Work Package 6 – D6.2: Recommendations on how to analyse and interpret non-randomised studies

Authors: Seamus Kent[1], Pall Jonsson[1], Maximilian Salcher-Konrad[2], Mary Nguyen[2]

[1] National Institute for Health and Care Excellence (NICE)

[2] London School of Economics and Political Science

In this deliverable, we provide recommendations regarding the design, analysis, and reporting of non-randomised studies, drawing on the meta-epidemiological study reported in Deliverable 6.1 and other literature. In order to support the independent publication of this work, **we request that this deliverable will only be made available publicly once the paper reporting the results of our WP has been published.**

Contents

Summary

Background: Due to the increased use of expedited access programmes by regulators, the greater availability of observational healthcare data, and new analytical methods to address confounding, there is an impetus to estimate treatment effects from non-randomised studies for health technology assessment (HTA).

Methods: We critically discuss the challenges in the wider use of non-randomised studies in informing reimbursement and/or pricing decisions and offer recommendations to support the better use of such data by HTA bodies. This is informed by published reports on the use of non-randomised studies in HTA and regulation, empirical studies of different methodologies for estimating treatment effects from non-randomised studies, including the work undertaking by the IMPACT-HTA consortium, and a project workshop involving representatives from several HTA agencies.

Results: Although HTA bodies retain a strong preference for randomised evidence, reimbursement and/or pricing decisions are increasingly reliant on non-randomised evidence. The overall quality of non-randomised studies submitted to HTA bodies is, however, poor. Although empirical investigations of non-randomised studies in comparison to RCTs have found mixed evidence about the extent of bias and the circumstances in which bias is minimised, there is some evidence that high-quality studies are at lower risk of bias than low-quality studies, and that the risk of bias is particularly high for studies using external controls.

Conclusions: To ensure HTA processes remain rigorous and robust, HTA bodies should demand clear, extensive, and structured reporting of non-randomised studies. This should include an in-depth assessment of the risk of bias, and HTA bodies should issue clear guidelines for study conduct in order to support such assessments. HTA bodies should, in recognition of the additional uncertainty imparted by non-randomised designs in estimates of clinical effectiveness, strengthen early scientific advice offered to mitigate such uncertainty. Furthermore, managed entry agreements to ensure high-quality and relevant evidence is generated could facilitate patient access to treatment with high uncertainty. This will require HTA bodies to ensure that staff are equipped with the requisite skills to assess, and possibly even design, non-randomised studies.

Introduction

Most health technology assessment (HTA) bodies profess a strong preference for evidence on treatment effects from randomised controlled trials (RCTs) over non-randomised studies due to a lower risk of bias by design [1]. However, with an increasing number of medicines receiving regulatory approval based on non-randomised evidence, particularly in oncology and rare diseases, HTA bodies are being tasked with issuing recommendations on new technologies in the absence of RCTs or with limited RCT data, and even commissioning and designing non-randomised studies as part of managed entry agreements [2–7]. For other health technologies like medical devices, non-randomised studies already provide the predominant source of evidence [8]. Non-randomised studies are defined broadly here to include all study designs without randomisation, including non-randomised clinical trials, observational studies, and trials with external controls [9].

The quality of non-randomised studies so far accepted by HTA bodies for decision-making has been variable. Most non-randomised studies have used simple analytical methods which are unlikely to fully address the manifold risks of bias [2–5,10,11], and there is a widely acknowledged lack of transparency in governance and reporting of possible sources of bias in such studies [12]. Clear evidence-based guidance on the design, analysis, and interpretation of non-randomised studies and robust mechanisms of enforcement are needed to ensure the integrity of reimbursement and pricing decisions [13].

As part of the IMPACT-HTA project (work package 6), we undertook a meta-epidemiological study to compare estimates of comparative effectiveness from RCTs and non-randomised studies addressing the same clinical questions [14]. Based on our findings and those from other comparative methodology studies, as well as published reports on the use of non-randomised studies in HTA and regulatory decision making, we critically discuss the challenges in the use of non-randomised studies by HTA bodies and identify key considerations to support the better use of such data.

The use of non-randomised evidence to estimate treatment effects

RCTs are widely recognised as the gold standard for estimating treatment efficacy because of the power of randomisation in ensuring any differences in baseline characteristics between groups are due to chance, blinding (where applied) in preventing knowledge of treatment allocation from influencing behaviours, and standardised protocols in ensuring consistent data collection [15]. RCTs are not, however, immune to bias. More than half of the pivotal studies in oncology submitted to the European Medicines Agency (EMA) between 2014-2016 were judged to be at high risk of bias for their primary outcome [4]. Further, many well-conducted RCTs have limitations with regard to establishing comparative effectiveness for HTA including comparisons to treatments (or placebo) other than usual care, selected patient populations, insufficient follow-up, treatment protocols that deviate from usual care, and the use of surrogate outcomes which may not predict the outcome(s) of ultimate interest [16–19]. Against this background, there is at least the potential that well conducted non-randomised studies based on high-quality, relevant (e.g. inclusion of patient relevant outcomes), and representative data could support decision making.

In addition, RCTs may be considered unethical (e.g. last line cancer therapies) or infeasible (e.g. due to small numbers of eligible patients) [20]. There has been an increase in the use of expedited access programmes by regulatory authorities to facilitate faster patient access to innovative treatments, particularly in oncology and rare diseases [6,21]. Around 80% of new medicines in the US are approved through such schemes [22]. These schemes are associated with the submission of lower quality evidence, including widespread use of single-arm trials [5,18]. It has been argued that many of these studies recruited enough patients to support randomisation [18]. Following approval through expedited access programmes, HTA bodies are required to make initial reimbursement and/or pricing decisions based on this limited evidence, sometimes as part of managed entry agreements in which further data collection is mandated, whether in the form of further trials or observational data (usually registries), to support reassessment at a later date [23–25]. Therefore, methods for estimating the relative effectiveness of treatments using non-randomised data are of interest to HTA bodies.

Non-randomised studies are typically at a higher risk of bias than RCTs because patients are not randomly allocated to treatment and treatment allocation is known. Instead, physicians decide on a patient's treatment based on their expectation of the benefit-risk profile of different treatments for

that patient and their preferences. As such, any observed differences in outcomes between patients receiving different treatments may be due to patient characteristics and not the treatment alone (i.e. confounding by indication) [26,27]. Bias may also arise from poor data quality including errors in data entry, measurement error, misclassification of exposures and outcomes, missing data, or from poor analytical choices, coding errors, or selective reporting and/or publication of results [28,29].

Despite the increasing availability of observational data and advances in analytical methods, RCTs remain the main source of evidence on treatment effects for medicines [3,30–33]. An evaluation of evidence submissions of medicines to international HTA bodies typically find non-randomised evidence on effectiveness to be used in around 5% of submissions, with some variation by country and disease areas [3,30–32]. Most submissions used evidence from single-arm trials and were predominantly in oncology, infectious diseases, and for orphan drugs. Although the proportion of submissions using non-randomised evidence remains low, it is increasing: for instance, more than half of all NICE's medicines appraisals between 2010 and 2016 using non-randomised data on clinical effectiveness were in 2015-16 alone [3]. For other types of health technologies like medical devices, there has long been a greater reliance on non-randomised studies to estimate treatment effects as regulatory requirements do not generally demand RCTs and RCTs may be difficult to perform [8,18,34].

Despite the pervasive risk of bias from confounding in non-randomised studies, many HTA bodies and regulators have accepted submissions using no or only simple methods of adjustment [2–5,10,11]. Anderson et al. [3] found that of the 22 medicines appraised by NICE using non-randomised data between 2010 and 2016, only five used a regression approach to adjust for confounding while two-thirds used naïve unadjusted indirect comparison to aggregate data. Despite these limitations, there was no evidence that the chance of a positive recommendation differed in these submissions compared to those relying on RCTs, though this comparison could be confounded by other factors. A persistent challenge is that for clinical questions where RCTs cannot be performed due to small patient numbers, there is also likely to be limited observational data, inhibiting the extent of adjustment possible in analyses.

Empirical assessments of non-randomised studies vs RCTs

In this section, we summarise some of the literature examining the internal validity of different types of non-randomised studies and analytical methods.

Meta-epidemiological studies explore the impact of study design characteristics on estimated treatment effects [35]. While most commonly used to understand the key design characteristics of RCTs, several studies have compared estimates of relative treatment effects from RCTs and non-randomised studies. A 2017 pragmatic review of fourteen meta-epidemiological studies reported that seven found no systematic differences between RCTs and non-randomised studies, five found estimates from non-randomised studies to systematically exceed those from RCTs, and the remaining two were inconclusive [36]. A 2014 systematic review found overall no systematic difference in treatment effects but substantial variation in these differences for specific clinical questions [37].

In IMPACT-HTA work package 6 we undertook a novel meta-epidemiological study containing 2,500 studies of pharmacological interventions across 350 clinical topics [14]. While no evidence of a systematic difference in estimates of treatment effects between RCTs and non-randomised studies was found (relative odds ratio = 0.96, 95% CI: 0.89-1.03), this masked substantial variation across clinical topics. In 38% of clinical topics conflicting conclusions were reached about treatment effectiveness, and 38% of clinical topics had treatment effects that differed by a factor of 2 or greater; for 16% of topics, differences were statistically significant. Estimates from experimental non-randomised studies (e.g. quasi-randomised or non-randomised trials) were on average 21% more favourable compared to RCTs, while no significant difference was found for observational studies.

A large number of studies have attempted to replicate RCT evidence using non-randomised studies with mixed success [20]. Previous meta-epidemiological and other studies have found that higher-quality, prospective non-randomised studies and RCTs produce estimates more similar to each other [36]. Further prospective empirical work is required to assess the performance of observational studies against as-yet unreported RCTs and identify those characteristics of non-randomised studies that lead to more reliable estimates [20,38].

Where confounding is present, appropriate adjustment for confounders does ameliorate bias to some extent, but may not always fully eliminate it, even where confounders are known and observed [39]. There are numerous methods available to reduce the effects of confounding as described in the preceding section, and it is important to understand the relative performance of different approaches. In its meta-epidemiological study, the IMPACT-HTA consortium compared RCTs with non-randomised studies by method of adjustment, namely covariate adjustment (excluding those using propensity scores), use of propensity scores (for adjustment or matching), or other matching, but found information on analytical methods to be often lacking. The average difference in treatment effect between randomised and non-randomised studies did not differ according to the methods used [14]. Many other studies have compared variants of these techniques for specific clinical questions with mixed results. Overall, there does not appear to be strong evidence that matching and propensity score methods reliably produce less biased estimates than traditional covariate adjustment [37,40,41]. More recent methods like the use of negative controls or falsification endpoints, i.e. assessment of outcomes where no association with the intervention is expected, as qualitative evidence for or against residual confounding or to formally calibrate the treatment effect have shown promise [42,43].

There is stronger evidence that non-randomised studies relying on external controls, such as single-arm trials, are associated with greater bias on average than other types of non-randomised studies [36]. It can be challenging to identify an appropriate control group, and external control data may differ in terms of patient characteristics, diagnostic criteria, disease severity, treatment patterns, and outcome ascertainment, particularly when using historical controls. Moreover, there is good evidence that external control groups have worse outcomes than control groups in RCTs [44,45]. In some disease areas, there is no defined treatment standard against which to compare a novel treatment, and treatment patterns are frequently changing. External controls are most useful where the treatment effect is large and the condition is well characterised and predictable [45,46]. However, there is a lack of consensus of what constitutes a large treatment effect and even modest variations in outcomes in the controls inflates the false positive rate [36,39]. There is limited guidance on best practice methods when using external control data but methods based on adjustment using individual patient data, are preferred to naïve adjustment methods or population adjustment using aggregate data [45,47,48].

Considerations for the use of non-randomised studies in comparative effectiveness estimation

Based on the results of the IMPACT HTA meta-epidemiological study and other comparative methodology studies, as well as published reports on the use of non-randomised studies in HTA and regulatory decision making, we discuss some of the key considerations for those undertaking non-randomised studies, usually study sponsors, and for HTA bodies and payers to ensure the most appropriate use of such evidence in decision-making. An initial set of considerations was discussed at a project workshop in June 2020 with representatives of various HTA bodies and IMPACT consortium members (see acknowledgements). The views expressed in this manuscript do not necessarily represent the views of all participants, nor the agencies they represent.

Conduct of non-randomised studies

1. *Planning and design*

    1.1. *Justify the need for a non-randomised study and demonstrate that the research question is amenable to being answered using non-randomised data*

Non-randomised studies may be required in the absence of sufficient or robust RCT data or to complement evidence from RCTs, particularly in the post-marketing phase [20,29,49]. All research questions should be clearly defined, for instance using the PICOTS (population, intervention, comparators, outcomes, timing, and setting) and estimand frameworks (an appendix to ICH-E9) [50,51]. In most cases a control group will be required to demonstrate comparative effectiveness [45]. Where observational data is considered for use as part of a non-randomised study, it is important to ascertain that: 1) the question is amenable to being answered using observational data (e.g. studies based on hard clinical endpoints are easier to conduct than studies based on changes in severity of a chronic disease), because only a subset are [52,53], and; 2) that one or more data sources containing sufficient information (e.g. on patients, exposures, objective clinical outcomes, controls, and confounding factors) of high-quality and relevant to the decision context are available and accessible [29,44,54]. Datasets should be identified through a systematic, transparent, and reproducible process to ensure the most appropriate data is used [55,56]. This avoids the selection of datasets based on convenience or the knowledge or expectation of deriving particular results.

1.2. *Prospectively plan studies and engage with early scientific advice procedures*

Non-randomised studies should be planned prospectively to negate the possibility of selective methodology and results [12,55,57]. Many HTA bodies, separately or in collaboration with regulatory bodies, offer scientific advice to sponsors to help guide evidence generation throughout a product's lifecycle [58–60]. The use of non-randomised evidence should be discussed early in a product's development where sponsors consider RCTs to be infeasible, unethical, or otherwise contraindicated, or for post-launch evidence generation to resolve outstanding uncertainties or extend access [7]. Scientific advice helps ensure that evidence generation plans will provide data of relevance to HTA bodies and regulators, potentially increasing the chance of successful marketing authorisation and reimbursement [61].

2. *Analysis*

2.1. *Understand potential risks of bias and address using appropriate analytical strategies*

Bias may arise in non-randomised studies for numerous reasons including patient selection, confounding by indication, or poor quality data [40]. The potential mechanisms of bias for any application should be clearly articulated, and the analysis plan designed to elucidate and minimise potential bias.

NICE's decision support unit developed a flowchart to help guide the selection of an appropriate analytical method where individual patient level data is available based on the treatment effect of interest and whether confounding is expected to be on observable or unobservable characteristics [10]. Adjustment using propensity score matching is widely recommended to adjust for differences in observed characteristics [62,63]. Though there is limited evidence that it providers better estimates than traditional covariate adjustment [37,40,41,64] the ability to explicitly assess the balance of observed covariates between treatment groups provides greater transparency [39].

Single arm trials do not, by construction, provide evidence on comparative effectiveness, and instead comparisons must be made with external data which introduces additional risks of bias [36,45]. Where possible high-quality individual patient level data (rather than aggregate data) from contemporaneous (or, where necessary, historical) controls should be used for adjustment as it provides greater scope to control for differences between patients [65]. Potential confounders should typically be defined prior

to study conduct and based on the scientific literature and through engagement with clinical experts [66]. In some situations, large-scale propensity score methods can be used without full pre-specification of covariates [67].

Analysts and reviewers must be aware that methods of adjustment based on observed covariates remain at risk of residual bias, whether because of unobserved confounders or poor measurement of observed confounders (e.g. adjustment for smoking status may not capture differential risks by smoking intensity) [68]. Instrumental variable regression has been suggested as an approach to provide causal estimates in the presence of unobserved confounding by identifying and controlling for an exogenous source of variation strongly related to the exposure and not the outcome except through the exposure (e.g. between provider variation in prescribing preference) [69]. However, identifying sufficiently strong instruments is difficult in practice and the practical usefulness of these methods has been contested [70].

Some other risks of bias can be addressed by trying to replicate RCT designs [44,63,71]. Such so-called 'target trials' form the basis of prominent risk of bias tools [72]. They involve numerous design principles but key among these are restricting cohorts to new users, using active comparators (as opposed to non-users), adjusting based on pre-treatment confounders (e.g. using propensity score matching), performing or utilising outcome validation studies, using biologically informed exposure effect windows and induction periods informed by biology and the reality of healthcare delivery, performing on-treatment and intention-to-treat analyses, and prespecifying sensitivity analyses. In practice this can be difficult to achieve given the limitations of many observational datasets [66]. In addition, for novel therapies and for certain conditions, it may not be possible to identify appropriate active comparators, while early users of a technology may differ from the target population [73]. Further empirical evidence is still required to fully demonstrate the value of these approaches [38].

2.2. Perform extensive sensitivity analyses

Non-randomised studies involve many decisions and assumptions including in data curation and analysis, each of which, alone or in combination, could have substantial effects on the resulting estimates. It is therefore essential that extensive and pre-specified sensitivity analyses are undertaken to understand the robustness of the results to these assumptions and characterise the uncertainty in the treatment effect [71]. Of central importance is understanding the impact of alternative estimation

strategies to control for differences between treatment groups, sometimes known as triangulation [74]. This could be complemented by quantitative bias analysis which includes techniques such as negative controls, the use of external information, and threshold analysis, i.e. identifying the extent of bias sufficient to change decisions [75–77].

Because non-randomised studies are at a high risk of bias from multiple avenues, an estimate of comparative effectiveness from a single data source is subject to considerable uncertainty [39]. Studies should ideally be performed in more than one dataset [57]. The ability to replicate results in different datasets may increase confidence in the validity of the results, though it should be recognised that treatment channelling and other data limitations may lead to a systematic bias across datasets. The inability to replicate findings in other datasets would require further explanation.

3. Reporting

3.1. Register protocols before study conduct

Detailed study protocols including statistical analysis plans should be registered before the beginning of the study on publicly accessible platforms using structured reporting templates [12,40,51]. This would improve study transparency and allay concerns about selective analyses and selective reporting as well as publication bias, which are major impediments to the acceptability and wider use of non-randomised evidence [12,57,78,79].

3.2. Report data, methods, and results transparently

Reporting checklists play an important role in the transparent reporting of non-randomised study methods and results by ensuring that key information is reported [40,80–82]. While they should be used in evidence submissions, they are not generally sufficient to support reproduction [62] and are not an indicator of quality. Ideally reviewers of submitted evidence, including HTA bodies or independent review groups, would also have access to the data and analytical code to ensure the replicability of the submitted results and assess the impact of alternative analytical decisions or data on the resulting estimate(s). However, there remain substantial governance, technical, and practical challenges to sharing data, including a lack of in-house expertise in many HTA agencies [82].

Poor quality data and the absence of a standard protocol for data collection may impart bias into analyses [29]. Data quality should be considered at the level of the data source as well as in relation to

each application, and this should be clearly reported. There is no universal consensus for defining data quality but it is commonly characterised in terms of data completeness, validity, consistency, timeliness, and accuracy [83,84]. Dimensions of data quality can be summarised quantitively, e.g. percentages of missing data, sensitivity and specificity for binary variables, and mean error for continuous variables [44]. A number of groups have designed data quality assessment tools that seek to provide vast amounts of information on data quality across quality domains, and further work is ongoing in this area [83]. Tools can also be used to ensure appropriate data governance is in place [85].

Validated software, for instance that developed through the open-source OHDSI community [83], support the transparent and comprehensive reporting of analyses, reduce the risk of coding errors, and impose good analytical practices. Along with auditing trails that ensure a correct ordering and comprehensive account of data preparation and analysis, they have been proposed as a complement to pre-registration [29,71,78,86].

3.3. Describe potential biases and report the overall risk of bias

Study sponsors should clearly articulate potential causes of bias and their impact on estimated treatment effects. The overall risk of bias should also be formally assessed using well-validated checklists [72,87]. For non-randomised studies, the ROBINS-I tool is recommended by the European network for HTA (EUnetHTA) and assesses the risk of bias by specifying the research question as a target trial and considering risks from the seven domains of bias namely bias due to confounding, in selection of participants into the study, in classification of interventions, due to deviations from intended interventions, due to missing data, in measurement of outcomes, and in the selection of the reported result [72]. NICE's decision support unit also recommends the Queens checklist to assess the risk of bias pertaining to the specific analytical methods used [10].

3.4. Convey and ideally quantify the uncertainty

Uncertainty is pervasive in HTA, particularly where estimates of effects and costs are required over the long-term [88]. Although several meta-epidemiological studies have found no evidence of systematic differences in treatment effect estimates between RCTs and non-randomised studies, there was great variation in estimates across clinical questions, and few clear predictors of bias [14,36]. Unless a decision maker is able to identify when differences are likely to be present and the direction and magnitude of any differences, then they must acknowledge the sizeable additional uncertainty involved in the use

of non-randomised evidence which will not be fully captured by the statistical uncertainty in the estimated effectiveness parameter: some studies have suggested the true uncertainty may be as much as five to ten times greater [39]. Sensitivity analysis including quantitative bias analysis and triangulation by study design and data set can support a deeper understanding of the uncertainty in the estimated treatment effects. This uncertainty should be appropriately conveyed, and ideally quantified, in evidence submissions [89–91].

Further considerations for HTA bodies

4. Strengthening systems

4.1. Strengthen and standardise scientific advice procedures

Scientific advice provides an essential function in allowing HTA bodies and regulators to work with stakeholders to ensure that plans for evidence generation deliver information that supports and improves decision-making in both pre-market and post-market settings [7,18,19]. Most scientific advice to date has related to the generation of evidence from trials. As scientific advice committees are asked to provide more guidance to design non-randomised studies there needs to be clear and consistent guidelines in place to ensure that the evidence generated can support decision making across jurisdictions [58,59,92].

4.2. Strengthen conditional reimbursement processes to ensure generating of further informative evidence after initial reimbursement decisions

Uncertainty poses the risk that incorrect decisions will be made that are detrimental to population health. Managed entry agreements of various forms have the potential to ameliorate the impact of uncertainty and are widely used throughout Europe [21,24,25,93]. Following the taxonomy of Kanavos et al. [25], managed entry agreements include financial schemes that seek to cap expenditure on a new treatment at the population or individual level using policy tools such as discounts and price-volume agreements, and outcomes based schemes that either link reimbursement to observed patient or health system outcomes or seek to reduce uncertainty through the collection of additional data, e.g. coverage with evidence development. Such schemes have been widely used across Europe [7,25,94,95].

A lack of transparency around such schemes means there remains an incomplete understanding of their strengths and limitations [25,94]. A study of the conditional reimbursement scheme used in the

Netherlands between 2006 and 2012 found that the success of the scheme was compromised by several procedural and methodological limitations with poor quality studies commissioned and limited impact on decision making [96]. Similar issues have been experienced by other HTA bodies [7,95,97]. The experience of regulators with post-marketing studies shows long delays or even failure to conduct such studies, even in the presence of theoretically strong incentives for pharmaceutical companies to comply with mandatory post-marketing obligations, such as financial penalties or threats to revoke marketing authorisations [22,98–100]. Further empirical evidence is needed on the net benefits of different types of managed entry agreements and their ability to generate meaningful evidence [25].

These experiences underscore the importance of having high-quality evidence available at the time of the initial approval. When such schemes are used, it is imperative that clear responsibilities for data collection and analysis are defined and that appropriate enforcement mechanisms are available to HTA bodies to ensure timely delivery of high-quality evidence. Data collection and analysis should follow best practice guidance and be reported transparently.

4.3. Invest in and develop staff skills in the design, analysis, and interpretation of non-randomised studies

As an increasing number of HTA submissions are made on the basis of non-randomised data and as HTA bodies respond by commissioning and designing studies, and perhaps even analysing data, it is imperative that their staff and decision-making committees possess the requisite skills to conduct this work [10,49]. This is likely to require both training and recruitment.

5. Issuing and enforcing best practice guidance

To ensure the generation of high-quality evidence suitable for decision-making, HTA should issue clear guidance on data quality standards and best practice methods for the design, conduct, and reporting of non-randomised studies, and ensure that these are followed. An example of this is the ReQuest tool developed by EUnetHTA [85], which ensures a minimum level of quality control for registries and registry studies used in managed entry agreements. Gliklich and Leavy [101] have called for a unified set of quality criteria for *real-world* data sources drawing on the experience of registries.

There are several ongoing initiatives to establish frameworks for the use of real-world evidence in decision making including by the EMA and NICE [102,103]. It is essential that these tools are built in a

collaborative fashion to streamline evidence generation and ensure adoption. They should aspire to be simple to apply and interpret so as to enhance transparency and reduce the burden on decision makers and payers. This would also benefit smaller HTA bodies and payers who have fewer resources to conduct detailed assessments themselves. This work could also include clear guidance as to when non-randomised studies will be considered and where they are expected. Because this is an evolving field, it is important that these frameworks are regularly revised.

6. Supporting future research and initiatives

6.1. Support initiatives that seek to enhance the validity, availability, and usability of routine healthcare data

While there remain many legitimate concerns about the growing use of non-randomised evidence in healthcare decision making, recent years have seen advancements in our collective understanding about what constitutes reliable evidence. Impediments to the greater use of observational data in research include a lack of data interoperability, data fragmentation, data governance processes, and concerns about data quality and risk of bias [79]. HTA bodies should encourage and support initiatives in all these areas [19]. This could include developing appropriate registries and processes for pre-registration of study protocols [12,40], developing and maintaining comprehensive and detailed registries of observational datasets including substantial meta-data [104,105], supporting data linkages and validation exercises, common data models to enhance data interoperability [106], definition of core datasets, data collection standards [54,101], distributed data networks [107], and the development of tools to understand data and evidence quality [102,103].

6.2. Support further investigations into best practice methodology and empirical assessments of comparative methods performance

While there is a growing consensus about the use of the target trial approach to effectiveness estimation in non-randomised studies, there remains a lack of high-quality prospective evidence to support it. Similarly, there is limited understanding about which of these decision features or which analytical strategies contribute most to alleviating bias. Some research in this area is ongoing [44]. HTA bodies should support such work and encourage further research.

In this report, we concentrated predominantly on the threat to internal validity posed by confounding by indication, which is the most common challenge and the focus of much research. However internal validity may also be compromised by other factors including appropriately defining time-zero for non-active comparators, treatment multiplicity, sequential treatment decisions, intercurrent events (e.g. treatment switching), and time-varying confounding [50,108]. Further work is required to better understand such risks and methods to mitigate them [64].

Finally, we have focused on the conduct and interpretation of individual non-randomised studies, but multiple such studies may exist and HTA bodies are interested in the totality of evidence. Further methodological research is required to identify optimal approaches of combining evidence from randomised and non-randomised studies, particularly when there are unanchored comparisons [109].

Conclusions

While most HTA bodies have a strong preference for evidence on treatment effects to be derived from RCTs, they are increasingly being asked to make reimbursement and/or pricing decisions based on non-randomised studies. These studies are at higher risk of bias than RCTs meaning estimates of clinical effectiveness are often highly uncertain. In these situations, rigorous and extensive processes should be followed to ensure that evidence derived from non-randomised studies is of high-quality and those conducting such studies adhere to best practices, including the use of high-quality data, addressing confounding using appropriate methods, and transparency in study design, conduct, analysis, and reporting. Even with high-quality research, HTA bodies should recognise the uncertainty inherent in non-randomised studies and establish robust mechanisms to mitigate the risks for population health thereby imposed.

References

1. Makady A, Ham R ten, de Boer A, Hillege H, Klungel O, Goettsch W. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. *Value Heal*. 2017;20(4):520-532. doi:10.1016/j.jval.2016.12.003
2. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: Analysis of EMA and FDA approvals 1999-2014. *BMJ Open*. 2016;6(6):e011666. doi:10.1136/bmjopen-2016-011666
3. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. *J Clin Epidemiol*. 2019;105:50-59. doi:10.1016/j.jclinepi.2018.09.003
4. Naci H, Davis C, Savović J, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. *BMJ*. 2019;366:l5221. doi:10.1136/bmj.l5221
5. Goring S, Taylor A, Müller K, et al. Characteristics of non-randomised studies using comparisons with external controls submitted for regulatory approval in the USA and Europe: A systematic review. *BMJ Open*. 2019;9(2):e024895. doi:10.1136/bmjopen-2018-024895
6. Bouvy JC, Sapede C, Garner S. Managed entry agreements for pharmaceuticals in the context of adaptive pathways in Europe. *Front Pharmacol*. 2018;9:280. doi:10.3389/fphar.2018.00280
7. Moseley J, Vamvakas S, Berntgen M, et al. Regulatory and health technology assessment advice on postlicensing and postlaunch evidence generation is a foundation for lifecycle data collection for medicines. *Br J Clin Pharmacol*. 2020;86(6):1034-1051. doi:10.1111/bcp.14279
8. Crispi F, Naci H, Barkauskaite E, Osipenko L, Mossialos E. Assessment of Devices, Diagnostics and Digital Technologies: A Review of NICE Medical Technologies Guidance. *Appl Health Econ Health Policy*. 2019;17(2):189-211. doi:10.1007/s40258-018-0438-y
9. Grimes DA, Schulz KF. An overview of clinical research: The lay of the land. *Lancet*. 2002;359(9300):57-61. doi:10.1016/S0140-6736(02)07283-5
10. NICE Decision Support Unit. *NICE DSU Technical Support Document 17: The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data*.; 2015. http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf
11. Hatswell AJ, Freemantle N, Baio G. Economic Evaluations of Pharmaceuticals Granted a Marketing Authorisation Without the Results of Randomised Trials: A Systematic Review and Taxonomy. *Pharmacoeconomics*. 2017;35(2):163-176. doi:10.1007/s40273-016-0460-6
12. ISPOR. *Improving Transparency in Non-Interventional Research for Hypothesis Testing—WHY, WHAT, and HOW: Considerations from The Real-World Evidence Transparency Initiative (Draft White Paper)*.; 2019. https://www.ispor.org/docs/default-source/strategic-initiatives/improving-transparency-in-non-interventional-research-for-hypothesis-testing_final.pdf?sfvrsn=77fb4e97_6
13. Bell, H, Wailoo, A J, Hernandez, M, Grieve, R, Faria, R, Gibson, L, Grimm S. *The Use of Real World Data for the Estimation of Treatment Effects in NICE Decision Making*.; 2016. http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf
14. IMPACT-HTA Work Package 6. https://www.impact-hta.eu/work-package-6
15. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med*. 2020;382(7):674-678. doi:10.1056/NEJMsb1901642

16. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;15(7):677-687. doi:10.1002/hec.1093
17. 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 Heal*. 2007;10(5):326-335. doi:10.1111/j.1524-4733.2007.00186.x
18. Naci H, Salcher-Konrad M, Kesselheim AS, et al. Generating comparative evidence on new drugs and devices before approval. *Lancet*. 2020;395(10228):986-997. doi:10.1016/S0140-6736(19)33178-2
19. Cipriani A, Ioannidis JPA, Rothwell PM, et al. Generating comparative evidence on new drugs and devices after approval. *Lancet*. 2020;395(10228):998-1010. doi:10.1016/S0140-6736(19)33177-0
20. Eichler HG, Koenig F, Arlett P, et al. Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation. *Clin Pharmacol Ther*. 2019;107(4):773-779. doi:10.1002/cpt.1638
21. Husereau D, Henshall C, Jivraj J. Adaptive approaches to licensing, health technology assessment, and introduction of drugs and devices. *Int J Technol Assess Health Care*. 2014;30(3):241-249. doi:10.1017/S0266462314000191
22. Darrow JJ, Avorn J, Kesselheim AS. FDA Approval and Regulation of Pharmaceuticals, 1983-2018. *JAMA - J Am Med Assoc*. 2020;323(2):164-176. doi:10.1001/jama.2019.20288
23. Nicod E, Annemans L, Bucsics A, Lee A, Upadhyaya S, Facey K. HTA programme response to the challenges of dealing with orphan medicinal products: Process evaluation in selected European countries. *Health Policy (New York)*. 2019;123(2):140-151. doi:10.1016/j.healthpol.2017.03.009
24. Bruegger U. *A Review of Coverage With Evidence Development (CED) in Different Countries: What Works and What Doesn't*.; 2014. https://htai.org/wp-content/uploads/2018/02/CED_Report_Bruegger_Final_Version.pdf
25. Kanavos P, Ferrario A, Tafuri G, Siviero P. Managing Risk and Uncertainty in Health Technology Introduction: The Role of Managed Entry Agreements. *Glob Policy*. 2017;8:84-92. doi:10.1111/1758-5899.12386
26. Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in establishing efficacy. *Lancet*. 2019;393(10168):210-211. doi:10.1016/S0140-6736(18)32840-X
27. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359(9302):248-252. doi:10.1016/S0140-6736(02)07451-2
28. Bowrin K, Briere JB, Levy P, Millier A, Clay E, Toumi M. Cost-effectiveness analyses using real-world data: an overview of the literature. *J Med Econ*. 2019;22(6):545-553. doi:10.1080/13696998.2019.1588737
29. Asche C V., Seal B, Kahler KH, Oehrlein EM, Baumgartner MG. Evaluation of Healthcare Interventions and Big Data: Review of Associated Data Issues. *Pharmacoeconomics*. 2017;35(8):759-765. doi:10.1007/s40273-017-0513-5
30. Makady A, van Veelen A, Jonsson P, et al. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. *Pharmacoeconomics*. 2018;36(3):359-368. doi:10.1007/s40273-017-0596-z
31. Griffiths EA, Macaulay R, Vadlamudi NK, Uddin J, Samuels ER. The Role of Noncomparative Evidence in Health Technology Assessment Decisions. *Value Heal*. 2017;20(10):1245-1251. doi:10.1016/j.jval.2017.06.015
32. Jao R, Jaksa A, Pontynen A, Wang X. Health Techonolgy Assessment (HTA) Agencies Consideration of Real World Evidence (RWE). *Value Heal*. 2018;21:S7.

33. Bullement A, Podkonjak T, Robinson MJ, et al. Real-world evidence use in assessments of cancer drugs by NICE. *Int J Technol Assess Health Care*. 2020;36(4):388-394. https://doi.org/10.1017/S0266462320000434
34. George E. How real-world data compensate for scarce evidence in HTA. *Z Evid Fortbild Qual Gesundhwes*. 2016;112:S23-S26. doi:10.1016/j.zefq.2016.04.012
35. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med*. 2017;22(4):139-142. doi:10.1136/ebmed-2017-110713
36. Woolacott N, Corbett M, Jones-Diette J, Hodgson R. Methodological challenges for the evaluation of clinical effectiveness in the context of accelerated regulatory approval: an overview. *J Clin Epidemiol*. 2017;90:108-118. doi:10.1016/j.jclinepi.2017.07.002
37. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014;2014(4). doi:10.1002/14651858.MR000034.pub2
38. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther*. 2017;102(6):924-933. doi:10.1002/cpt.857
39. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F. Evaluating non-randomised intervention studies. *Health Technol Assess (Rockv)*. 2003;7(27):1-186. doi:96-26-99 [pii]
40. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). *The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology (Revision 8)*.; 2020. http://www.encepp.eu/standards_and_guidances/documents/GuideMethodRev8.pdf
41. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol*. 2004;57(12):1223-1231. doi:10.1016/j.jclinepi.2004.03.011
42. Prasad V, Jena AB. Prespecified falsification end points: Can they validate true observational associations? *JAMA - J Am Med Assoc*. 2013;309(3):241-242. doi:10.1001/jama.2012.96867
43. Schuemie MJ, Ryan PB, Dumouchel W, Suchard MA, Madigan D. Interpreting observational studies: Why empirical calibration is needed to correct p-values. *Stat Med*. 2014;33(2):209-218. doi:10.1002/sim.5925
44. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther*. 2019;105(4):867-877. doi:10.1002/cpt.1351
45. International Conference on Harmonization. *ICH Topic E10—Choice of Control Group in Clinical Trials. Note for Guidance on Choice of Control Group in Clinical Trials (CPMP/ICH/364/96)*.; 2000. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf
46. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. *Br Med J*. 2003;327(7429):1459-1461. doi:10.1177/154510970400300401
47. Cucherat M, Laporte S, Delaitre O, et al. From single-arm studies to externally controlled studies. Methodological considerations and guidelines. *Therapies*. 2019;75(1):21-27. doi:10.1016/j.therap.2019.11.007
48. Ghadessi M, Tang R, Zhou J, et al. A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). *Orphanet J Rare Dis*. 2020;15(1):69. doi:10.1186/s13023-020-1332-x

49. HTAi Global Policy Forum. *Real-World Evidence in the Context of Health Technology Assessment Processes – from Theory to Action*.; 2018. https://htai.org/wp-content/uploads/2019/02/HTAiGlobalPolicyForum2019_BackgroundPaper.pdf
50. ICH. Addendum on Estimands and Sensitivity Analysis in Clinical Trials To the Guideline on Statistical Principles for Clinical Trials E9(R1). *Int Conf Harmon*. Published online 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
51. Agency for Healthcare Research and Quality. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*.; 2013. https://effectivehealthcare.ahrq.gov/products/observational-cer-protocol/research
52. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Netw open*. 2019;2(10):e1912869. doi:10.1001/jamanetworkopen.2019.12869
53. Fralick M, Bartsch E, Darrow JJ, Kesselheim AS. Understanding when real world data can be used to replicate a clinical trial: A cross-sectional study of medications approved in 2011. *Pharmacoepidemiol Drug Saf*. Published online 2020:1-6. doi:10.1002/pds.5086
54. Duke-Margolis Center. *A Framework for Regulatory Use of Real-World Evidence*.; 2017. https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf
55. Eichler HG, Bloechl-Daum B, Bauer P, et al. "Threshold-crossing": A Useful Way to Establish the Counterfactual in Clinical Trials? *Clin Pharmacol Ther*. 2016;100(6):699-712. doi:10.1002/cpt.515
56. 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*. 2019;106(1):103-115. doi:10.1002/cpt.1480
57. 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. *Value Heal*. 2017;20(8):1003-1008. doi:10.1016/j.jval.2017.08.3019
58. Tafuri G, Pagnini M, Moseley J, et al. How aligned are the perspectives of EU regulators and HTA bodies? A comparative analysis of regulatory-HTA parallel scientific advice. *Br J Clin Pharmacol*. 2016;82:965-973. doi:10.1111/bcp.13023
59. Tafuri G, Lucas I, Estevão S, et al. The impact of parallel regulatory–health technology assessment scientific advice on clinical development. Assessing the uptake of regulatory and health technology assessment recommendations. *Br J Clin Pharmacol*. 2018;84(5):1013-1019. doi:10.1111/bcp.13524
60. Maignen F, Osipenko L, Pinilla-Dominguez P, Crowe E. Integrating health technology assessment requirements in the clinical development of medicines: the experience from NICE scientific advice. *Eur J Clin Pharmacol*. 2017;73(3):297-305. doi:10.1007/s00228-016-2174-2
61. Ng T, Ziomek J, Delaitre-Bonnin C, Bending MW. PHP276 - A REVIEW OF THE IMPACT OF INTEGRATED SCIENTIFIC ADVICE FOR THE OPTIMISATION OF EVIDENCE GENERATION FOR HTA APPRAISALS. *Value Heal*. 2018;21:S197. doi:10.1016/j.jval.2018.09.1170
62. Schneeweiss S, Glynn RJ. Real-world data analytics fit for regulatory decision-making. *Am J Law Med*. 2018;44(2-3):197-216. doi:10.1177/0098858818789429
63. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254

64. Ewald H, Ioannidis JPA, Ladanie A, Mc Cord K, Bucher HC, Hemkens LG. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. *J Clin Epidemiol*. 2020;118:29-41. doi:10.1016/j.jclinepi.2019.10.012
65. European Medicines Agency. *Workshop on Single-Arm Trials in Oncology*.; 2016. https://www.ema.europa.eu/en/events/workshop-single-arm-trials-oncology
66. Institute for Quality and Efficiency in Health Care. *[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*.; 2020. https://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
67. Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf*. 2012;21(S1):41-49. doi:10.1002/pds.2328
68. Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: A simulation study. *Am J Epidemiol*. 2007;166(6):646-655. doi:10.1093/aje/kwm165
69. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf*. 2010;19(6):537-554. doi:10.1002/pds.1908
70. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med*. 2014;161(2):131-138. doi:10.7326/M13-1887
71. Schneeweiss S. Real-World Evidence of Treatment Effects: The Useful and the Misleading. *Clin Pharmacol Ther*. 2019;106(1):43-44. doi:10.1002/cpt.1405
72. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. doi:10.1136/bmj.i4919
73. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: Methodological challenges and implications for drug development. *Clin Pharmacol Ther*. 2011;90(6):777-790. doi:10.1038/clpt.2011.235
74. Lawlor DA, Tilling K, Smith GD. Triangulation in aetiological epidemiology. *Int J Epidemiol*. 2016;45(6):1866-1886. doi:10.1093/ije/dyw314
75. Van Der Weele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-Value. *Ann Intern Med*. 2017;167(4):268-274. doi:10.7326/M16-2607
76. Lash TL, Fox MP, Maclehose RF, Maldonado G, Mccandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969–1985. doi:10.1093/ije/dyu149
77. Sammon CJ, Leahy TP, Gsteiger S, Ramagopalan S. Real-world evidence and nonrandomized data in health technology assessment: use existing methods to address unmeasured confounding? *J Comp Eff Res*. Published online 2020. doi:10.2217/cer-2020-0112
78. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: Opportunities and challenges. *J Comp Eff Res*. 2018;7(12):1133-1143. doi:10.2217/cer-2018-0066
79. Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clin Pharmacol Ther*. 2019;106(1):36-39. doi:10.1002/cpt.1426

80. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885
81. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349. doi:10.1016/j.jclinepi.2007.11.008
82. Wang S V., Schneeweiss S, Berger ML, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Value Heal*. 2017;20(8):1009-1022. doi:10.1016/j.jval.2017.08.3018
83. Observational Health Data Sciences and Informatics. *The Book of OHDSI*.; 2019. https://ohdsi.github.io/TheBookOfOhdsi/
84. Duke-Margolis Centre. *Determining Real-World Data's Fitness for Use and the Role of Reliability*.; 2019. https://healthpolicy.duke.edu/sites/default/files/u31/rwd_reliability.pdf
85. EUnetHTA. REQueST® Tool and its vision paper. Published 2019. Accessed March 25, 2020. https://eunethta.eu/request-tool-and-its-vision-paper/
86. Oortwijn W, Sampietro-Colom L, Trowman R. How to Deal with the Inevitable: Generating Real-World Data and Using Real-World Evidence for HTA Purposes - From Theory to Action. *Int J Technol Assess Health Care*. 2019;35(4):346-350. doi:10.1017/S0266462319000400
87. Dreyer NA, Velentgas P, Westrich K, Dubois R. The grace checklist for rating the quality of observational studies of comparative effectiveness: A tale of hope and caution. *J Manag Care Pharm*. 2014;20(3):301-308. doi:10.18553/jmcp.2014.20.3.301
88. Grimm SE, Pouwels X, Ramaekers BLT, et al. Development and Validation of the TRansparent Uncertainty ASsessmenT (TRUST) Tool for Assessing Uncertainties in Health Economic Decision Models. *Pharmacoeconomics*. 2020;38(2):205-216. doi:10.1007/s40273-019-00855-9
89. Schünemann H. All evidence is real world evidence. *BMJ Opin*. Published online 2019. doi:https://blogs.bmj.com/bmj/2019/03/29/holger-j-schunemann-all-evidence-is-real-world-evidence/
90. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.ad
91. Thornton J, Alderson P, Tan T, et al. Introducing GRADE across the NICE clinical guideline program. *J Clin Epidemiol*. 2013;66(2):124-131. doi:10.1016/j.jclinepi.2011.12.007
92. Solà-Morales O, Volmer T, Mantovani L. Perspectives to mitigate payer uncertainty in health technology assessment of novel oncology drugs. *J Mark Access Heal Policy*. 2019;7(1):1562861. doi:10.1080/20016689.2018.1562861
93. Garrison LP, Towse A, Briggs A, et al. Performance-based risk-sharing arrangements - Good practices for design, implementation, and evaluation: Report of the ISPOR good practices for performance-based risk-sharing arrangements task force. *Value Heal*. 2013;16(5):703-719. doi:10.1016/j.jval.2013.04.011
94. Wenzl M, Chapman S. *Performance-Based Managed Entry Agreements for New Medicines in OECD Countries and EU Member States: How They Work and Possible Improvements Going Forward*.; 2019. doi:10.1787/6e5e4c0f-en
95. Dabbous M, Chachoua L, Caban A, Toumi M. Managed Entry Agreements: Policy Analysis From the European Perspective. *Value Heal*. 2020;23(4):425-433. doi:10.1016/j.jval.2019.12.008

96. Makady A, Nijmeijer H, de Boer A, Hillege J, Klungel O, Goettsch W. Implementation of Conditional Reimbursement Schemes in HTA Practice: Experiences from the Netherlands. *Value Heal*. 2016;19(7):A348. doi:10.1016/j.jval.2016.09.011
97. van de Wetering EJ, van Exel J, Brouwer WBF. The Challenge of Conditional Reimbursement: Stopping Reimbursement Can Be More Difficult Than Not Starting in the First Place! *Value Heal*. 2017;20(1):118-125. doi:10.1016/j.jval.2016.09.001
98. Woloshin S, Schwartz LM, White B, Moore TJ. The fate of FDA postapproval studies. *N Engl J Med*. 2017;377(12):1114-1117. doi:10.1056/NEJMp1705800
99. Banzi R, Gerardi C, Bertele' V, Garattini S. Approvals of drugs with uncertain benefit-risk profiles in Europe. *Eur J Intern Med*. 2015;26(8):572-584. doi:10.1016/j.ejim.2015.08.008
100. Hoekman J, Klamer TT, Mantel-Teeuwisse AK, Leufkens HGM, De Bruin ML. Characteristics and follow-up of postmarketing studies of conditionally authorized medicines in the EU. *Br J Clin Pharmacol*. 2016;82(1):213-226. doi:10.1111/bcp.12940
101. Gliklich RE, Leavy MB. Assessing Real-World Data Quality: The Application of Patient Registry Quality Criteria to Real-World Data and Real-World Evidence. *Ther Innov Regul Sci*. 2020;54(2):303-307. doi:10.1007/s43441-019-00058-6
102. Head of Medicines Agencies and European Medicines Agency. *HMA-EMA Joint Big Data Taskforce Phase II Report: 'Evolving Data-Driven Regulation.'*; 2020. https://www.ema.europa.eu/en/documents/other/hma-ema-joint-big-data-taskforce-phase-ii-report-evolving-data-driven-regulation_en.pdf
103. National Institute for Health and Clinical Excellence. *Increasing Use of Health and Social Care Data in Guidance Development*.; 2019. https://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
104. Lovestone S. The European medical information framework: A novel ecosystem for sharing healthcare data across Europe. *Learn Heal Syst*. 2020;4(2):e10214. doi:10.1002/lrh2.10214
105. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). ENCePP Resources Database. Published 2020. http://www.encepp.eu/encepp/resourcesDatabase.jsp
106. European Medicines Agency. *A Common Data Model for Europe? - Why? Which? How?*; 2018. https://www.ema.europa.eu/en/documents/report/common-data-model-europe-why-which-how-workshop-report_en.pdf
107. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: A practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(6 SUPPL.). doi:10.1097/MLR.0b013e3181d9919f
108. Gill J, Prasad V. Improving observational studies in the era of big data. *Lancet*. 2018;392(10149):716-717. doi:10.1016/S0140-6736(18)31619-2
109. Efthimiou O, Mavridis D, Debray TPA, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med*. 2017;36(8):1210-1226. doi:10.1002/sim.7223