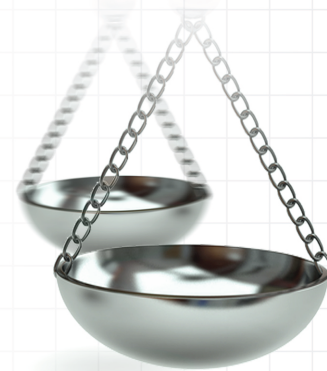




RWE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 16

Journal of **Comparative Effectiveness Research**

Alejandra Castanon¹, Antonia Tsvetanova¹ & Sreeram V Ramagopalan^{*,1,2}

¹ Lane Clark & Peacock LLP, London, W1U 1DQ, UK

² Centre for Pharmaceutical Medicine Research, King's College London, London, SE1 9NH, UK

*Author for correspondence: sreeram.ramagopalan@lcp.uk.com

In this update, we discuss recent US FDA guidance offering more specific guidelines on appropriate study design and analysis to support causal inference for non-interventional studies and the launch of the European Medicines Agency (EMA) and the Heads of Medicines Agencies (HMA) public electronic catalogues. We also highlight an article recommending assessing data quality and suitability prior to protocol finalization and a *Journal of the American Medical Association*-endorsed framework for using causal language when publishing real-world evidence studies. Finally, we explore the potential of large language models to automate the development of health economic models.

First draft submitted: 12 June 2024; Accepted for publication: 26 June 2024; Published online: 5 July 2024

Keywords: causal inference • causal language framework • data quality assessment • FDA guidance • health technology assessment • large language models • non-interventional studies • public electronic catalogues • real-world data • real-world evidence • study design

Recently the US FDA released guidance for manufacturers looking to submit a non-interventional study aiming to contribute to a demonstration of substantial evidence of effectiveness and/or evidence of safety of a drug [1]. This document builds on previous FDA guidance that outlined expectations for real-world evidence (RWE) studies, including transparency in data source selection, study design and study conduct. This earlier guidance called for clear documentation of study design choices in the study protocol, public registration of protocols and ensuring that the FDA is able to access source data [2–4]. The FDA's latest guidance [1] provides more detail on using RWE for causal inference, in particular details around identifying and addressing confounding and other sources of bias when planning and conducting non-interventional studies. Key recommendations for addressing confounding and bias include:

- Prespecifying the research question, study design and analytic approach in the study protocol and statistical analysis plan before initiating the study. This includes providing a rationale for the chosen non-interventional study design and discussing alternative approaches considered and reasons for not choosing them.
- Identifying relevant covariates and developing strategies to address potential bias.
- Using causal diagrams (e.g., directed acyclic graphs or single-world intervention graphs) to specify the theorized causal relationship.
- Carefully defining the index date (time zero) for all study arms and addressing any immortal time bias.
- Defining the start and end of follow-up period and the planned approach for censoring.
- Specification of the estimand and detailing the statistical approach to account for potential confounding factors.
- Assessment of unmeasured confounding and the evaluation of potential overadjustment of intermediate variables on the causal pathway.
- Addressing the potential for unequal detection of outcomes across comparison groups (differential surveillance or misclassification) and reverse causality.

- Handling missing or misclassified data and multiplicity issues.
- Conducting planned sensitivity analyses to assess the robustness of study findings to changes in key factors or assumptions.

This guidance aligns with previous FDA guidance on design considerations for external control arms [5] by specifying the need for detailed analytical approaches to support causal inference and accounting for potential confounding and bias [6]. However, the external control guidance explicitly called out the estimand framework to support best practice for doing this, and this guidance highlights target trial emulation as one of the causal frameworks that could be used. Evidence submitted to regulators typically forms the main body of evidence used to support subsequent market access activities. Broadly there is some alignment across many regulators and health technology assessment (HTA) agencies for using the target trial emulation framework when conducting RWE studies to measure treatment effectiveness. There are however some minor differences – for example, single-world intervention graphs are not mentioned in guidance from the National Institute for Health and Care Excellence (NICE) or Canada's Drug Agency (formerly CADTH) [7,8]. Whether or not the target trial framework is the framework used, what is crucial is that there is transparency in all aspects of conducting a study – from choosing a data source to analyzing the results [9]. It remains to be seen if there ends up being a preference by decision makers for one causal framework over another, and whether or not the individual nuances of guidance differences between HTA agencies and regulators impacts acceptance of RWE.

On the theme of protocol development, in a recent editorial, Wang and Schneeweiss [10] discuss the importance of carrying out data checks prior to registering a protocol for a RWE study seeking to answer a causal question, based on their learnings from being key members of the RCT-Duplicate initiative [11,12]. They argue that without these evaluations, studies often prove unfeasible due to poor data quality, suitability or reliability. Wang and Schneeweiss [10] suggest data quality evaluation should focus on the measurement of outcomes and confounders to ensure that these elements are accurately captured and reliable for the study's objectives. Additionally, a thorough feasibility assessment should be conducted. This includes determining the size of the study population after applying eligibility criteria and examining the incidence of study end points in the comparator group to ensure sufficient event rates for meaningful (well-powered) comparisons. Assessing the distribution of pre-treatment patient characteristics before and after applying statistical balancing techniques such as propensity score matching or weighting is also recommended, ensuring that groups can be appropriately compared. An argument for these data checks is that by not estimating study end points in the exposure group one can assess the suitability of data sources without undermining the study itself. To ensure transparency and eliminate the risk of post-hoc data mining, it is essential to register the study protocol after appropriate data checks have been completed. As co-authors of the PRINCIPLED framework, it is unsurprising that this editorial aligns with it [13]. By following this approach, it should allow manufacturers to identify the best data source, be guided on the best timing as to when to conduct a study (i.e., when there is enough power to study outcomes) and to minimize bias before study initiation.

While randomized clinical trials (RCTs) remain the preferred approach for answering causal questions, a recent article in the *Journal of the American Medical Association* (JAMA) argues that carefully designed analyses of observational data can provide valuable evidence when strong assumptions are met and trials are not feasible [14]. As such, this should allow for the use of causal language when publishing these studies. The authors propose a framework for observational studies looking to show causal inference based on six core questions: what is the causal question, what is the causal estimand, what is the study design (including what data to be used), what causal assumptions are being made, how can the observed data be used to answer the causal question (including statistical methods) and is causal interpretation tenable? The framework practically allows for the use of causal language to state research questions, describe methods and assumptions, as well as to determine whether causal interpretations are appropriate. As journals are often reluctant to use causal language for RWE studies, this article reflects a major step forward in terms of the recognition of how RWE can provide causal inference. The article aligns well with published HTA frameworks, which support a similar transparent reporting of methods and results of RWE studies. The endorsement of causal language in RWE studies by journals, in combination with a plethora of RWE guidance from HTA agencies, regulators and medical societies underscores the growing importance of RWE for decision making.

To support good quality and reliable RWE generation the Heads of Medicines Agency (HMA) and the European Medicines Agency (EMA) have introduced two public electronic catalogues – one for real-world data (RWD) sources and one of RWD studies [15]. The initiative builds on two former databases. The catalogue for RWD

sources replaces the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Resources Database [16]. Currently the catalogue contains 217 data sources from a variety of countries (including countries outside of Europe such as the USA). Types of data sources include registries, biobanks, electronic medical records and administrative claims. The catalogue for RWD studies expands and replaces the European Union electronic register of post-authorization studies (EU PAS Register). Both catalogues adhere to 'FAIR' data principles (findable, accessible, interoperable and reusable) [17] and employ a standardized set of metadata to describe and connect data sources to studies. The launch of these catalogues brings the European regulator a step closer to achieving their ambition of more data-driven decisions by improving the discoverability of data. While the data source catalogue is currently not exhaustive, manufacturers should welcome its availability in order to assist with the identification of sources of RWD. It is unclear if these data sources are all fit for purpose however, and for HTA submissions it is important they are assessed using a tool such as the NICE Data Suitability Assessment Tool (DataSAT) to ensure they are the most appropriate data source to address specific research questions [18].

We have previously discussed [19] the opportunities machine learning methods may provide in health economics and outcomes research (HEOR). Demonstration projects that highlight the usefulness of these methods are however needed. Such a study was recently published by Reason *et al.* [20]. The study shows how health economic models can be developed using Generative Pre-Trained Transformer 4 (GPT-4), a large language model (LLM). Reason *et al.* selected two published partitioned survival models evaluating interventions in non-small-cell lung cancer (NSCLC) and renal cell carcinoma (RCC) for replication using GPT-4. LLMs generate text content based on inputs known as 'prompts'. Text-based prompts can use any text-based form, however there is ongoing research into 'effective' prompts, where these prompts are those most likely to produce an output of the desired form. Reason and colleagues iteratively developed prompts instructing GPT-4 to program the NSCLC and RCC models in the R programming language. These prompts included descriptions of each model's methods, assumptions and parameter values. To assess the accuracy and variability of GPT-4's output, 15 replicas of each model were created. The results from the artificial intelligence (AI)-generated models were then compared with the published values obtained from the original, human-programmed models. GPT-4 accurately replicated the NSCLC model, with 93% (14/15) of AI-generated models being error-free. The RCC model was closely replicated but required simplification of one input calculation. With this change, 60% (9/15) of AI-generated RCC models were error-free. Error-free models reproduced published cost-effectiveness ratios within 1%. The development of a health economic model typically involves four phases: conceptualization of the model, estimating parameter values, constructing the model and validating the model. Reason and colleagues were able to show how the third phase, model construction, can be automated using an LLM. The use of LLMs to develop models offers the potential to significantly accelerate the time taken for model development. Further work is needed to improve accuracy and testing in other disease areas; however, it remains to be seen how HTA agencies will respond to LLM developed models. Although this demonstration study did not include RWE, LLM will nevertheless be used in this context. This will necessitate HTA RWE guidance on appropriate usage of AI models when generating evidence.

In conclusion, recent developments in RWE guidance from the FDA, causal language guidance from JAMA and the launch of the HMA-EMA catalogues of RWD sources demonstrate the growing importance of RWE for decision making. Furthermore, the potential of LLMs to automate health economic model development highlights the expanding role of AI in HEOR. As RWE continues to gain traction, it is crucial for manufacturers to stay informed about evolving guidance and best practices to ensure the generation of high-quality, reliable evidence that meets the standards of regulatory and HTA bodies.

Financial disclosure

SV Ramagopalan has received an honorarium from Becaris Publishing for the contribution of this work. The authors have no other 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 apart from those disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity 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.

Writing disclosure

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 <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. US Food & Drug Administration. Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry. (March 2024). Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-evidence-considerations-regarding-non-interventional-studies-drug-and-biological-products>
2. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 6. *J. Comp. Eff. Res.* 11(7), 473–475 (2022).
3. US Food & Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products. (2021). Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory>
4. US Food & Drug Administration. Real-World Data: Assessing Registries To Support Regulatory Decision-Making for Drug and Biological Products. (2023). Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products>
5. Motrinchuk AS, Belousov D. Considerations for the design and conduct of externally controlled trials for drug and biological products. *Real-World Data Evid.* 3(4), 29–40 (2023).
6. Bray B, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 11. *J. Comp. Eff. Res.* 12(5), e230008 (2023).
7. National Institute for Health and Care Excellence. NICE real-world evidence framework. Overview. (2022). Available from: <https://www.nice.org.uk/corporate/ecd9/chapter/overview>
8. Canada's Drug Agency. Guidance for Reporting Real-World Evidence. Available from: <https://www.cadth.ca/guidance-reporting-real-world-evidence>
9. Castanon A, Duffield S, Ramagopalan S, Reynolds R. Why is target trial emulation not being used in health technology assessment real-world data submissions? *J. Comp. Eff. Res.* 13(8), e240091 (2024).
10. Wang SV, Schneeweiss S. Data checks before registering study protocols for health care database analyses. *JAMA* 331(17), 1445–1446 (2024).
11. Bray B, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 12. *J. Comp. Eff. Res.* 12(7), e230092 (2023).
12. Wang SV, Schneeweiss S. RCT-DUPLICATE Initiative. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 329(16), 1376–1385 (2023).
13. Desai RJ, Wang SV, Sreedhara SK *et al.* Process guide for inferential studies using healthcare data from routine clinical practice to evaluate causal effects of drugs (PRINCIPLED): considerations from the FDA Sentinel Innovation Center. *BMJ* 384, e076460 (2024).
14. Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA* 331(21), 1845–1853 (2024).
15. European Medicines Agency. HMA-EMA Catalogues of real-world data sources and studies. Available from: <https://catalogues.ema.europa.eu/>
16. ENCePP. Available from: https://encepp.europa.eu/index_en
17. Nicholson C, Kansa S, Gupta N, Fernandez R. Will It Ever Be FAIR?: Making archaeological data Findable, Accessible, Interoperable, and Reusable. *Adv. Archaeologic. Pract.* 11(1), 63–75 (2023).
18. National Institute for Health and Care Excellence. NICE real-world evidence framework. Appendix 1 – Data Suitability Assessment Tool (DataSAT). (2022). Available from: <https://www.nice.org.uk/corporate/ecd9/chapter/appendix-1-data-suitability-assessment-tool-datasat>
19. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 9. *J. Comp. Eff. Res.* 11(16), 1147–1149 (2022).
20. Reason T, Rawlinson W, Langham J, Gimblett A, Malcolm B, Klijn S. Artificial intelligence to automate health economic modelling: a case study to evaluate the potential application of large language models. *PharmacoEconomics Open* 8(2), 191–203 (2024).