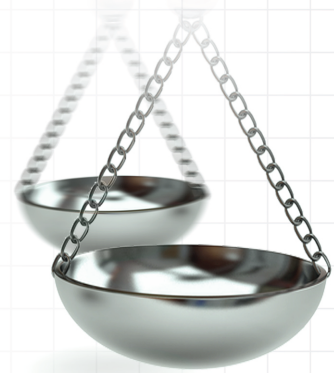
Industry Update

For reprint orders, please contact: reprints@futuremedicine.com



Journal of Comparative Effectiveness Research

R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 9

Alex Simpson[1] & Sreeram V Ramagopalan*,[1]
[1]Global Access, F Hoffmann-La Roche, Basel, Switzerland
*Author for correspondence: sreeram.ramagopalan@roche.com

In this latest update we highlight a recent International Society of Pharmacoeconomics and Outcomes Research Good Practice Report on machine learning (ML) for health economics and outcomes research. We specifically discuss use cases of ML that offer opportunities in the generation of evidence using real-world data, including improvements in the identification of study cohorts, confounder identification and adjustment and estimating treatment effect heterogeneity. Barriers to the wider adoption of ML methods are also discussed.

First draft submitted: 30 May 2022; Accepted for publication: 7 June 2022; Published online: 23 August 2022

Keywords: comparative effectiveness research • health technology assessment • HTA • International Society of Pharmacoeconomics and Outcomes Research • ISPOR • machine learning • real-world evidence • RWE

Randomized controlled trials (RCTs) remain the gold standard tool for assessing effectiveness of treatments for use in health technology assessment (HTA). However, the use of real-world data (RWD) for this purpose plays a crucial role, offering the opportunity to estimate comparative effectiveness when well-powered RCTs are not possible and supplementing RCT evidence by providing estimates over the longer-term and in population subgroups typically excluded from RCTs.

However, not exploiting randomization has limitations, most importantly the increased risk of unmeasured confounding and other biases. Quantitative bias analysis methods can help to explore how unmeasured confounding is likely to impact estimates of treatment and cost–effectiveness [1–3], but perhaps more important are approaches that help to limit rather than address residual confounding.

A recently published International Society of Pharmacoeconomics and Outcomes Research (ISPOR) Good Practice Report outlines the opportunities machine learning (ML) methods may provide in health economics and outcomes research and highlights multiple approaches that may specifically help with minimizing confounding in RWD studies, including cohort selection, confounder identification and adjustment and estimating treatment effect heterogeneity [4].

Cohort selection is highlighted in the report as an area where ML methods offer particular promise. RWD sources, particularly electronic health records, often derive variables from unstructured data such as free-text clinician notes within patient records. Manual abstraction has traditionally been used to extract information from such sources, which is costly and time consuming. The report highlights how ML methods such as natural language processing can increase the efficiency of cohort selection. This involves practitioners first labeling the data, training ML models on this data and then applying them to new data to determine probabilities of cohort eligibility. Detailed human abstraction is then only conducted on patients with probabilities over a certain threshold, enabling more targeted and detailed extraction. A concern raised in the report is that this approach maximizes specificity over sensitivity (generating few false-positives, but many false-negatives), leading to selection bias if false negatives are not randomly sampled. However, work by Flatiron Health suggests bias can be avoided, with an algorithm derived to identify patients with metastatic breast cancer demonstrating considerable sensitivity with no differences in

Future Medicine

average outcomes and baseline/clinical characteristics relative to those in test cohorts generated using only manual abstraction [5].

A second promising use case highlighted in the report is in the identification of confounders. HTA RWD guidelines recommend that confounders are identified using systematic searches of prognostic studies and/or use of expert opinion [6]. However, literature and knowledge gaps are possible, particularly in indications that are less researched, such as rare diseases or disease subgroups. Small sample sizes may also limit the ability to adjust for all potential confounders even if data on them is available. In processes called feature selection and extraction, the report explains how ML methods can be used alongside study sample data to identify new relevant confounders and/or reduce the covariate set to those that are the most prognostic. This may be particularly useful in small samples where traditional statistical techniques are prone to overfitting.

The report also details how ML methods can enhance approaches used to adjust for observed confounding. Assumptions regarding the nature of the relationship between confounders and outcomes should always be guided by causal reasoning and subject matter knowledge. However, just like information on whether a given variable is a confounder, functional forms are often unknown, with many practitioners assuming linear and additive effects and ignoring potential non linearity and interaction effects. The report highlights how common ML methods for prediction and classification can flexibly estimate functional forms semi- or nonparametrically.

Causal ML methods can also flexibly estimate treatment effect heterogeneity [7,8] and the report highlights the usefulness of this in supporting a move toward individualized treatment decisions. However, these approaches may have wider applications. The absence of unmeasured effect modification is a key assumption for methods used to generalize or 'transport' treatment effects from a study to different populations [9] and ML methods are being increasingly recommended for identifying effect modifiers in this setting [10]. The issue of transportability is likely to become increasingly important as the use of RWD becomes more prevalent in HTA, particularly as RWD data sources from one country are commonly used to generate evidence for HTA submissions in another [11].

Despite the benefits and potential uses of ML methods, the report highlights some limitations. ML algorithms are only as good as the data from which they learn, meaning if training data are flawed, so too will the algorithms. This can lead to unintended biases, such as reinforcing racial inequalities where input data are impacted by racial bias such as unequal access to treatment [12]. Transportability may also be a concern for certain ML algorithms. For example, a deep-learning system for identifying diabetic retinopathy developed by Google Health demonstrated high accuracy both in the lab and in clinical practice across multiple countries [13,14], but faced implementation difficulties in low- and middle-income countries, where lower quality imaging was more prevalent [15,16].

ML methods are also inherently complex and the report highlights concerns over the transparency of these methods and their 'explainability' to patients and decision makers. There is a feeling that RWD in and of itself suffers from a lack of acceptability by HTA agencies [17]. If it is felt that complexity and difficulty in understanding is a key driver of non acceptance, manufacturers may be hesitant to use ML methods in HTA submissions. The report encourages ML developers to be transparent about the development and execution of ML methods and develops a 'PALISADE' checklist to aid in this. This may be a helpful first step in encouraging acceptance, but demonstration projects that highlight the usefulness of these methods in a HTA setting may also be needed. More studies examining and addressing potential limitations of ML methods and also exploring in what cases ML methods are superior to existing approaches, may be required to generate the confidence needed to enable wider adoption. Many HTA agencies are yet to provide concrete guidance on when they are willing to accept RWD for guiding their decisions. Reducing or addressing key biases is crucial for improving the acceptance of comparative effectiveness studies using RWD, but will require the application of increasingly complex analytics, including the use of methods such as ML. HTA agencies will therefore need to maintain awareness of methods development in the area and update guidelines accordingly.

Financial & competing interests disclosure

The author SV Ramagopalan has received an honorarium from Future Science Group for the contribution of this work. A Simpson and SV Ramagopalan are employees of F Hoffmann-La Roche. 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.

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

References

1. Leahy TP, Duffield S, Kent S *et al.* Application of quantitative bias analysis for unmeasured confounding in cost–effectiveness modelling. *J. Comp. Eff. Res.* 11(12), 861–870 (2022).
2. Leahy TP, Kent S, Sammon C *et al.* Unmeasured confounding in nonrandomized studies: quantitative bias analysis in health technology assessment. *J. Comp. Eff. Res.* 11(12), 851–859 (2022).
3. Popat S, Liu SV, Scheuer N *et al.* Addressing challenges with real-world synthetic control arms to demonstrate the comparative effectiveness of pralsetinib in non-small-cell lung cancer. *Nat. Commun.* 13(1), 3500 (2022).
4. Padula WV, Kreif N, Vanness DJ *et al.* Machine learning methods in health economics and outcomes research-the PALISADE checklist: a good practices report of an ISPOR task force. *Value Health* 25(7), 1063–1080 (2022).
5. Birnbaum B, Nussbaum N, Seidl-Rathkopf K *et al.* Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. *arXiv* (2020). https://arxiv.org/abs/2001.09765
6. National Institute of Health and Care Excellence. NICE real-world evidence framework (2022). www.nice.org.uk/corporate/ecd9/chapter/overview
7. Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann. Stat.* 47(2), 1148–1178 (2019).
8. Kreif N, Diazordaz K, Moreno-Serra R, Mirelman A, Hidayat T, Suhrcke M. Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia. *Health Serv. Outcomes Res. Methodol.* 22(2), 192–227 (2022).
9. Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernán MA. Extending inferences from a randomized trial to a new target population. *Stat. Med.* 39(14), 1999–2014 (2020).
10. Degtiar I, Rose S. A review of generalizability and transportability. *arXiv* (2021). https://arxiv.org/abs/2102.11904
11. Beal B, Altomare I, Ray J, Bargo D, Adamson B. HTA3 Passport for Travel: proposed framework for transportability of oncology real world evidence. *Value Health* 25(7), S504 (2022).
12. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464), 447–453 (2019).
13. Gulshan V, Peng L, Coram M *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22), 2402–2410 (2016).
14. Raumviboonsuk P, Krause J, Chotcomwongse P *et al.* Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit. Med.* 2, 25 (2019).
15. Beede E, Baylor E, Hersch F *et al.* A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. Presented at: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, Hawaii, USA (2020).
16. Ruamviboonsuk P, Tiwari R, Sayres R *et al.* Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit. Health* 4(4), e235–e244 (2022).
17. Economist Impact. Value of real-world evidence in health technology assessment: lost in translation? (2022) https://impact.economist.com/projects/rwe-in-hta/.