommendations in the future.

Two authors (AJ and SV) independently extracted the RWE recommendations. Discrepancies were discussed and resolved with the help of a third author (NG), if needed. Recommendations were categorized based on the building block that it addressed and a summary of the recommendations were developed (see the [Supplementary document](#)). Some documents included recommendations for multiple building blocks and were included accordingly. For example, Health Canada and CADTH’s ‘Elements of Real World Data/Evidence Quality throughout the Prescription Drug Product Life Cycle’ included recommendations on study protocol development and fit-for-purpose data [22]. Noting agreement between stakeholders, recommendations were summarized by building block (Table 1). To identify areas that may need further development, we then compared the existing recommendations to the building block components we believe are necessary for comprehensive guidance (Table 2). We propose that comprehensive guidance should include three pillars. First a clearly articulated, ideally consensus-based, minimally sufficient justification or explanation of the components of each building block. Second, a step-by-step process that researchers can follow to meet decision-maker expectations. Last, a decision tool or checklist to allow researchers to cross-check the inclusion of all necessary criteria and justify their study design choices to decision-makers.

**Table 1. Real-world evidence recommendations and gaps within building blocks.**

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Study design: fit-for-purpose design	<p>FDA and EMA generally agree upon the following criteria for a persuasive ECA:</p> <ul style="list-style-type: none"> <li>well defined natural history of a disease;</li> <li>external control population is very similar to that of the treatment group;</li> <li>concomitant treatments affecting the primary end point are not substantially different between the external control population and the trial population; and</li> <li>results provide compelling evidence of a change in the established disease progression</li> </ul>	–	Lack of detail on how criteria are measured, and on the minimum necessary criteria for persuasive ECA	[23,24]
Study design: fit-for-purpose design	<p>FOCR recommends the following considerations for the design of ECA studies:</p> <ul style="list-style-type: none"> <li>carefully consider confounding control, including a priori identification of key baseline prognostic factors through DAGs (further imbalances should be mitigated through pre-specified statistical analysis and sensitivity analysis);</li> <li>use objective end points that would be less affected by different measurement techniques, timing, and setting;</li> <li>pre-specify protocols and statistical analyses to improve confidence and transparency; and</li> <li>engage in early discussions with regulators.</li> </ul>	<p>Gatto <i>et al.</i> &amp; ISPOR agree on the need to control for confounding, the usefulness of DAGs, and the need for transparency, but in relation to all RWE study design, not just ECAs</p>	<p>Lack of detail on:</p> <ul style="list-style-type: none"> <li>Preferred methods to control for confounding</li> <li>What good control for confounding looks like (e.g., what level of balance between ECA and clinical trial is sufficient?)</li> <li>What qualifies as an objective end point and how should the objectivity be measured and communicated?</li> <li>What should be communicated to regulators and when? Should HTAs/payers also be included?</li> </ul>	[25,26,28]
Study design: fit-for-purpose design	<p>Hernán <i>et al.</i> recommend emulating a target RCT when designing an RWE study, using strategies such as:</p> <ul style="list-style-type: none"> <li>selecting a new user design;</li> <li>emulating random assignment and ensuring comparability by adjusting for all confounding factors;</li> <li>using outcomes that can be independently validated; and</li> <li>creating a careful definition of time zero</li> </ul>	<p>Gatto <i>et al.</i> and IQWiG include emulating a target RCT as a step in their process for designing RWE studies, although IQWiG's recommendations focus on registry-based RWE studies only</p>	–	[20,28,29]
Study design: fit-for-purpose design	<p>Franklin <i>et al.</i> recommend a framework for determining whether an RWE study is appropriate, and how to design the study to promote validity, which includes:</p> <ul style="list-style-type: none"> <li>data-dependent considerations;</li> <li>avoiding known design and analytic flaws;</li> <li>including robustness checks; and</li> <li>using software developed for RWD analysis.</li> </ul>	<p>Duke-Margolis mostly agrees and recommends justification for why an RCT is not feasible when determining if an RWE is appropriate</p>	<p>While this framework offers high level considerations for determining when an RWE study is appropriate, it lacks a detailed process for evaluating such considerations. For example, Franklin <i>et al.</i> list the ability to measure outcomes and exposure in the database as a key criterion in determining whether to run an RWE study, but do not specify how to evaluate the quality and validity of this measurement</p>	[21,22]
Study design: fit-for-purpose design	<p>Gatto <i>et al.</i> propose a step-by-step process to design and justify RWE study design choices in a way that ensures validity and establishes trust. Steps include:</p> <ul style="list-style-type: none"> <li>conceptualizing a target RCT and determining what elements can be more pragmatic when articulating a research question and operationalizing design elements;</li> <li>drawing a causal diagram and describing capture of potential confounders;</li> <li>specifying minimum criteria that must be met to validly capture each design element; and</li> <li>using a flow chart to determine if an RWE study design is appropriate.</li> </ul>	<p>Hernán <i>et al.</i> developed the recommendation to emulate a target RCT</p> <p>ISPOR agrees with the usefulness of causal diagrams to identify all potential confounders</p>	<p>Lacking a process and decision guide on how to vet and select the most appropriate data source for the study</p>	[26,28,29]

DAG: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FOCR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

**Table 1. Real-world evidence recommendations and gaps within building blocks (cont.).**

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Study design: fit-for-purpose design	<p>ISPOR recommends that <b>to avoid classification bias</b>, researchers should:</p> <ul style="list-style-type: none"> <li>• identify potential sources of misclassification and how they might impact results;</li> <li>• choose an exposure time-window based on when the medication may cause the outcome vs actual drug intake; and</li> <li>• use measures validated with external sources or a population most similar to the study, or at least perform sensitivity analysis on various definitions of the results if validation is not possible</li> </ul>	<p><b>Duke-Margolis</b> agrees with using validated measures, and also recommends considering a new user design to minimize classification bias</p>	<p>Lacking specifics on how to operationalize these considerations and the minimum criteria necessary for a valid study. For example, the recommendation reads: “when measuring comorbidity, select a measure that has been validated in a population most similar to the study and for the outcome under investigation.” The recommendations do not specify what researchers should do if there is a validated measure in a dissimilar population or if the measure hasn't been validated</p>	[21,26]
Study design: fit-for-purpose design	<p>ISPOR, Duke-Margolis, and Gatto <i>et al.</i> recommend that, to understand the potential impact of <b>confounding</b>, researchers should:</p> <ul style="list-style-type: none"> <li>• define a DAG <i>a priori</i></li> <li>• perform sensitivity analysis for different assumptions regarding the confounding structure</li> </ul>	<p><b>IQWiG</b> states that researchers should prespecify confounders and use adjustment to establish similarity between groups</p>	<p>Does not describe how to determine what variables should be included in pre-specified sensitivity analysis vs post-hoc sensitivity analysis</p>	[20,21,26,28]
Study design: fit-for-purpose design	<p>Franklin <i>et al.</i> offer timing and measurement <b>recommendations for a single, continuous-valued confounder</b> that varies over time, noting that:</p> <ul style="list-style-type: none"> <li>• Researchers should carefully consider the timing of confounder measurement relative to exposure initiation and the expected rate of change in the confounder before and after exposure</li> <li>• To minimize measurement error and bias, researchers should assess confounders in close temporal proximity to the start of treatment. Ideally, these measurements should precede exposure initiation; however, measurements taken shortly after exposure initiation may also be useful for confounder adjustment</li> <li>• If the timing of exposure initiation relative to confounder measurements is unknown, all potential approaches can yield badly biased estimates of treatment effect</li> <li>• Restricting the sample to patients who initiate exposure shortly after a confounder measurement is a safe approach, but investigators should be aware of how this restriction impacts the target treatment effect parameter</li> </ul>	<p>–</p>	<p>Very focused discussion on measurement and study design choices for single-continuous valued confounders</p>	[27]
Study design: protocol development	<p>Stakeholders have published recommendations on <b>protocol development</b>, ranging from high-level tools and lists to in-depth considerations for observational study design and protocol development. Overall, there is high level agreement between recommendations, though some discrepancies exist</p>	<p><b>EMA PASS template, ENCePP checklist, and CADTH/Health Canada</b> provide high-level tools and lists that guide what should be included in a study protocol</p> <p><b>AHRQ and PCORI</b> provide in-depth considerations for observational study design and protocol development</p>	<p>With multiple tools, it is unclear which tool is most appropriate and preferred by decision-makers</p>	[12,31,34–36]

DAG: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FOCR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

Table 1. Real-world evidence recommendations and gaps within building blocks (cont.).

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Study design: protocol development	<p>ISPOR-ISPE recommends seven good <b>procedural practices for hypothesis testing studies</b>, the majority of which apply to protocol development and general study planning. Practices include:</p> <ul style="list-style-type: none"> <li>• determining an <i>a priori</i> hypothesis;</li> <li>• registering protocol;</li> <li>• publishing study results and noting any deviations from protocol and analysis plans;</li> <li>• enabling opportunities to replicate studies;</li> <li>• performing study on different data sources to confirm findings;</li> <li>• publicly addressing methodological criticisms; and</li> <li>• including key stakeholders in the designing, conducting of, and disseminating of studies</li> </ul>			[33]
Data source: fit-for-purpose	<p>Duke-Margolis states that <b>elements of a fit-for-purpose dataset</b> include:</p> <ul style="list-style-type: none"> <li>• data relevancy, including the availability of key data elements (i.e., exposure, outcomes and covariates), ability for patient-level linking, representativeness, sufficient subjects, and longitudinally; and</li> <li>• data reliability, including accuracy, validity, conformance, plausibility, completeness, data provenance, and transparency in data processing</li> </ul>	<p>FDA generally agrees, but its recommendation omits the following components: sufficient subjects, longitudinally, and validity and plausibility of key data elements.</p> <p>EMA generally agrees, but its recommendation omits covariate availability, conformance and plausibility</p> <p>CADTH/Health Canada generally agrees, but its recommendation omits conformance and plausibility</p> <p>Hall <i>et al.</i> generally agrees but lacks discussion on data provenance</p> <p>IQWiG generally agrees, but its recommendations don't include the need for sufficient subjects, and, while implied, don't specifically require accuracy of outcomes and exposure</p> <p>IQWiG also omits conformance and plausibility</p>	<p>Recommendations are high-level and lack criteria on how to operationalize and meeting each element and minimum criteria</p>	[12,20,37,42,44,45]
Data source: fit-for-purpose	<p>Duke-Margolis recommends <b>operationalizing the evaluation of data reliability</b> with a minimum set of verification checks to assess data reliability (i.e., conformance, completeness and plausibility), including:</p> <ul style="list-style-type: none"> <li>• verification checks for data conformance to assess the structure of the data and how compliant the data are with internal relational, formatting or computational definitions or standards;</li> <li>• verification checks for data completeness to assess the presence values within data elements;</li> <li>• verification checks for data plausibility to assess the range of values within a variable, and whether two or more variables have an expected context-dependent relationship;</li> <li>• verification checks for data plausibility to assess whether time-related and time-varying variables change as expected; and</li> <li>• adequate data documentation to preserve end users' ability to check their mapping and transformations, both of which should be prespecified and justified.</li> </ul>			
Data source: fit-for-purpose			<p>Lacks thresholds for assessments of data conformance, completeness or plausibility</p>	[43]

DAG: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FOQR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

**Table 1. Real-world evidence recommendations and gaps within building blocks (cont.).**

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Data source: data quality	FDA states the following as necessary components of EHR data in order to be used in clinical investigations (not specific to RWE studies): <ul style="list-style-type: none"> <li>• data quality;</li> <li>• data integrity;</li> <li>• data provenance; and</li> <li>• transparency in data collection and data modifications</li> </ul>	MHRA agrees with these recommendations, but note they are minimum requirements for compliance of all data, not just EHR data	FDA lists general considerations but does not provide thorough detail on how to evaluate each	[38,39]
Data source: data quality	EUNetHTA recommends the REQuest tool for international organizations (HTA and regulatory) considering whether to use registry data in evidence development. REQuest evaluates methodological information, essential registry standards, and additional requirements (a collection of 23 questions)	–	This document collects detailed information about registry data only, and REQuest does not offer recommendations on how to evaluate the sufficiency of each component to meet the needs of regulators or HTA agencies	[41]
Data source: data quality	According to CanREValue, an RWD set to be used for oncology RWE must include: <ul style="list-style-type: none"> <li>• baseline demographics (e.g., age, sex);</li> <li>• clinical variables (e.g. performance status, prior treatments);</li> <li>• subsequent treatments (e.g., chemotherapy, radiation);</li> <li>• for research questions on outcomes, information such as overall survival or other time-to-event end points;</li> <li>• for research questions on safety, hospitalization and/or ED visits;</li> <li>• for research questions on cost-effectiveness, cost data;</li> <li>• for research questions on budget impact outcomes, cost of drug, update and characteristics of use; and</li> <li>• for research questions QoL, patient-reported outcomes</li> </ul>	–	CanREValue offers oncology-specific recommendations but does not discuss how these criteria might change based on oncologic condition. It also excludes a discussion of needs for data outside oncology, and where there may or may not be overlap with other therapeutic areas	[40]
Data source: data quality	EMA recommends an RWD set completeness scoring algorithm, designed to enable filtering of databases based on a number of key characteristics. The algorithm scores each dataset, but not specifically in terms of the research questions of interest. In general, datasets with the highest scores were more comprehensive and likely to be of interest for regulatory decision-making. <ul style="list-style-type: none"> <li>• Score 1 is computed based on data source characteristics, namely collaboration, longitudinal data, recording of exposure and recording of clinical events</li> <li>• Score 2 refers to data source elements such as size of the data source, access to and analysis of data, linkage potential, presence of hospital data and pediatric data, patient characteristics and disease characteristics included, validation studies, and the potential transformation of the data to a common data model</li> </ul>	–	EMA does not offer a cut off for what is appropriate for regulatory decision-making	[37]
Fit-for-purpose analytic methods	NICE recommends using an algorithm for deciding what analytical methods to choose for RWE studies. The Decision Support Unit (on behalf of NICE) created an algorithm of a number of sequential steps to help choose the appropriate method to estimate treatment effect from comparative individual patient data. Since it is nearly impossible to know which method is best, results obtained from alternative methods that vary plausible assumptions should be presented as sensitivity analyses. Although the choice of method should be driven by the treatment effect of interest, the availability of data and the mechanism by which people were assigned to treatment or otherwise will play a central role in what can and cannot be identified	–	Omits high-dimensional propensity scores as an analytic option	[46]

DAG: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FOCR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

Table 1. Real-world evidence recommendations and gaps within building blocks (cont.).

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Fit-for-purpose analytic methods	Schneeweis <i>et al.</i> recommend high-dimensional propensity score matching as an analytical option to improve effect estimates. In typical pharmacoepidemiologic studies, high-dimensional propensity score can result in improved effect estimates compared with adjustment limited to predefined covariates, when benchmarked against results expected from randomized trials	–	Lacking criteria to determine when high-dimensional propensity score matching is most appropriate, and best practices for the method	[49]
Fit-for-purpose analytic methods	<p>ISPOR reviews the following methods to control for confounding:</p> <ul style="list-style-type: none"> <li>• stratification;</li> <li>• regression;</li> <li>• propensity score analysis;</li> <li>• marginal structural models;</li> <li>• instrumental variable (IV) analysis;</li> <li>• structural equation modeling; and</li> <li>• sensitivity analysis related to residual confounding</li> </ul>	IQWiG agrees that regression, propensity score analysis, and IV analysis are suitable ways to control for confounding. Duke-Margolis agrees with ISPOR's recommendations	Missing high-dimensional propensity scores, and only offers a cursory discussion of sensitivity analysis Does not provide detail on best practices for executing these analyses	[20,21,47]
Fit-for-purpose analytic methods	Schneeweis recommends using sensitivity analyses to improve understanding of the effects of residual confounding in pharmacoepidemiologic studies. Four basic approaches can be applied: <ul style="list-style-type: none"> <li>• sensitivity analyses based on an array of informed assumptions;</li> <li>• analyses to identify the strength of residual confounding that would be necessary to explain an observed drug-outcome association;</li> <li>• external adjustment of a drug-outcome association given additional information on single binary confounders from survey data using algebraic solutions; and</li> <li>• external adjustment considering the joint distribution of multiple confounders of any distribution from external sources of information, using propensity score calibration</li> </ul>	ISPOR recommends using sensitivity analyses to control for confounding, and Schneeweis expands on this recommendation and details a number of different approaches. Duke-Margolis adopts Schneeweis's recommendations and agrees on the use of sensitivity analysis to evaluate confounding	Missing decision criteria for which approach to choose and how to evaluate the validity of the approach	[21,47,48]
Transparency and reproducibility: transparency	<p>ISPOR/ICPE recommends the following to ensure transparency:</p> <ul style="list-style-type: none"> <li>• registering study protocols and posting protocols before study analysis begins;</li> <li>• posting results tables;</li> <li>• including date-stamps on protocols and all revisions; and</li> <li>• using structured report templates</li> </ul>	Duke-Margolis agrees with pre-registering study protocols and statistical analysis plans to mitigate information bias	Details goals for comprehensive study registries and directs researchers to post protocols online before study initiation but does not indicate where researchers should post protocols in the interim while more comprehensive RWE protocol registries are developed	[21,52]
Transparency and reproducibility: Transparency	<p>ISPOR/ICPE outlines a roadmap for how observational researchers can improve transparency over time, including:</p> <ul style="list-style-type: none"> <li>• establishing a study registry;</li> <li>• defining what information registries should include and</li> <li>• determining how incentive structures can be created to encourage researchers to post study protocols</li> </ul>	–	Outlines future directions for transparency in observational research but provides little guidance on how a present-day researcher can improve transparency	[52]

DAG: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FOCR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

**Table 1. Real-world evidence recommendations and gaps within building blocks (cont.).**

Building block	Recommendation	Authors that agree	Gaps in recommendations	Ref.
Transparency and reproducibility: transparency	<p>The RECORD-PE expands upon the RECORD and STROBE recommendations to provide a detailed checklist for writing a transparent study report on observational research. Key aspects include:</p> <ul style="list-style-type: none"> <li>• detailed information on the data source, data cut, linked supplementary data, and data cleaning step;</li> <li>• detailed methods sections with descriptions of statistical methods, covariate analyses, subgroup analyses and sensitivity analyses;</li> <li>• detailed inclusion and exclusion criteria, codes, algorithms and enrollment definitions;</li> <li>• detailed exposure, outcome and covariate assessments including codes, algorithms, windows of assessment and validation;</li> <li>• results of study population selection including a patient flow diagram;</li> <li>• detailed results section including baseline characteristics, follow-up time and key results;</li> <li>• a discussion of limitations, potential biases, interpretations and implications;</li> <li>• disclosure of funding; and</li> <li>• where supplemental materials may be found.</li> </ul>	<p>The RECORD-PE agrees fully with the RECORD and STROBE recommendations but adds additional recommendations. Recommendations from Wang <i>et al.</i> overlap with some aspects of the RECORD-PE checklist detailing requirements for the methods section. Since the RECORD-PE focuses primarily on achieving transparency, rather than reproducibility, Wang <i>et al.</i> list several minimum necessary criteria needed to replicate a study that are not detailed in the RECORD-PE</p>	<p>Regulators and HTA organizations have not commented on whether these study method details are recommended/required for decision-making</p>	[51,53–55]
Transparency and reproducibility	<p>Wang <i>et al.</i> lay out a framework for understanding the minimum necessary information needed in a study report methods section sufficient for replication of a study and study results by an independent researcher in healthcare database studies. Key components in the methods section include:</p> <ul style="list-style-type: none"> <li>• detailed information on the data source, data cut, linked supplementary data and data cleaning steps;</li> <li>• a study design diagram featuring key temporal anchors;</li> <li>• a detailed inclusion/exclusion criteria section including codes, algorithms, windows of assessment and enrollment definitions;</li> <li>• a detailed exposure, outcome, and covariate assessment, including codes, algorithms and windows of assessment;</li> <li>• sampling strategies, matching ratio, and matching criteria; and</li> <li>• details on software and versions used for implementation of analyses</li> </ul>	<p>Schneeweiss <i>et al.</i> expands upon Wang <i>et al.</i> to detail the importance of the study design diagram for allowing readers to quickly digest and understand key time windows Components of a study diagram should include:</p> <ul style="list-style-type: none"> <li>• source data range and study data range in calendar time;</li> <li>• detailed first order temporal anchors in patient time; and</li> <li>• detailed second order temporal anchors as they relate to first order anchors.</li> </ul> <p>RECORD-PE provides a checklist that includes some, but not all, of the components in this framework</p>	<p>While comprehensive, regulators and HTA organizations have not commented on whether these study method details are recommended/required for decision-making</p>	[50,51,54]
Final report evaluations	<p>There are a number of tools to assist decision makers in evaluating the quality of RWE studies, however there is no consensus on a gold standard tool cited. We collected and grouped the major categories from all tools, including:</p> <ul style="list-style-type: none"> <li>• a priori considerations;</li> <li>• data sources;</li> <li>• study design;</li> <li>• results reporting; and</li> <li>• interpretation of findings.</li> </ul> <p>Only two tools (ISPOR's "A Checklist for Retrospective Database Studies" and ISPOR-AMCP-NPC's "A Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform HealthCare Decisions") included at least one question from each category</p>	<p>–</p>	<p>Only NICE has a tool, regulators and other HTAs have not commented on what they are looking for in RWE study quality. No gold standard tool endorsed by HTAs or regulators</p>	[46,61,62]

DA.G: Directed acyclic graph; ECA: External control arm; ED: Emergency department; EHR: Electronic health record; FO.CR: Friends of cancer research; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data; RWE: Real-world evidence; QoL: Quality of life;

**Table 2. Comparison of components of comprehensive guidance with current real-world evidence recommendations.**

Building block	Proposed components of comprehensive guidance	Details of proposed components	Do existing recommendations provide sufficient guidance?	Ref.
Study design: fit-for-purpose design and protocol development	Minimally sufficient justification/explanation	A high-level, but clearly articulated, ideally consensus-based framework for determining when RWE studies, including ECAs, are likely to produce valid evidence for decision-making	Partial; EMA and FDA have issued guidance on high-level criteria for a persuasive ECA, but other study design considerations are missing guidance	[23,24]
		Criteria for determining situations in which RWE studies are necessary, justified and can provide valid evidence for decision-making if principled database epidemiology is applied (i.e., assessing the appropriateness of RWE). For example, should RWE only be used when RCTs are not ethical? Can ECAs be used to increase power in RCTs or should ECAs only be used with single-arm studies?	No; RCT DUPLICATE is exploring the validity of RWE in specific use cases, but guidance on acceptable situations for RWE is necessary from decision-makers	[56]
	Step-by-step process	Step-by-step process for designing RWE studies with a focus on feasibility and setting and meeting minimum criteria necessary for a valid study	Yes	
		Step-by-step process for making study design choices to mitigate major sources of bias in RWE studies (e.g., confounding, selection bias). This process should note the pros and cons of each approach, outline steps to decide which approach is best, and define minimum criteria as necessary	No; A number of groups have outlined basic considerations for components of study design (e.g., how to mitigate confounding), but details on how to select the correct methods, and apply and validate the methods used, are absent	
	Tool/checklist to aid researchers	To facilitate transparency, a tool/checklist to assist researchers through the process of selecting the best RWE study design and record/justify all decisions, including how each choice met minimum criteria necessary for a valid study	Partial; Gatto <i>et al.</i> provides a template to document study design decisions, but it is unclear if this template is accepted by decision-makers	[28]
		Tool/checklist to document all final study design decisions and outline the statistical analysis plan in a formal study protocol	Partial; multiple tools exist and there is high-level agreement between recommendations, however, it is unclear which tool is most appropriate and preferred by decision-makers	
Data source: Data quality	Minimally sufficient justification/explanation	Clearly articulated and ideally consensus-base definition of a quality RWD source and the components necessary to determine quality. Do components change based on the therapeutic area? How should patient-relevant and patient-reported outcomes be incorporated into RWD sources? Are there recommendations to data stewards for uniformity in RWD capture?	No	
		A list of appropriate datasets that meet quality criteria and that could be used for regulatory and HTA questions per country or region	Partial; EMA has evaluated European RWD sources for regulatory decision-making while CanREValue has evaluated Canadian cancer RWD sources for HTA decision-making	[37,40]
		Minimum criteria necessary for a real-world dataset to be determined high-quality.	No	
	Step-by-step process	A step-by-step process to evaluate overall RWD quality	No	
	Tool/checklist to aid researchers	Tool/checklist to communicate and facilitate transparency in the evaluation	Partial; REQuEST tool offers considerations for registries but does not offer guidance on how to determine if the considerations are met. The EMA has created an algorithm to evaluate the completeness of datasets, but this is only one component of quality and the EMA does not offer score thresholds for what is considered complete and relevant for decision-making	[37,41]

ECA: External control arm; EHR: Electronic health record; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data;

**Table 2. Comparison of components of comprehensive guidance with current real-world evidence recommendations (cont.).**

Building block	Proposed components of comprehensive guidance	Details of proposed components	Do existing recommendations provide sufficient guidance?	Ref.
Data source: Fit-for-purpose data	Minimally sufficient justification/explanation	Clearly articulated and ideally consensus-base definition of a fit-for-purpose RWD set that allows the researcher to make inferences in regard to the research question	Partial; While there is slight disagreement between recommendations, Duke-Margolis's definition is most inclusive. However, it is unclear if global decision-makers agree and adopt the Duke-Margolis recommendations	[42]
		Minimum criteria necessary for a dataset to be determined fit-for-purpose. Including: <ul style="list-style-type: none"> <li>• Criteria to meet data relevancy and data reliability components specified in the study design</li> <li>• Criteria for ensuring the generalizability of the RWD to current clinical practice</li> <li>• Criteria for the applicability of international RWD for decision-making. When is using international data acceptable vs not, and what are the best practices for using international datasets?</li> </ul>	No	
	Step-by-step process	Step-by-step process on how to evaluate data relevancy and data reliability (major components of the Duke-Margolis definition)	Partial; Duke-Margolis recommends a minimum set of verification checks but does not provide thresholds to determine if the check was met	[43]
	Tool/checklist to aid researchers	Tool/checklist for researchers to document the process of justifying and selecting a fit-for-purpose dataset	No	
Fit-for-purpose analytic methods	Minimally sufficient justification/explanation	Clear description and potentially examples of what statistical methods could be appropriate to minimize the risk of bias for different types of research questions and study designs	Partial; NICE developed an algorithm to help select methods, but it is unclear if other decision makers agree with the recommendations	[46]
		Recommendations and thresholds for determining when controlling for confounding has been sufficient (e.g., what level of balance is needed between treatment groups?)	No	
	Step-by-step process	An outline of the pros and cons of each method to control for confounding (including residual confounding) and a step-by-step guide on how to select and apply the appropriate methods	Partial; Multiple authors have outlined pros and cons of methods	
	Tool/checklist to aid researchers	Tool/checklist for researchers to document the process of justifying and selecting analytic method(s)	No	
Transparency & Reproducibility	Minimally sufficient justification/explanation	Decision-makers set clear expectations for researchers on what defines transparency in RWE studies including: <ul style="list-style-type: none"> <li>• best practices in engaging with decision makers and other stakeholders in the planning and execution of RWE studies, including the timing of the engagement</li> <li>• best practices in allowing decision makers to have access to patient-level data to pressure test assumptions and validity of results</li> <li>• best practices in protocol pre-registration</li> </ul>	No	
		Minimum criteria necessary in the study report to facilitate complete understanding of how results were achieved and if necessary, to replicate the study	Partial; Wang <i>et al.</i> lay out a framework for understanding the minimum necessary information needed in a study report methods section sufficient for replication of a study	[51]
	Step-by-step process	Step-by-step process for facilitating transparency from study design through final report evaluation including a study protocol registration process	No; ISPOR Transparency Initiative is currently working toward a culture of transparency starting with protocol registration, but it has not yet issued process recommendations yet	[52]
	Tool/checklist to aid researchers	Tool/checklist for researchers to ensure they are following the step-by-step process and have met all study transparency requirements	Partial; Wang <i>et al.</i> lay out a framework of the minimum necessary information needed for study replication. While comprehensive, it only focuses on the study report and regulators and HTA organizations have not commented on whether these study method details are recommended/ required for their decision-making	[51]

ECA: External control arm; EHR: Electronic health record; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data;

Table 2. Comparison of components of comprehensive guidance with current real-world evidence recommendations (cont.).

Building block	Proposed components of comprehensive guidance	Details of proposed components	Do existing recommendations provide sufficient guidance?	Ref.
Final report evaluation	Minimally sufficient justification/explanation	Clearly articulated and ideally consensus-base definition of a high-quality RWE study	No; Multiple tools exist to evaluate quality, thus defining it by proxy, but there is a lack of agreement between tools on what components are necessary	
		Minimum criteria for a high-quality RWE study	No	
	Step-by-step process	Step-by-step process for how decision-makers will evaluate the components of the study and determine overall quality	No	
	Tool/checklist to aid decision-makers	Tool/checklist for decision-makers to communicate study quality to external stakeholders	No; Multiple tools exist for decision-makers to evaluate quality, but no gold standard tool is endorsed by HTAs or regulators	
ECA: External control arm; EHR: Electronic health record; HTA: Health technology assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; RCT: Randomized control trial; RWD: Real-world data;				

Here, we provide two examples – within separate building blocks – to illustrate the fragmented nature of the current guidance documents, the gaps in existing recommendations documents and development is needed within the building block in order to progress to comprehensive guidance.

*Example 1: A spotlight on fit-for-purpose design ('study design, fit-for-purpose design' building block) & use of external control arms*

Single-arm trials have become more prevalent in regulatory and HTA submissions [63], especially for conditions with high unmet need in which RCTs are not feasible. There is pressure on regulatory and HTA bodies to determine how to best use single-arm trial evidence, which is sometimes supplemented with external control arms (ECAs) from real-world data (RWD), in their decision-making. To date, three stakeholders have been active in developing recommendations for the use of ECAs, which is one type of RWE study design, (Figure 2, 'study design' building block); two regulatory bodies, the FDA and EMA and one multi-stakeholder collaboration, the US's Friends of Cancer Research (FOCR).

Embedded in the FDA and EMA published recommendations on choosing an appropriate control group for any study, not specifically RWE studies, both agencies describe the conditions necessary for a persuasive ECA (Table 1); both agree that ECAs may be useful when there exists a well-defined natural history, objective end point(s), patient group comparability, sufficient covariate measurement and a large effect size [15,16]. The ECA study supporting blinatumomab (Blinicyto) for treatment of adults with Philadelphia chromosome negative relapsed or refractory B-precursor acute lymphoblastic leukemia met these criteria and the FDA and EMA included the ECA as part of the body of evidence used in their approvals [64,65].

Building upon the FDA and EMA's recommendations, FOCR convened a multi-stakeholder consortium to address potential for bias in the control arm, a key challenge in developing valid ECA studies, and subsequently published a white paper identifying potential biases that may arise with ECAs and outlining study design methods to mitigate these biases (Table 1 [17]). One of these methods, the use of a directed acyclic diagram (DAG) to identify key baseline prognostic factors that should be included as confounders in the design and analysis of the ECA, was echoed by both ISPOR [18] and Gatto *et al.* [20] in their recommendations on designing RWE studies (not just ECAs) to understand the potential effects of confounding (Table 1). While the FDA, EMA and FOCR recommendations are a starting point for ECA study design, more direction from decision-makers is needed (Table 1). The FDA and EMA could provide examples of when the persuasive criteria are and are not met. For example, what are examples of when the natural history of a disease is well defined versus not? What are the minimum criteria necessary for a persuasive control arm? Similarly, the FOCR recommendations could go further and discuss what 'good' control for confounding looks like (e.g., what level of balance between the control arm and single-arm study is necessary).

As ECAs are only one type of RWE study design, it is important to understand how the totality of recommendations within the fit-for-purpose study design building block match what we might expect for comprehensive guidance (Table 2). One gap is the lack of guidance on how to select the most appropriate study design components

to limit bias and how to confirm the method chosen was appropriate. A number of stakeholders (e.g., ISPOR [18], FOCR [17], Franklin *et al.* [19]) outlined basic considerations to avoid types of bias that impact RWE studies (e.g., classification, selection, confounding), but the recommendations stop short of operationalizing a process for selecting an approach and validating that approach. While we do not expect an exact recipe for what study design elements to select for every situation, researchers could benefit from more guidance on what elements might be appropriate for certain situations (e.g., what makes a persuasive ECA) and how the choices made by the researcher will be evaluated. As regulatory and HTA decision-makers are developing their RWE recommendations regarding study design considerations, they should understand the current landscape of recommendations and consider closing current gaps by providing more detailed guidance on RWE study design.

*Example 2: A spotlight on fit-for-purpose data ('data sources' building block) & matching of data to the research question*

The 'data sources, fit-for-purpose' data building block (Figure 2) has contributions from a diverse set of stakeholders including regulators (FDA, EMA, Health Canada), HTAs (CADTH, IQWiG), RWE Initiatives (Duke-Margolis) and independent researchers. Duke-Margolis and their multi-stakeholder consortium crafted recommendations in 'Characterizing RWD Quality and Relevancy for Regulatory Purposes,' which focused on describing elements of 'fit-for-purpose' data such as data relevancy and data reliability [35]. We cross-referenced five guidance documents against the more comprehensive Duke-Margolis' recommendations: EMA and Heads of Medicines Agency Big Data Task Force subgroup report 'Observational Data (Real World Data)' [30], FDA's 'Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices' [371] joint recommendation from the Canadian Agency for Drugs and Technologies in Health (CADTH) and Health Canada, 'Elements of Real World Data/Evidence Quality throughout the Prescription Drug Product Life Cycle' [22], 'Guidelines for Good Data Selection and use of Pharmacoepidemiology Research' [38], and IQWiG's 'Routine Practice Data for the Benefit Assessment of Drugs [12].' We assessed concordance on individual elements of data relevancy and reliability that were deemed necessary for data fitness-for-purpose, for example, representativeness, availability of key data elements, data provenance and accuracy.

While we observed fairly good agreement across these six recommendation papers, some small differences emerged (Table 1). For example, unlike the other regulatory and HTA bodies, EMA did not discuss the availability of covariates as a criterion in assessing data's fitness-for-purpose. This discrepancy was likely due to the motivation of the EMA's publication, which was to catalog RWD sources within the European Union. By evaluating each RWD source against a predetermined list of essential characteristics (e.g., longitudinally, transparency in data processing), the EMA by proxy identified criteria for fit-for-purpose data. The omission of covariate availability as an important aspect of fit-for-purpose data assessment could have been deliberate or simply not relevant to the goal of the EMA's publication. Of note, EMA has not issued specific formalized recommendations on their definition for fit-for-purpose RWD and thus it leaves researchers and manufacturers to interpret EMA's position through other publications; this type of extrapolation regarding a decision-makers position on criteria for high-quality RWE was common in our analysis.

Multiple stakeholders have offered recommendations on criteria for defining data fitness-for-purpose. Subtle differences in these recommendations could reflect differences in what is considered important or that recommendations were developed in silos without consulting previously published pieces or other stakeholders on the topic. For example, CADTH and Health Canada did not reference Duke-Margolis's white paper or Hall *et al.* on defining fitness-for-purpose; the EMA's subgroup report did not reference Duke, FDA, CADTH or Health Canada's recommendations on fitness-for-purpose, but did reference Hall *et al.* These stakeholders may be focused on building upon the work of their regional colleagues (e.g., Duke-Margolis building on FDA's work on fitness-for-purpose and not Canada or Europe's work) or are focused on in country recommendations due to logistical challenges with aligning with multinational stakeholders, however, with the global drug development community and stakeholder commitment to RWE, this is a missed opportunity to consolidate effort and guide the development of global RWE recommendation. The proposed organized structure could have helped stakeholders identify previous recommendations, likely published in the gray literature, in order to build upon the recommendations and avoid duplicative work. Adopting similar existing recommendations could have allowed the stakeholder to move on to focus on less defined areas of fit-for-purpose data and make further progress toward comprehensive guidance.

While there is high level agreement on defining what is fit-for-purpose data, gaps in other aspects of fitness-for-purpose exist (Table 2). For example, there are no operational definitions for researchers to determine if the RWD source meets each data relevance or data reliability criteria. Duke-Margolis does offer recommendations on the verifications checks for data reliability (e.g., blood pressure recordings in the dataset that are biologically

plausible), however, it does not offer thresholds for passing the checks [36]. Another consideration that is relevant in a global community is the acceptability of international RWD in decision-making. When is data from outside the country relevant for in country decision-making? What are best practices for using international data? Should healthcare systems be similar? How is similarity measured? Researchers would benefit from guidance in these areas and decision-makers should consider evaluating these topics for inclusion in their future RWE guidelines.

### Looking ahead: transforming fragmented recommendations to comprehensive guidance

The completeness of recommendations within each building block differs, but each building block has gaps that need to be closed in order to progress toward comprehensive guidance. We encourage stakeholders to use this analysis to both close gaps within their current recommendations and prioritize future recommendation development work. While the end goal is to eventually develop comprehensive RWE guidance, a more realistic starting point could be to focus first on complete guidance within each building block.

### Building blocks to prioritize for further development or consensus

To begin, we should strategically prioritize future guidance development to focus on addressing the major concerns of critics, like transparency and data quality. Transparency and reproducibility (Figure 2) have been identified by many stakeholders as ways to enhance the credibility of RWE by allowing others to understand and replicate methods from RWD studies and remove any appearance of being a ‘black box.’ Great progress has been made by the REPEAT Initiative [61] and Wang *et al.*’s publication ‘Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies,’ [44] which showcased what elements are necessary to include in a study report to facilitate RWE study reproducibility. However, further work is needed to reduce the possibility that a research may have ‘cherry picked’ study designs elements and results that conform to the researcher’s hypothesis. The multi-stakeholder RWE Transparency Initiative, is working to ‘establish a culture of transparency for study analysis and reporting of hypothesis evaluating real-world evidence studies [45].’ A key objective of the initiative is to develop recommendations for RWE study protocol registration on platforms like clinicaltrials.gov and the European Union’s electronic Register of Post-Authorization Studies; these recommendations are intended to increase understanding of how researchers arrived at the results presented. Pre-registering the study protocol is a big step toward transparency, but decision-makers and stakeholders still need to determine what components are necessary to include in that protocol. While multiple tools exist, there is a lack of consensus on what tool is most appropriate. We would also encourage decision-makers and researchers to develop best practices around engagement with decision-makers during study design and execution in order to facilitate transparency in the RWE generation process.

Another critical area that has received recent attention is the heterogeneity of RWE quality and challenges in data access. Groups like the European Health Data and Evidence Network (EHDEN) are working on harmonizing RWD across the EU and building an infrastructure for easier data access [66]. While these are larger structural data quality issues, groups, like FOCCR, are also working on expanding and improving RWD capture to meet the needs of more personalized healthcare research questions, like how to incorporate patient reported outcomes (PRO) data in RWD (‘data sources, data quality’ building block, Table 2). With the patient at the center of any healthcare system, ensuring that treatments are positively impacting patient-relevant outcomes (e.g., quality of life) is essential and there is evidence that PRO capture during routine clinical care can also improve treatment outcomes [67]. Typical RWD sources like claims and electronic medical records are not always well-structured to capture PRO elements. In oncology, PRO-Common Terminology Criteria for Adverse Events (PRO-CTCAE), an FDA validated PRO instrument, was developed by the National Cancer Institute to capture self-reported symptoms of toxicity in clinical trials [68]. Unfortunately, this instrument does not appear to be well captured in RWD sources. Understanding how data providers and intermediaries can enhance the capture and validation of PROs, like the PRO-CTCAE or other instruments like PROMIS [69] or the EQ-5D [70], in RWD/RWE requires collaboration across multiple stakeholders. Not all RWE research questions will require PRO outcomes, but a larger push to incorporate PROs in RWD sources can expand the applicability of RWE as the industry moves to more personalized medicine. FOCCR has assembled a coalition to validate common oncology end points, including PROs, in RWD in order to improve the capture of RWE in routine practice, where it can be a reliable metric to evaluate oncology product effectiveness [71].

### Agreeing on what is 'high-quality' RWE & offering consistent comprehensive guidance for researchers

In the move toward comprehensive guidance, a consensus, where possible, should be reached in defining high-quality RWE. Historically, agreement on evidence requirements (i.e., what evidence is needed to make decisions like placebo controlled studies vs active comparator studies) between regulators and HTA agencies, and even among HTA agencies, has been low; in the EU alone, there are approximately 23 countries that have HTA systems with different assessment frameworks and varying requirements for clinical and economic evidence [72]. Differences in evidence requirements among HTAs and regulators is not surprising due to the differences in the questions being put to the data. Regulatory agencies are focused on safety and efficacy of new treatments, while HTA bodies are concerned with the relative effectiveness and value of treatments compared with those already being used. Additionally, regulatory agencies have the ability to require evidence; HTAs, on the other hand, are often limited to the evidence available during the regulatory approval process and this evidence often lacks long-term outcomes, which are more relevant for value assessments. For example, while NICE has issued a Methods Guide – a broad high-level summary for its evidence preferences – and supporting 'Technical Support Documents' to provide recommendations on technical issues such as adjusting for bias, NICE does not dictate what evidence is required and gives broad leeway to the manufacturer to suggest their own value proposition. Furthermore, the HTAs across Europe do not have a unified structure for their assessment process and reimbursement decision-making. In fact, the 23 different HTA bodies in Europe have different legal remits, assessment methods and acceptance levels to using non-randomized data.

Stakeholders have recognized the lack of alignment to date in evidence requirements among decision-makers and there is a push for increased coordination, especially in the area of RWE use. Multi-country and multi-national collaborations currently exist across regulatory and HTA agencies. For example, the EMA and FDA recently strengthened their collaboration and agreed to develop a 'roadmap for international collaboration on real world evidence' [73]. The IMI GetReal project brings together EU regulators, HTAs, patient advocacy organizations and academics – stakeholders with potentially competing goals – to collaborate in a 'safe harbor' environment, identify key issues, and co-develop solutions for RWE generation and use [74]. ISPOR-ISPE's RWE Task Forces similarly spans multiple stakeholders with the goal of advancing collaboration on recommendations for building blocks of high-quality RWE standards [75]. Collaboration between regulators, HTA bodies and other stakeholders to harmonize evidence requirements where appropriate is in progress and this eye toward harmonization should continue for RWE guidance. While there may always be some gap between the evidence required to make regulatory versus HTA decisions, both groups agree in requiring high-quality evidence for their decision-making. Consensus on what is high-quality RWE is not out of reach even if regulators and HTAs require different types of high-quality RWE. Collaboration is needed to ensure comprehensive guidance is not created in silos and, where possible, can be applied across decision-making forums. Comprehensive guidance that focuses on key elements of quality RWE (e.g., data quality, study design, appropriate analytical methods, etc.) should also transcend geographic borders; as long as the methodology is robust, sound methodological guidance developed in one region could be applied universally. The proposed organized structure can help identify disparate recommendations and areas where consensus is needed.

### Demonstration projects to accelerate progress to comprehensive guidance

For building blocks that need further development, demonstration projects can be used to improve methodology through prospective validation of methods which will bolster credibility by showcasing where RWE is and is not an appropriate choice. These projects are concrete use cases to pressure test the applicability and validity of RWE for decision-making. For example, several on-going demonstration projects are funded by FDA and serve to provide additional evidence for the FDA's RWE guidance document that is expected in 2021. RCT DUPLICATE is recreating 30 already completed RCTs and 7 prospective RCTs using RWD [60]. The RCTs selected for duplication span multiple therapeutic areas, including those outside the traditional uses of RWE (i.e., rare disease and oncology). The goal is to mimic the RCT as much as possible to understand the conditions necessary for RWE studies to come to the same causal conclusions. While the benefit of RWE is that it reflects the real-world experience of patients and often includes different patient populations and treatment pathways than RCTs, it is important to start with a foundational understanding of RWE's ability to study causal questions and where RWE 'gets it right' and importantly, where RWE 'gets it wrong.' The FDA will be using the insights from RCT DUPLICATE to set guidelines on relevant use cases for RWE in regulatory decision-making (e.g., the applicability of RWE in label

expansions). The FDA, through the Reagan-Udall Foundation is also collaborating with FOCCR on the COVID-19 Evidence Accelerated to use RWE to advance the US's response to the pandemic [76]. The EMA and Health Canada are co-leading work on real-world evidence to support decision-making in COVID-19. This work focuses on vaccine vigilance, building international cohorts to study disease epidemiology, the performance of drugs and pregnancy surveillance. This work will help to build consensus between international regulators on what defines 'high-quality' RWE and on its place in decision-making [77]. Demonstration projects can provide valuable insights to establish the utility of RWE and pressure test assumptions as to where RWE can be used. Insights from these projects can bolster recommendations in the building blocks and pave the way to comprehensive guidance.

## Conclusion

This structure is the first to organize the myriad RWE recommendations in order to understand progress made toward comprehensive guidance and to identify gaps where more research, especially demonstration projects, is needed. For example, the use of ECAs (Figure 2, 'study design, fit-for-purpose design' building block), 'transparency and reproducibility' (Figure 2) and methods around PRO data capture (Figure 2, 'data sources, data quality' building block) are areas that would benefit from additional development and were described in detail here. In addition to these selected spotlights, gaps in current recommendations were identified in each building block (Table 2). Substantial headway has been made for a number of aspects of RWE quality (e.g., defining what is 'fit-for-purpose data,' Figure 2), however using the proposed organized structure, it is time to collectively assess the current status of recommendations for RWE quality and use. Public health will advance most quickly if stakeholders and researchers can agree on the parameters for each building block. This will allow researchers to conduct better RWE studies, which will be more likely to support decision-making.

As decision-makers are developing their RWE guidance, they should take the lead on adopting the recommendations that meet their expectations for 'quality.' Researchers and Professional Organizations, like ISPOR, should also use the organized structure to identify areas where their research initiatives can make a contribution to RWE standard setting. Collaborations between stakeholders can include formal cross-organizational projects to develop consensus within the building blocks, sharing learnings from recommendation development to avoid silos, and co-authoring publications and presentations to discuss rationales for recommendations. While collaborative and demonstration projects will bring substantial progress, smaller more experienced research groups may be able to contribute to filling smaller gaps via peer reviewed research with greater speed.

With the amount of attention on RWE and the investment from a wide variety of stakeholders to get RWE right, there is incentive for stakeholders to collaborate and make rapid progress toward comprehensive guidance on what is high-quality decision-grade RWE and the conditions for when it should be used in decision-making. This consensus and pathway for RWE use will help propel the generation of high-quality RWE where we can further learn by applying the principles of use in decision-making, which will help address skepticism around the utility of RWE in healthcare decision-making. Once there is trust in RWE as a relevant and valid option for healthcare decision-making, stakeholders can amend the traditional hierarchical pyramid of evidence and move toward greater adoption of RWE where it is most impactful.

## Future perspective

As healthcare decision-makers look beyond traditional evidence sources (i.e., RCTs), RWE will become a more widely used source of evidence in regulatory and HTA decision-making and the hierarchical pyramid of evidence sources will be updated to include a number of methodologies and data sources. Guided by our organized structure and recommendation gap analysis, the ongoing work by decision-makers to set methodological standards around the generation and use of RWE and the 'learn by doing' approach of demonstration projects, especially in response to the COVID pandemic, will increase trust and quell skepticism in the reliability and validity of high-quality RWE.

### Executive summary

#### Comprehensive guidance on real-world evidence generation & use is needed

- With the increased focus on real-world evidence (RWE) to aid regulatory and HTA decisions, guidance is needed to define 'high-quality' RWE and determine how it should and should not be used in decision-making.
- Decision-makers and other stakeholders have issued numerous position pieces and recommendations on different aspects of RWE, however, to date, the recommendations are fragmented; no comprehensive guidance has been issued.

#### Moving from fragmented RWE recommendations to comprehensive guidance

- We offer a structure for organizing the landscape of current RWE recommendations which is structured by the key building blocks of high-quality RWE and follows the typical research workflow of hypothesis generation through study design and execution. Within the building blocks, we summarized the current state of RWE recommendations and identified gaps.
- In order to move from fragmented RWE recommendations to comprehensive guidance, we suggest decision-makers utilize the organized structure to adopt existing recommendations, close gaps in their current recommendations, and identify areas where further work is needed.
- Stakeholders should collaborate on demonstration projects to improve methodology through prospective validation of methods which will bolster credibility by showcasing where RWE gets it right and where it is not helpful.

### Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: [www.futuremedicine.com/doi/suppl/10.2217/cer-2020-0228](http://www.futuremedicine.com/doi/suppl/10.2217/cer-2020-0228)

### Author contributions

N Gatto, J Wu, P Jonsson, H-G Eichler suggested recommendation documents for inclusion. A Jaksa completed the targeted search. A Jaksa and S Vititoe extracted data from the recommendation documents. A Jaksa and N Gatto designed the organized structure and drafted the article. J Wu, P Jonsson and H-G Eichler offer the manufacturer, HTA and regulatory perspective, respectively. All authors contributed critical revisions to the article. A Jaksa, N Gatto, J Wu, P Jonsson, S Vititoe, and H-G Eichler had final approval of the version to be published.

### Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

### References

Papers of special note have been highlighted as: • of interest

1. National Institute for Health and Care Excellence. Increasing use of health and social care data in guidance development (2020). [www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/how-we-develop-nice-guidelines/data-and-analytics-statement-of-intent](http://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/how-we-develop-nice-guidelines/data-and-analytics-statement-of-intent)
2. US Food and Drug Administration. Framework for FDA's Real-World Evidence Program (2018). [www.fda.gov/media/120060/download](http://www.fda.gov/media/120060/download)
  - **FDA's RWE Framework outlines key considerations that it will explore when developing guidance documents to assist manufacturers interested in using real-world evidence (RWE).**
3. European Medicines Agency. EMA Regulatory Science to 2025 (2020). [www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection\\_en.pdf](http://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf)
  - **The European Union's focus on understanding and using Big Data in regulatory and HTA decision-making is a cornerstone of the EMA's strategic focus through 2025.**

4. Office of the Commissioner. Coronavirus (COVID-19) Update: FDA Collaborations Promote Rigorous Analyses of Real-World Data to Inform Pandemic Response. FDA (2020). [www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-collaborations-promote-rigorous-analyses-real-world-data-inform](http://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-collaborations-promote-rigorous-analyses-real-world-data-inform)
5. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N. Engl. J. Med.* 382(7), 674–678 (2020).
6. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular disease, drug therapy, and mortality in Covid-19. *N. Engl. J. Med.* 382(25), e102 (2020).
7. Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet* (2020). [www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)31180-6/abstract](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31180-6/abstract)
8. Avorn J. In Defense of Pharmacoepidemiology — Embracing the Yin and Yang of Drug Research. *N. Engl. J. Med.* 357(22), 2219–2221 (2007).
9. Griffiths EA, Macaulay R, Vadlamudi NK, Uddin J, Samuels ER. The role of noncomparative evidence in health technology assessment decisions. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 20(10), 1245–1251 (2017).
10. Makady A, Ham RT, de Boer A *et al.* Policies for use of real-world data in health technology assessment (HTA): a comparative study of six HTA agencies. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 20(4), 520–532 (2017).
11. Morton SC, Costlow MR, Graff JS, Dubois RW. Standards and guidelines for observational studies: quality is in the eye of the beholder. *J. Clin. Epidemiol.* 71, 3–10 (2016).
12. IQWiG. Development of scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a Social Code Book V – rapid report (2020). [www.iqwig.de/en/projects-results/projects/drug-assessment/a19-43-development-of-scientific-concepts-for-the-generation-of-routine-practice-data-and-their-analysis-for-the-benefit-assessment-of-drugs-according-to-35a-social-code-book-v-rapid-report.11901.html](http://www.iqwig.de/en/projects-results/projects/drug-assessment/a19-43-development-of-scientific-concepts-for-the-generation-of-routine-practice-data-and-their-analysis-for-the-benefit-assessment-of-drugs-according-to-35a-social-code-book-v-rapid-report.11901.html)
13. Duke-Margolis Center for Health Policy. Understanding the need for non-interventional studies using secondary data to generate real-world evidence for regulatory decision making, and demonstrating their credibility (2019). <https://healthpolicy.duke.edu/publications/understanding-need-non-interventional-studies-using-secondary-data-generate-real-world>
14. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin. Pharmacol. Ther.* 105(4), 867–877 (2019).
15. US Food and Drug Administration. Demonstrating substantial evidence of effectiveness for human drug and biological products (2019). [www.fda.gov/media/133660/download](http://www.fda.gov/media/133660/download)
16. European Medicines Agency. Choice of control group in clinical trials (2001). [www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5\\_en.pdf](http://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf)
17. Friends of Cancer Research. Characterizing the use of external controls for augmenting randomized control arms and confirming benefit (2019). [www.focr.org/sites/default/files/Panel-1\\_External\\_Control\\_Arms2019AM.pdf](http://www.focr.org/sites/default/files/Panel-1_External_Control_Arms2019AM.pdf)
18. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value Health.* 12(8), 1053–1061 (2009).
19. Franklin JM, Schneeweiss S, Solomon DH. Assessment of confounders in comparative effectiveness studies from secondary databases. *Am. J. Epidemiol.* 185(6), 474–478 (2017).
20. Gatto NM, Reynolds RF, Campbell UB. A structured preapproval and postapproval comparative study design framework to generate valid and transparent real-world evidence for regulatory decisions. *Clin. Pharmacol. Ther.* 106(1), 103–115 (2019).
21. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available: table 1. *Am. J. Epidemiol.* 183(8), 758–764 (2016).
22. Canada H. Elements of real world data/evidence quality throughout the prescription drug product life cycle (2020). [www.canada.ca/en/services/health/publications/drugs-health-products/real-world-data-evidence-drug-lifecycle-report.html](http://www.canada.ca/en/services/health/publications/drugs-health-products/real-world-data-evidence-drug-lifecycle-report.html)
23. Center for Drug Evaluation and Research. Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data sets. US Food Drug Adm (2020). [www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic](http://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic)
24. *Developing a protocol for observational comparative effectiveness research: a user's guide.* Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. (Eds). Agency for Healthcare Research and Quality (US), MD, USA, [www.ncbi.nlm.nih.gov/books/NBK126190/](http://www.ncbi.nlm.nih.gov/books/NBK126190/)
25. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP) (2015). [www.pharmacoepi.org/resources/policies/guidelines-08027/](http://www.pharmacoepi.org/resources/policies/guidelines-08027/)
26. Berger ML, Sox H, Willke RJ *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(9), 1033–1039 (2017).

27. Newhouse R, Goodman S, Aronson N *et al*. PCORI Methodology Report (2019). [www.pcori.org/sites/default/files/PCORI-Methodology-Report.pdf](http://www.pcori.org/sites/default/files/PCORI-Methodology-Report.pdf)
28. European Medicines Agency. Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies (2012). [www.ema.europa.eu/en/documents/other/guidance-format-content-protocol-non-interventional-post-authorisation-safety-studies\\_en.pdf](http://www.ema.europa.eu/en/documents/other/guidance-format-content-protocol-non-interventional-post-authorisation-safety-studies_en.pdf)
29. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. ENCePP Checklist for Study Protocols (2018). [www.encepp.eu/standards\\_and\\_guidances/checkListProtocols.shtml](http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml)
30. HMA/EMA Joint Task Force on Big Data. Observational data (Real World Data) Subgroup report (2019). [www.ema.europa.eu/en/documents/report/observational-data-real-world-data-subgroup-report\\_en.pdf](http://www.ema.europa.eu/en/documents/report/observational-data-real-world-data-subgroup-report_en.pdf)
31. US Food and Drug Administration. Use of electronic health record data in clinical investigations guidance for industry (2018). [www.fda.gov/media/97567/download](http://www.fda.gov/media/97567/download)
32. Medicines & Healthcare Products Regulatory Agency. Medicines & Healthcare products Regulatory Agency (MHRA) 'GXP' Data Integrity Guidance and Definitions (2018). [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/687246/MHRA\\_GxP\\_data\\_integrity\\_guide\\_March\\_edited\\_Final.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf)
33. CanREValue. Mapping Canadian Provincial Data Assets to Conduct Real-World Studies on Cancer Drugs CanREValue Collaboration Data Working Group Interim Report 2020 (2020). [https://cc-arcc.ca/wp-content/uploads/2020/04/The-CanREValue-Data-WG-Interim-Report-Revision\\_Final\\_v1.pdf](https://cc-arcc.ca/wp-content/uploads/2020/04/The-CanREValue-Data-WG-Interim-Report-Revision_Final_v1.pdf)
34. EUNetHTA JA3. Registry Evaluation and Quality Standards Tool (REQueST<sup>®</sup>). EUNetHTA (2019). <https://eunetha.eu/request-tool-and-its-vision-paper/>
35. Duke-Margolis Center for Health Policy. Characterizing RWD Quality and Relevancy for Regulatory Purposes (2018). <https://healthpolicy.duke.edu/sites/default/files/2020-08/Characterizing%20RWD%20for%20Regulatory%20Use.pdf>
36. Duke-Margolis Center for Health Policy. Determining Real-World Data's Fitness for Use and the Role of Reliability (2019). [https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd\\_reliability.pdf](https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf)
37. US Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices - guidance for industry and food and drug administration staff (2017). [www.fda.gov/media/99447/download](http://www.fda.gov/media/99447/download)
38. Hall GC, Sauer B, Bourke A *et al*. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol. Drug Saf.* 21(1), 1–10 (2012).
39. Faria R, Alava MH, Manca A, Wailoo AJ. The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data (2015). <http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf>
40. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good Research Practices for Comparative Effectiveness Research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value Health* 12(8), 1062–1073 (2009).
41. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol. Drug Saf.* 15(5), 291–303 (2006).
42. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 20(4), 512–522 (2009).
43. Schneeweiss S, Rassen JA, Brown JS *et al*. Graphical depiction of longitudinal study designs in health care databases. *Ann. Intern. Med.* 170(6), 398 (2019).
44. Wang SV, Schneeweiss S, Berger ML *et al*. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol. Drug Saf.* 26(9), 1018–1032 (2017).
45. Orsini LS, Berger M, Crown W *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(9), 1128–1136 (2020).
46. Benchimol EI, Smeeth L, Guttman A *et al*. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med.* 12(10), e1001885 (2015).
47. Langan SM, Schmidt SA, Wing K *et al*. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* k3532 (2018).
48. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 335(7626), /bmj/335/7626/0.2.atom (2007).
49. Sterne JA, Hernán MA, Reeves BC *et al*. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. i4919 (2016).
50. GRACE Principles: Recognizing High-Quality Observational Studies of Comparative Effectiveness. *AJMC*. [www.ajmc.com/journals/issue/2010/2010-06-vol16-n06/ajmc.10jundreyer.467to471](http://www.ajmc.com/journals/issue/2010/2010-06-vol16-n06/ajmc.10jundreyer.467to471)

51. Motheral B, Brooks J, Clark MA *et al.* A checklist for retrospective database studies—report of the ispor task force on retrospective databases. *Value Health* 6(2), 90–97 (2003).
52. Berger ML, Martin BC, Huseareu D *et al.* A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: An ISPOR-AMCP-NPC Good Practice Task Force Report. *Value Health* 17(2), 143–156 (2014).
53. Duke-Margolis Center for Health Policy. A framework for regulatory use of real-world evidence (2017). [https://healthpolicy.duke.edu/sites/default/files/2020-03/rwe\\_white\\_paper\\_2017.09.06.pdf](https://healthpolicy.duke.edu/sites/default/files/2020-03/rwe_white_paper_2017.09.06.pdf)
54. JA3 Work Package 5 - Lifecycle approach to improve evidence generation. EUnetHTA. <https://eunetha.eu/ja3-archive/work-package-5-life-cycle-approach-to-improve-evidence-generation/>
55. IMI GetReal - Workpackage 3. [www.imi-getreal.eu/About-GetReal/Workpackage-3](http://www.imi-getreal.eu/About-GetReal/Workpackage-3)
56. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. *J. Comp. Eff. Res.* 7(12), 1133–1143 (2018).
57. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR real-world data task force report. *Value Health* 10(5), 326–335 (2007).
58. Bell H, Wailoo AJ, Hernandez M *et al.* The use of real world data for the estimation of treatment effects in NICE decision making (2016). <http://nice.dsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf>
59. RWE4Decisions. Real-world evidence to support HTA/payer decisions about highly innovative technologies (2019). [https://8c3e11d9-5f36-452f-abe3-c95befd6e85d.filesusr.com/ugd/e1a359\\_56d20ef27bba4a949a4adf7cae77cd35.pdf](https://8c3e11d9-5f36-452f-abe3-c95befd6e85d.filesusr.com/ugd/e1a359_56d20ef27bba4a949a4adf7cae77cd35.pdf)
60. Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology (RCT DUPLICATE). RCT Duplic. [www.rctduplicate.org/](http://www.rctduplicate.org/)
- **FDA funded demonstration project to duplicate 37 randomized control trials using RWE. Results from all 37 replication studies are expected at the end of 2020.**
61. Reproducible Evidence: Practices to Enhance and Achieve Transparency (REPEAT). REPEAT Initiat. [www.repeatinitiative.org/](http://www.repeatinitiative.org/)
62. Gustafsson S, Johansson P, Ponten J, Stromgren A, Viber A. Follow-up of drug utilisation and treatment effects in clinical practice (2018). [www.tlv.se/in-english/reports/arkiv/2020-12-09-follow-up-of-drug-utilisation-and-treatment-effects-in-clinical-practice.html](http://www.tlv.se/in-english/reports/arkiv/2020-12-09-follow-up-of-drug-utilisation-and-treatment-effects-in-clinical-practice.html)
63. Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz HM, Ross JS. Assessment of Clinical Trials Supporting US Food and Drug Administration Approval of Novel Therapeutic Agents, 1995-2017. *JAMA Netw. Ope* 3(4), e203284 (2020).
64. Committee for Medicinal Products for Human Use (CHMP). Assessment report: blincyto (2015). [www.ema.europa.eu/en/documents/assessment-report/blincyto-epar-public-assessment-report\\_en.pdf](http://www.ema.europa.eu/en/documents/assessment-report/blincyto-epar-public-assessment-report_en.pdf)
65. Center for Drug Evaluation and Research. Statistical Review(s): Application Number 125557Orig1s000 (2014). [www.accessdata.fda.gov/drugsatfda\\_docs/nda/2014/125557Orig1s000StatR.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000StatR.pdf)
66. EHDEN. EHDEN Vision, mission & objectives. [ehden.eu](http://ehden.eu) (2020). [www.ehden.eu/vision-and-mission/](http://www.ehden.eu/vision-and-mission/)
67. Basch E, Deal AM, Dueck AC *et al.* Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 318(2), 197 (2017).
68. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)= (2020). <https://healthcaresdelivery.cancer.gov/pro-ctcae/>
69. PROMIS: Clinical Outcomes Assessment (2019). <https://commonfund.nih.gov/promis/index>
70. EQ-5D instruments – EQ-5D (2021). <https://euroqol.org/eq-5d-instruments/>
71. Real-World Evidence | Friends of Cancer Research. [www.focr.org/rwe](http://www.focr.org/rwe)
72. Chamova J. Europäische Kommission, Generaldirektion Gesundheit und Lebensmittelsicherheit. Mapping of HTA national organisations, programmes and processes in EU and Norway (2017). <https://doi.org/10.2875/5065>
73. European Commission, EMA and FDA agree new priorities to strengthen their collaboration on medicines. *Eur. Med. Agency* (2020). [www.ema.europa.eu/en/news/european-commission-ema-fda-agree-new-priorities-strengthen-their-collaboration-medicines](http://www.ema.europa.eu/en/news/european-commission-ema-fda-agree-new-priorities-strengthen-their-collaboration-medicines)
74. IMI GetReal - New methods for RWE collection and synthesis (2021). [www.imi-getreal.eu/About-GetReal/Overall-objectives](http://www.imi-getreal.eu/About-GetReal/Overall-objectives)
75. Joint ISPE-ISPOR Special Task Force (2021). [www.ispor.org/member-groups/task-forces/joint-ispe-isor-special-task-force](http://www.ispor.org/member-groups/task-forces/joint-ispe-isor-special-task-force)
76. Evidence Accelerator (2021). <https://evidenceaccelerator.org/>
77. Global regulatory workshop on COVID-19 real-world evidence and observational studies. *Eur. Med. Agency* (2020). [www.ema.europa.eu/en/news/global-regulatory-workshop-covid-19-real-world-evidence-observational-studies](http://www.ema.europa.eu/en/news/global-regulatory-workshop-covid-19-real-world-evidence-observational-studies)
- **With the urgent need to understand the implications of the COVID-19, observational studies have become a prominent tool of researchers. Global regulators are collaborating to use RWE in studying vaccines, treatments and cohorts typically not included in RCTs (e.g., pregnant women).**