

Target trial emulation for transparent and robust estimation of treatment effects for health technology assessment using real-world data: opportunities and challenges.

Manuel Gomes^{1*}, Nick Latimer², Marta Soares³, Sofia Dias⁴, Gianluca Baio⁵, Nick Freemantle⁶, Dalia Dawoud^{7 8}, Allan Wailoo², Richard Grieve⁹

1 Department of Applied Health Research, University College London.

2 School of Health and Related Research, University of Sheffield.

3 Centre for Health Economics, University of York

4 Centre for Reviews and Dissemination, University of York

5 Department of Statistical Science, University College London.

6 Institute of Clinical Trials and Methodology, University College London

7 Science, Policy and Research group, National Institute for Health and Care Excellence.

8 Faculty of Pharmacy, Cairo University, Egypt.

9 Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine.

*Corresponding author: m.gomes@ucl.ac.uk

Funding: NL is supported by Yorkshire Cancer Research (Award S406NL).

Conflicts of interest: The authors have no conflicts or competing interests to declare

Abstract

Evidence about the relative effects of new treatments is typically collected in randomised controlled trials (RCTs). In many instances, evidence from RCTs falls short of the needs of health technology assessment (HTA). For example, RCTs may not be able to capture longer-term treatment effects, or include all relevant comparators and outcomes required for HTA purposes. Information routinely collected about patients and the care they receive have been increasingly used to complement RCT evidence on treatment effects. However, such routine (or real-world) data are not collected for research purposes, so investigators have little control over the way patients are selected into the study or allocated to the different treatment groups, introducing biases for example due to selection or confounding.

A promising approach to minimise common biases in non-randomised studies that use real-world data (RWD) is to apply design principles from RCTs. This approach, known as 'target trial emulation' (TTE), involves i) developing the protocol with respect to core study design and analysis components of the hypothetical RCT that would answer the question of interest, and ii) applying this protocol to the RWD so that it mimics the data that would have been gathered for the RCT. By making the 'target trial' explicit, TTE helps avoid common design flaws and methodological pitfalls in the analysis of non-randomised studies, keeping each step transparent and accessible. It provides a coherent framework that embeds existing analytical methods to minimise confounding, helps identify potential limitations of RWD, and the extent to which these affect the HTA decision. This paper provides a broad overview of TTE and discusses the opportunities and challenges of using this approach in HTA. We describe the basic principles of trial emulation, outline some areas where TTE using RWD can help complement RCT evidence in HTA, identify potential barriers to its adoption in the HTA setting and highlight some priorities for future work.

Key points for decision makers

- Deriving treatment effects from real-world data (RWD) can be viewed as an attempt to emulate a randomised experiment, the target trial, that would answer the question of interest. By explicitly describing the target trial and emulating it using RWD, target trial emulation (TTE) can help prevent common pitfalls and biases in non-randomised studies.
- We identified several areas where TTE can be used to complement randomised evidence to inform health technology assessment (HTA): 1) to provide a structured approach to aid the analysis of uncontrolled studies, 2) to support the review of HTA recommendations, 3) to improve indirect treatment comparisons, and 4) to evaluate personalised, adaptive treatment strategies.
- Further investments to improve the quality and accessibility of RWD are required to help TTE support the expanding role of RWD in HTA. More methodological guidance and worked examples of TTE in HTA settings are needed to encourage its uptake in practice.

1. Introduction

Routinely-collected data in electronic health records, registries and administrative datasets (hereafter denoted as 'real-world data' - RWD) are increasingly used to evaluate the relative effectiveness and cost-effectiveness of health interventions[1-3]. For example, 38% (45 out of 118) of NICE health technology appraisal (HTAs) completed between 2010 and 2015 have considered non-randomised studies to derive treatment effects[4]. This included, for instance, analysing data from a single-arm trial for the new treatment, and registry data for the comparator group. More generally, international HTA agencies are beginning to develop frameworks to guide the use of RWD in reimbursement decisions[5-7]. Concurrently, data harmonisation and linkage of large-scale real-world datasets have considerably improved in recent years, offering major opportunities to inform healthcare decision-making[8, 9].

Deriving valid estimates of treatment effects from RWD is challenging, not least because RWD are usually not collected for research purposes. Recurrent challenges include: i) data quality and completeness - for example real-world datasets may include vast amounts of information, but there may be issues with inconsistent, inaccurate or incomplete records; ii) defining the comparator and target population, and ensuring alignment of eligibility, treatment assignment and start of follow-up, which often introduce selection biases when done inappropriately[10, 11]; iii) potential for confounding due to the presence of both measured and unmeasured factors that may affect the choice of treatment and outcomes of interest[12].

Recent methods guides called for a more coherent approach to the design and analysis of non-randomised studies in HTA[13-16]. These highlighted a general lack of awareness about appropriately designing such studies for estimating treatment effects. Traditionally, effectiveness and cost-effectiveness studies using RWD tend to focus on confounding adjustments made at the analysis stage. However, conventional adjustments are often conducted inappropriately, and their underlying assumptions are often implausible or not readily understood[14, 17]. By focusing on the analysis stage, researchers often overlook important design aspects, such as a clear definition of the comparator and target population, and correct alignment of eligibility, treatment assignment and start of follow-up, which can introduce severe biases[1]. These methodological guidelines suggested that, as HTA agencies move towards greater use of RWD, approaches that can minimise biases at the design stage should be prioritised.

An approach increasingly used in epidemiological studies to minimise common biases in the analysis of RWD is to use design principles from RCTs. This approach is known as 'target trial emulation' (TTE)[18, 19]. It involves specifying the protocol for the hypothetical RCT (the 'target trial') that would answer the question of interest (e.g. treatment effect), and emulating

this target trial using RWD. Many studies have demonstrated that by using a TTE framework, combined with appropriate analytical methods, it is possible to approximate the results from well-conducted RCTs[11, 18, 20-23]. These included initiatives where different research teams emulated the same target trial[24], or where trial emulation was completed before the RCT results were known[25]. In some instances where discrepancies between the RCT and the emulated target trial were reported[26], studies illustrated how TTE could help articulate the reasons for those differences and understand where the limitations of RWD might lie. For example, RWD might not allow emulation of all eligibility criteria of the RCT. Some studies demonstrated that differences in the results between the emulated target trial and the RCT were due to differences in design features, for example the definition of time zero (baseline), rather than the presence of unmeasured confounding[18, 21-23]. In practice, however, non-randomised studies cannot exclude the possibility of unmeasured confounding, so TTE should be best viewed as complementary to well-conducted RCTs.

TTE is becoming an integral part of major guidelines for the use of real-world evidence. For example, Cochrane's methods guidance advocates the target trial approach to facilitate the assessment of risk of bias in non-randomised studies[27]. The updated NICE methods guidelines [15] are endorsing the use of TTE in HTA, but offer little guidance on how this approach might be useful and how it might complement existing analytical methods. The aim of this paper is to provide a broad overview of the target trial approach and discuss the opportunities and challenges of adopting TTE in HTA. The next section describes the basic principles of trial emulation, drawing on the methods and applications of TTE in the causal inference literature[18, 19, 22, 23, 26, 28]. A hypothetical trial to evaluate optimal treatment intensification for type-2 diabetes illustrates how TTE principles can be applied. Section 3 highlights some areas where TTE can complement RCT evidence, based on the authors' extensive experience in the use of RWD in HTA[2, 3, 13, 14, 29]. Section 4 identifies some potential barriers to the adoption of TTE in HTA, drawing on previous experience with applying TTE to the HTA setting[30-32]. Section 5 outlines the next steps to address these challenges, and some concluding remarks are provided in Section 6.

2. Target trial emulation

TTE entails three broad steps, which we describe below drawing on a hypothetical trial to estimate the effect of alternative intensification strategies for glycaemic control in type-2 diabetes patients who failed metformin as first-line therapy. The first intensification often involves a dual therapy combining metformin with Sulphonylureas (SU) or newer drugs, such as DPP4 and SGLT2 inhibitors. Table 1 provides an example of a target trial protocol for a hypothetical pragmatic RCT that would address this question.

Step 1) Specify the target trial protocol with respect to core study design and analysis components. This should draw on discussions with subject-matter experts (e.g. clinicians) and patients to ensure the protocol poses a relevant question to the target population, and closely reflects routine clinical practice.

1.1 Eligibility criteria: inclusion/exclusion criteria that one would use in the hypothetical RCT to define the type-2 diabetes population of interest. In our example, the target population is type-2 diabetic patients who have taken metformin (first-line therapy) for at least 3 months and had no previous treatment intensification.

1.2 Treatment strategies: treatment intensification strategies should be consistent with those observed in routine clinical practice, for example, in line with national guidelines for glycaemic control in type-2 diabetes.

1.3 Assignment procedures: patients and health professionals are typically aware of treatment intensification received, so blind assignment is not possible in this target trial.

1.4 Time zero: This is analogous to the point of randomisation (baseline), where eligibility assessment occurs, the treatment strategies are initiated, and follow up starts.

1.5 Outcomes: pre-specify outcomes of interest: in our example these include HbA1c, all-cause mortality, CVD events and costs

1.6 Estimands: the focus of the estimation should reflect the substantive question of interest[33]. For example, intention-to-treat (ITT) (the effect of *initiating* one of the treatment strategies, irrespective of compliance) is often of prime interest to HTA agencies. An alternative estimand is the per-protocol (PP) effect, i.e. the effect of *taking* one of them for a certain period of time.

1.7 Analysis plan: describe the methods required to estimate the effect of interest, including addressing any sources of biases, such as selection bias and loss to follow-up. Central to this is the approach taken to emulate randomisation, i.e. ensure comparability between groups. TTE does not mandate any specific analytical approach, but matching, inverse probability weighting or stratification are typically used to balance groups according to observed confounders. If unmeasured confounding is anticipated, then the TTE can be combined with approaches to address residual confounding, such as instrumental variables[34], outcome controls[35] and E-values[36]. The choice of approach will depend on the specific decision problem and RWD at hand, and should be supported by directed acyclic graphs (DAGs)[37] to characterise the potential unmeasured confounders.

Step 2) Emulate the target trial using RWD, i.e. apply the pre-specified protocol of the target trial to the RWD so that it mimics the data (and the analysis) that would have been gathered for the hypothetical RCT.

Suppose we had rich RWD about type-2 diabetes patients and the treatments they receive from electronic health records. In England, for example, these could come from routinely collected data in general practices (Clinical Practice Research Datalink-CPRD[38]), and hospital's administrative data (Hospital Episode Statistics-HES[39]). The right-hand side column of Table 1 summarises the extent to which these data would emulate the target trial.

2.1 Eligibility criteria: We would expect to be able to successfully emulate the eligibility criteria of the hypothetical RCT for patients who had relevant data on treatment history prior to baseline.

2.2 Treatment strategies: Treatments received in routine practice may or may not be consistent with one of the pre-specified intensification strategies. If baseline data indicates a patient is intensifying treatment with either Sulfonylureas, DPP4 or SGLT2 inhibitors, then the patient is assigned to that strategy. If the patient is starting an irrelevant comparator for the TTE (e.g. insulin) the patient is deemed ineligible, and excluded.

2.3 Assignment procedures: To emulate random treatment assignment, adjustment for confounders is required to ensure comparability between groups. Matching or other techniques may be used to minimise differences in baseline confounders between comparison groups. In practice, one cannot exclude the possibility of residual confounding, so analytical approaches to assess this, such as negative controls[35] or E-values[36], should be routinely considered.

2.4 Time zero: this would be defined at the time (or just before) eligible patients initiated a specific treatment intensification strategy.

2.5 Outcomes: All pre-specified outcomes were expected to be measured in linked CPRD-HES data.

2.6. Estimands: Decision-makers would be interested in both ITT (comparison of treatment initiators) and PP effects (comparison of treatment compliers)

2.7 Analysis plan: The ITT effect would be obtained by comparing groups defined by initiation of the intensification strategies, adjusting for baseline confounders (e.g. via matching). Estimation of the PP effect would require further adjustment for post-baseline factors associated with treatment adherence (e.g. using inverse probability weighting). Similar adjustments will be required if there is informative loss to follow-up or missing data.

Step 3) Estimate the treatment effects of interest in the emulated target trial using the methods specified in the analysis plan.

If the emulation of the target trial is successful, i.e. if we are able to fully replicate each component of the target trial protocol using RWD (Step 2), the resulting treatment effects will approximate that of the hypothetical RCT, had it been conducted. In practice, it will be difficult to establish if emulation failed due to unmeasured confounding. However, there will be situations where the analyst can be more confident about successful emulation of randomisation; for example, when the TTE successfully emulates an existing RCT in the relevant population [26], or when TTE is combined with a valid instrumental variable [34]. Missing data and measurement error in RWD may also affect our ability to successfully emulate some elements of the protocol, and may require imposing additional assumptions.

By explicitly describing and emulating the target trial, TTE has several strengths for deriving treatment effects from RWD for HTA purposes. Firstly, it can avoid common design flaws and apparent paradoxes in non-randomised studies[40]. For example, selection biases due to the misalignment between eligibility assessment and treatment assignment[18, 22, 23]. Secondly, it provides a structured process for identifying potential limitations and articulating trade-offs in non-randomised studies. In particular, the RWD may not include sufficient information on confounders to emulate randomisation. By clearly specifying the assignment procedures, TTE helps us think about appropriate analytical strategies to explore unmeasured confounding or redefine the target trial in meaningful ways[35, 40]. Thirdly, TTE explicitly ties the design and analysis of the non-randomised study to the target trial, facilitating the interpretation and communication of its underlying assumptions and findings. Fourth, by exploiting large-scale RWD of sufficient quality, TTE can include a larger, more diverse population, and longer follow-up compared to the hypothetical RCT. In addition, the types of patients and the care they receive in the trial emulation can reflect more closely those observed in clinical routine practice, compared to those included in actual RCTs.

3. How can trial emulation help leverage RWD for HTA?

We have identified four areas for which TTE can be helpful in HTA.

HTA decisions based on uncontrolled studies

For many reimbursement decisions, for example on new cancer drugs, evidence on treatment effectiveness and cost-effectiveness comes from uncontrolled studies, such as single-arm trials[2, 3]. Typically, where uncontrolled studies of new treatments exist, the major concern is the lack of a well-defined comparator group, which tends to introduce severe biases, for example due to systematic differences in populations and standards of care[4, 41].

Uncontrolled studies often derive treatment effects by comparing the treated patients with historical controls, i.e. a group of patients who did not receive the treatment. Control patients have been increasingly taken from RWD sources, such as disease registries[2]. Typically treated patients in the uncontrolled study are matched, for example using propensity score matching, with the real-world comparator group. However, these analyses often overlook important design aspects, such as the correct alignment of eligibility assessment, time zero and start of follow-up in the matched control group. TTE builds on existing research exploring how comparator groups can be generated using RWD[42, 43] by providing a structured approach to aid the analysis and interpretation of uncontrolled studies.

Supporting the review of HTA recommendations

In many instances, reimbursement decisions are made and given a set time at which they will be reviewed. For example, innovative treatments may be made available more quickly to patients through accelerated access initiatives[44]. Usually, recommendations include a caveat that they will be reviewed at a specified time-point in the future – usually around 2-3

years after the initial recommendation. However, the review of HTA recommendations rarely happens in a meaningful way. Once the treatment is recommended and widely used in the NHS, it is often difficult to stop its use or conduct a randomised experiment. For example, NHS funding for β interferon for multiple sclerosis continued after RWD found it did not improve patient outcomes[45]. In addition, the review of HTA recommendations is often based on immature data [46].

At the time of the review, useful information about the effect of the treatment in real-world clinical practice is often available. By carefully designing and emulating the hypothetical trial that would address the relevant treatment effectiveness questions using such RWD, TTE could help establish a more meaningful review of (conditional) reimbursement decisions. When considered at early stages (i.e. at the time of the initial recommendation), TTE used in combination with value of information analyses could help identify critical data items, such as key confounders, to be collected in RWD during the relevant period, which may require the linkage of different routine data sources. In this way, TTE could complement other sources of evidence (such as updated data-cuts from RCTs) to contribute to meaningful reviews of HTA recommendations.

Improving indirect treatment comparisons

The absence of head-to-head RCTs raises important challenges for evaluating alternative treatments competing for healthcare resources. Each treatment is typically evaluated in a separate RCT, and a common, relevant comparator may be lacking[29]. In addition, there may be broader systematic differences (e.g. eligibility criteria and care standards) between the individual RCTs. Relative treatment effects are commonly derived using indirect comparison approaches, such as network meta-analysis, but these are typically based on aggregate data from the individual RCTs, and hence unlikely to adequately account for important differences across studies[29, 47].

TTE can complement indirect treatment comparisons in HTA in several ways. Firstly, the TTE can allow for head-to-head comparisons of treatment options available in routine clinical practice. For example, NICE has recently appraised seven different biologics[48] for patients with rheumatoid arthritis who failed non-biologics, but only 3 out of 30 RCTs included in the appraisal were head-to-head RCTs. Biologic registries, such as FORWARD[49], collect large amounts of information about RA patients and biologic treatments, and would enable direct comparisons between these. Clearly, TTE will have limited applicability for indirect comparisons of new treatments (e.g. cancer drugs) for which RWD is not yet available. Secondly, TTE allows us to assess the plausibility of assumptions made by conventional indirect treatment comparison methods. For example, by explicitly specifying the target trial and emulating it using RWD, TTE can help examine potential differences between existing RCTs in terms of patient and clinical factors (target population), treatment pathways and relative effects over time. This is particularly valuable when there are differences in effect-modifiers between RCTs that cannot be properly adjusted for by conventional methods. Thirdly, TTE can enable meaningful comparisons of outcomes of prime interest to HTA agencies that may not be included across all RCTs. These may include clinical measures of disease severity or progression, complemented with health-related quality of life and cost data.

Evaluation of personalised, adaptive treatment strategies

Evidence from RCTs is often insufficient to inform the targeting of the right treatments for the right patients over time. By harnessing large-scale routine datasets, which include data on important individual risk factors over a long follow-up period and across a diverse population, TTE can provide valuable, robust evidence to inform personalised, adaptive treatment strategies. This is critical, for example, in the evaluation of treatment strategies for the management of long-term conditions, such as diabetes and hypertension. TTE can help optimise the use of such treatments using RWD because these reflect how chronic diseases are managed in practice and offer insights about treatment effectiveness that would not be

captured in RCTs. Again, this is limited to established interventions for which RWD is available.

The advantages of TTE are particularly pronounced in the evaluation of complex treatment strategies that are sustained over time and are dynamic in nature[50]. In these settings, treatment assignment and eligibility assessment typically occur at multiple time points; for example type-2 diabetes patients may need to adapt treatment intensification whenever glycaemic control is inadequate. TTE enables us to carefully consider the alignment of eligibility assessment, time zero and treatment initiation at different points in time, as well as an appropriate strategy to address the time-varying nature of the confounding (e.g. g-methods)[40].

4. Barriers to the adoption of TTE in HTA

While the potential benefits of TTE to derive treatment effects from RWD may be substantial, this approach has not yet permeated HTA practice. Drawing on our recent experience in applying TTE to the HTA setting[30-32], we have identified several potential barriers to the adoption of TTE in HTA. These concern the use of the TTE framework specifically, but also limitations associated with the use of RWD which may limit the use of TTE in HTA:

Study design - While TTE has been used for over a decade in other areas, such as Epidemiology, it is a relatively new concept in HTA. Although the idea of ‘target trial’ is implicit in many analyses of RWD for estimating treatment effects, non-randomised studies rarely characterise the target trial itself. Most published papers still focus their efforts at the analysis stage, most notably attempting to tackle confounding. More attention needs to be devoted to careful design of non-randomised studies, which should be grounded in trial emulation principles.

Causal inference methods - Some of the causal inference tools required for analysis stage of TTE are unlikely to be in HTA users’ standard toolkit. For example, g-methods[51] (e.g. marginal structural models), often required to make adjustments for informative loss to follow-up or time-dependent confounding, have received little attention in HTA. In addition, widely used epidemiological methods to detect potential residual confounding, such as outcome controls[36] and E-value approaches[35], are yet to permeate practice in HTA.

Richness of information - RWD of prime interest to HTA users, such as cancer registries, and electronic health records, differ in terms of the depth and breadth of information routinely collected. Some of these RWD sources may not be readily used for TTE because they lack sufficient information to successfully emulate the different components of the target trial, such as eligibility criteria and relevant confounders[42]. The level of richness of information required for successful emulation will crucially depend on the specific question set out by the target trial. However, recent TTE applications in comparative effectiveness research, using US claims data[26, 28], cancer registries[23, 52] or UK electronic health records[22, 50] suggested good levels of success in emulating the target trial, and agreement with actual RCTs[26].

Data access - Gaining access to RWD with the richest information, including those collected and funded by public bodies, tends to take a long time (up to 2 years) and is often highly costly.

Data management - Appropriate application of TTE requires a clear understanding of the underlying data-generating mechanism, which HTA users are often not familiar with. This includes knowledge about the temporality of how patients enter/exit the dataset, and how outcomes, exposures and confounders are operationally defined. This is particularly relevant when comparators come from different RWD sources, which often requires further data harmonisation[53]. For example, treatment and control interventions may not be given contemporaneously or may be given to different treatment populations.

5. Next steps

To help address these challenges, we suggest that further investments should prioritise: 1) improvements on the availability and quality of RWD, 2) better analytical methods and guidance.

An important next step to help TTE support the expanding role of RWD in healthcare decision-making is to improve the integration of existing RWD sources. For example, UK NHS Digital initiatives, such as Trusted Research Environments, are creating information ecosystems that bring together high-quality linked data across different care settings. Further integration of RWD will require partnerships between data controllers (e.g. healthcare organisations), tech providers, clinical and policy experts and the public[9]. Another crucial step to facilitate the use of RWD in HTA in a more tailored and timely fashion is to improve the infrastructures and data access mechanisms to enable users to readily access RWD. This may involve setting up (ideally open-source) digital platforms, such as OpenSAFELY [www.opensafely.org] and Federated Data Networks [www.ehden.eu], that support the analysis of large-scale linked health data. Such platforms can help standardise the data management pathway (principles for data provenance and curation)[53, 54] to encourage reproducibility and transparency, while the data stay 'local', overcoming data sharing barriers.

More methodological guidance and applications of TTE in HTA settings are also needed to encourage its uptake and acceptance by HTA agencies. Guiding principles on the implementation of TTE in other areas, such as epidemiology, are not directly applicable to the HTA context. For example, the flexibility of TTE for improving indirect treatment comparisons and the analysis of uncontrolled studies has not been exploited elsewhere, and further methodological work in these areas is needed. Calibration of RWD against existing RCTs (not necessarily the one we wish to emulate) should also be promoted to help us assess the extent to which RWD sources can replicate well conducted RCTs in the relevant population. If existing RCTs can be successfully emulated using a specific RWD source, researchers may be more confident that the 'no unmeasured confounding' assumption is satisfied in the trial emulation. In addition, methodological recommendations for RWD studies in HTA, such as ISPOR's Good Research Practices reports[55-57], will need to be updated to encourage future HTA studies to incorporate the principles of the target trial approach.

6. Conclusions

TTE provides a much-needed intuitive, general framework for the design and analysis of non-randomised studies in HTA. This approach supports current ambitions of HTA agencies for a more central role for RWD to better inform an iterative HTA process and enable more dynamic treatment guidelines. TTE can bring more clarity and confidence to the use of RWD in HTA, providing a common template for pre-registered study protocols, and minimising the potential for cherry-picking results.

RWD will not always enable us to emulate the ideal target trial, but the principles behind TTE ought to be adopted, nonetheless. These encourage researchers to carefully consider each element of study design to minimise common study design flaws, while keeping each step of the way transparent and accessible. While it is generally impossible to perfectly emulate randomisation and eliminate the risk of unmeasured confounding, TTE enables us to systematically articulate the trade-offs and compromises made in non-randomised studies. In addition, by explicitly describing the target trial, TTE facilitates alignment between the design, analysis and reporting of RCTs and non-randomised studies. This helps findings from the latter to be more readily understood by clinical experts, industry and decision makers and facilitates integration of both types of study in the decision-making process. TTE allows researchers to clearly identify potential limitations of RWD and help decision-makers understand the extent to which these affect the decision problem at hand.

Table 1 – Brief description of target trial emulation in the evaluation of optimal treatment intensification for type-2 diabetes

Component	Target trial protocol (pragmatic RCT)	Emulated trial using RWD
<u>Research question:</u> <i>Would HbA1c, risk of death and CVD, and costs differ according to intensification strategy for T2D patients who failed metformin?</i>		
1. Eligibility criteria	<p><i>Inclusion criteria:</i></p> <ul style="list-style-type: none"> - adult (18+) patients with diagnosis of type-2 diabetes - history of metformin for at least 3 months - inadequate glycaemic control (HbA1c \geq 7.5%) - no previous treatment intensification <p><i>Exclusion criteria:</i></p> <ul style="list-style-type: none"> - metformin not tolerated 	<p><i>Inclusion criteria (at time of eligibility assessment):</i></p> <ul style="list-style-type: none"> - same <p><i>Exclusion criteria:</i></p> <ul style="list-style-type: none"> - same, but patients with no available data on treatment history in the 6 months prior to baseline are also excluded
2. Treatment strategies	<p><i>Three treatment strategies:</i></p> <ol style="list-style-type: none"> 1) start intensification with SU 2) start intensification with DPP4i 3) start intensification with SGLT2i 	If treatment data at baseline is consistent with the intensification strategy on the left, patient is assigned to that strategy. If not (e.g. patient starts insulin), patient is deemed ineligible.
3. Assignment procedures	Participants are randomly assigned to a treatment strategy at baseline, and are aware of the assigned strategy (unblinded)	Randomisation is emulated via adjustment for all confounders to minimise differences between comparison groups.
4. Follow up	<p><i>Start:</i> time of treatment assignment (randomisation)</p> <p><i>End:</i> at the earliest of 2 years post-randomisation, death, cardiovascular disease (CVD) event, or drop-out</p>	<p><i>Start:</i> initiation of treatment strategy</p> <p><i>End:</i> at the earliest of 2 years post-baseline, death, CVD event, or drop-out.</p>
5. Outcomes	HbA1c, all-cause mortality, CVD events and costs	Same
6. Estimands	<ul style="list-style-type: none"> - <i>Intention-to-treat (ITT) effect:</i> effect of being assigned to one of the treatment strategies - <i>per-protocol (PP) effect:</i> effect of continuously taking one of the treatment strategies for 2 years. 	<ul style="list-style-type: none"> - observational analogue of ITT effect: comparison of initiators of different treatment strategies - observational analogue of PP effect: comparison of individuals who adhered to the assigned treatment over time
7. Analysis plan	<ul style="list-style-type: none"> - Intention-to-treat analysis: compare mean or risk differences between randomised groups - per-protocol analysis: Mean or risk differences between treatment groups with adjustment for baseline and post-baseline factors associated with treatment adherence. <p>Both need to adjust for censoring if follow up incomplete.</p>	<ul style="list-style-type: none"> - same, but with additional adjustments for baseline confounders - same, but with additional adjustment for both baseline and post-baseline factors associated with informative loss to follow up <p>Both analyses are assuming 'no unmeasured confounding'.</p>

SU: Sulfonylureas, DPP4: Dipeptidyl peptidase 4 inhibitor, SGLT2: Sodium-glucose transport protein 2 inhibitor, HbA1c: haemoglobin A1c.

References

1. Bullement A, Podkonjak T, Robinson MJ, Benson E, Selby R, Hatswell AJ, et al. Real-world evidence use in assessments of cancer drugs by NICE. *Int J Technol Assess Health Care*. 2020;1-7.
2. Goring S, Taylor A, Muller K, Li TJJ, Korol EE, Levy AR, 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.
3. 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.
4. Griffiths EA, Macaulay R, Vadlamudi NK, Uddin J, Samuels ER. The Role of Noncomparative Evidence in Health Technology Assessment Decisions. *Value Health*. 2017;20(10):1245-51.
5. Chan K, Nam S, Evans B, de Oliveira C, Chambers A, Gavura S, et al. Developing a framework to incorporate real-world evidence in cancer drug funding decisions: the Canadian Real-world Evidence for Value of Cancer Drugs (CanREValue) collaboration. *BMJ Open*. 2020;10(1):e032884.
6. Makady A, van Veelen A, Jonsson P, Moseley O, D'Andon A, de Boer A, et al. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. *Pharmacoeconomics*. 2018;36(3):359-68.
7. FDA. Framework for FDA's real-world evidence program. 2018.
8. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol*. 2012;41(6):1625-38.
9. Wood A, Denholm R, Hollings S, Cooper J, Ip S, Walker V, et al. Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ*. 2021;373:n826.
10. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther*. 2017;102(6):924-33.
11. Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, et al. Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*. 2019;188(8):1569-77.
12. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*. 2013;347:f6409.
13. Bell H, Wailoo A, Hernandez-Alava M, Grieve R, Faria R, Gibson L, et al. The use of real world data for the estimation of treatment effects in NICE decision making.; 2016.
14. Faria R, Hernandez-Alava M, Manca A, Wailoo A. The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. NICE DSU Technical Support Document No 17. 2015.
15. NICE. NICE health technology evaluations: the manual. Appendix 1 - Real world evidence framework. 2022.
16. Welton NJ, Phillippo DM, Owen R, Jones HE, Dias S, Bujkiewicz S, et al. CHTE2020 Sources and Synthesis of Evidence: Update to Evidence Synthesis Methods. 2020.
17. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ*. 2013;22(4):486-500.
18. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766-79.
19. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-64.

20. Admon AJ, Donnelly JP, Casey JD, Janz DR, Russell DW, Joffe AM, et al. Emulating a Novel Clinical Trial Using Existing Observational Data. Predicting Results of the PreVent Study. *Ann Am Thorac Soc*. 2019;16(8):998-1007.
21. Boyne DJ, Cheung WY, Hilsden RJ, Sajobi TT, Batra A, Friedenreich CM, et al. Association of a Shortened Duration of Adjuvant Chemotherapy With Overall Survival Among Individuals With Stage III Colon Cancer. *JAMA Netw Open*. 2021;4(3):e213587.
22. Dickerman BA, Garcia-Albeniz X, Logan RW, Denaxas S, Hernan MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med*. 2019;25(10):1601-6.
23. Emilsson L, Garcia-Albeniz X, Logan RW, Caniglia EC, Kalager M, Hernan MA. Examining Bias in Studies of Statin Treatment and Survival in Patients With Cancer. *JAMA Oncol*. 2018;4(1):63-70.
24. OPERAND. The Observational Patient Evidence for Regulatory Approval and uNderstanding Disease project. 2020.
25. Noseworthy PA, Gersh BJ, Kent DM, Piccini JP, Packer DL, Shah ND, et al. Atrial fibrillation ablation in practice: assessing CABANA generalizability. *Eur Heart J*. 2019;40(16):1257-64.
26. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies: First Results From the RCT DUPLICATE Initiative. *Circulation*. 2021;143(10):1002-13.
27. Sterne J, Hernan MA, McAleenan A, Reeves B, Higgins JPT. Chapter 25: Assessing risk of bias in a non-randomized study. *Cochrane Training Handbook*. 2021.
28. Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol*. 2017;32(6):495-500.
29. Phillippo DM, Dias S, Elsadat A, Ades AE, Welton NJ. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. *Int J Technol Assess Health Care*. 2019;35(3):221-8.
30. Katsoulis M, Lai AG, Diaz-Ordaz K, Gomes M, Pasea L, Banerjee A, et al. Identifying adults at high-risk for change in weight and BMI in England: a longitudinal, large-scale, population-based cohort study using electronic health records. *Lancet Diabetes Endocrinol*. 2021;9(10):681-94.
31. Katsoulis M, Stavola BD, Diaz-Ordaz K, Gomes M, Lai A, Lagiou P, et al. Weight Change and the Onset of Cardiovascular Diseases: Emulating Trials Using Electronic Health Records. *Epidemiology*. 2021;32(5):744-55.
32. Fotheringham J, Latimer N, Froissart M, Kronenberg F, Stenvinkel P, Floege J, et al. Survival on four compared with three times per week haemodialysis in high ultrafiltration patients: an observational study. *Clin Kidney J*. 2021;14(2):665-72.
33. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2017.
34. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17(4):360-72.
35. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-8.
36. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 2017;167(4):268-74.
37. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-32.
38. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-36.
39. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol*. 2017;46(4):1093-i.

40. Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-5.
41. 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;90:108-18.
42. Davies J, Martinec M, Delmar P, Coudert M, Bordogna W, Golding S, et al. Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *J Comp Eff Res.* 2018;7(9):855-65.
43. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and External Controls in Clinical Trials - A Primer for Researchers. *Clin Epidemiol.* 2020;12:457-67.
44. MHRA. Early access to medicines scheme (EAMS). 2014.
45. McCabe C, Chilcott J, Claxton K, Tappenden P, Cooper C, Roberts J, et al. Continuing the multiple sclerosis risk sharing scheme is unjustified. *BMJ.* 2010;340:c1786.
46. Tai TA, Latimer NR, Benedict A, Kiss Z, Nikolaou A. Prevalence of Immature Survival Data for Anti-Cancer Drugs Presented to the National Institute for Health and Care Excellence and Impact on Decision Making. *Value Health.* 2021;24(4):505-12.
47. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Med Decis Making.* 2018;38(2):200-11.
48. NICE. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed. Technology appraisal guidance [TA375]. 2016.
49. Wolfe F, Michaud K. The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank. *Rheumatology (Oxford).* 2011;50(1):16-24.
50. Danaei G, Garcia Rodriguez LA, Cantero OF, Logan RW, Hernan MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol.* 2018;96:12-22.
51. Hernan MA, Robins JM. Causal Inference: What if?: Boca Raton: Chapman & Hall/CRC; 2021.
52. Petito LC, Garcia-Albeniz X, Logan RW, Howlader N, Mariotto AB, Dahabreh IJ, et al. Estimates of Overall Survival in Patients With Cancer Receiving Different Treatment Regimens: Emulating Hypothetical Target Trials in the Surveillance, Epidemiology, and End Results (SEER)-Medicare Linked Database. *JAMA Netw Open.* 2020;3(3):e200452.
53. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Value Health.* 2017;20(8):1009-22.
54. Kent S, Burn E, Dawoud D, Jonsson P, Ostby JT, Hughes N, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *Pharmacoeconomics.* 2021;39(3):275-85.
55. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. *Value Health.* 2009;12(8):1044-52.
56. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, 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 Health.* 2017;20(8):1003-8.
57. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research

Practices for Retrospective Database Analysis - Task Force Report-Part II. Value Health. 2009;12(8):1053-61.