Antihyperglycemic Therapy and Cardiovascular Risk: Design and Emulation of a Target Trial Using Healthcare Databases

Published May 24, 2019

By Miguel Hernán

Miguel Hernán, MD, DrPH, is the Kolokotrones Professor of Biostatistics and Epidemiology at Harvard T.H. Chan School of Public Health.

He also is Gästprofessor i epidemiologi, Karolinska Institutet; special government employee, US Food and Drug Administration; data science adviser, *ProPublica*; and editor, *Epidemiology*.

His research focuses on learning what works for the treatment and prevention of diseases such as cancer, cardiovascular disease, and HIV infection.

Acknowledgements

Background: Knowledge Gaps on Cardiovascular Effects of Antihyperglycemic Therapy

Diabetes affects over 30 million Americans, with 1.5 million new cases in 2015,[1] and an estimated $327 billion in annual costs in 2017.[2] Individuals newly diagnosed with type 2 diabetes typically begin treatment with metformin and lifestyle modifications, but many require the addition of a second-line antihyperglycemic therapy within the first year. This choice of therapy after metformin is not straightforward because the multiple available drugs, distributed across six or more classes, have different benefits and risks.

While many randomized trials have been or are being conducted to help guide clinical decisions about second-line treatments for diabetes, important questions remain about their comparative effectiveness and safety. A particularly important concern is the effect of second-line agents on the risk of cardiovascular disease.

Several placebo-controlled randomized trials published since 2015 have quantified the cardiovascular effects of single agents in individuals with diabetes, on metformin, and at high cardiovascular risk, but these studies do not provide head-to-head comparisons. The EMPA-REG OUTCOME trial (2015) found a lower risk of cardiovascular death and heart failure hospitalization in patients assigned to empagliflozin, an SGLT2 inhibitor.[3] The CANVAS trial (2017) suggested similar cardiovascular benefits for canagliflozin, another SGLT2 inhibitor.[4] Two trials of GLP-1 receptor agonists, the LEADER trial (2016) for liraglutide[5] and the SUSTAIN-6 trial (2016) for semaglutide,[6] also found cardiovascular benefits. Nine other ongoing or completed trials studied other SGLT2 inhibitors and GLP-1 receptor agonists, as well as drugs in the DPP-4 class, with varying results.

An ongoing pragmatic trial (Glycemia Reduction Approaches in Diabetes or GRADE[7]) compares four second-line agents (glimepiride, sitagliptin, liraglutide, basal insulin glargine) among individuals with diabetes, on metformin, and at low-to-moderate cardiovascular risk, but this trial was designed to study glycemic control and is therefore too small to yield a precise estimation of cardiovascular effects. Also, it does not include an SGLT2 inhibitor arm.

Determining the comparative effectiveness and safety of drugs for second-line antihyperglycemic therapy requires a study that has not been designed yet. Such a study would require a large sample size, a long follow-up, and include individuals at low to moderate risk of cardiovascular disease. The viability of conducting a randomized trial with these characteristics is questionable because of practical and financial challenges. The best alternative to conducting that randomized trial would be to emulate it using high-quality observational data.

THE TARGET TRIAL: A General Framework for Causal Analyses of Observational Data

Observational healthcare databases are often used to address questions about the comparative effectiveness and safety of clinical interventions, that is, to estimate causal effects. The goal of these causal analyses is to inform the development of clinical guidelines and to support decision making by patients, clinicians, and other stakeholders. However, the individuals in healthcare databases are not randomly assigned to the treatment strategies they received. As a result, any outcome differences between groups of individuals receiving different treatments may be confounded, that is, fully or partly explained by differences between the individuals in each group rather than by differential health effects of the treatments.

Because of the possibility of confounding, observational analyses are generally not the preferred choice to study the causal effects of medicaments. All things being equal, a randomized trial is preferable. In fact, for any causal question about comparative effectiveness and safety, we can imagine a hypothetical randomized trial that would answer that question: the *target trial*.[8] Unfortunately, conducting the target trial may be costly, infeasible, unethical, or simply untimely. We resort to causal analyses of observational databases only because the randomized trial that would answer our causal question—the target trial—either cannot be conducted or cannot be conducted in time to inform decisions that need to be urgently made.

An implication of the above is that, for the purposes of decision making, we can regard causal inference from observational databases as an attempt to emulate a target trial. If the emulation is successful, the analysis of the observational data is expected to yield the same effect estimates as the target trial would have yielded had the latter been conducted. Note that, in general, observational data can only be used to emulate pragmatic target trials (e.g., no placebo control, no blinding, no intensive monitoring).

Though the concept of a target trial is implicit in many causal analyses of observational databases, the target trial itself is not always explicitly characterized. A typical consequence of the lack of explicit characterization of the target trial is the inability to evaluate the adequacy of the observational analyses: in the absence of a precise definition of the causal question, it is difficult to assess the validity of the answer.

More generally, all research on the comparative effectiveness and safety of clinical interventions has two key steps: (1) asking the causal question, and (2) answering the causal question. Step 1 can be precisely carried out by designing the target trial, and Step 2 by either conducting the target trial (when possible) or by explicitly emulating it using observational data. Both steps are important for valid causal inference from observational data because the choice of procedures for Step 2 follow naturally from the explicit causal questions articulated in Step 1. Neglecting Step 1 can lead to overreliance on statistical analyses that are difficult to interpret causally, which in turn leads to mistakes that put into question the validity of observational analyses.[9-11]

This document outlines a framework to study the cardiovascular effects of second-line antihyperglycemic agents in individuals with type 2 diabetes. The framework revolves around the explicit description and emulation of a target trial. The next section outlines the protocol of a pragmatic trial of second-line antihyperglycemic agents, that is, it articulates the causal question (Step 1). The last section describes the procedures and data required for the emulation of the target trial (Step 2).

The target trial described in this document is one of several target trials that could be reasonably proposed to estimate the effects of second-line antihyperglycemic agents. It would not be surprising if informed researchers wished to propose justified variations of, say, the trial's eligibility criteria or treatment strategies. Therefore, this document should not be viewed by researchers as a rigid description of the target trial that must be emulated, but rather as guidance on how to (1) structure the specification of the target trial that precisely characterizes the research question, (2) justify the adequacy of the available observational data to emulate the target trial, and (3) describe the data analyses to carry out the emulation. The intent of this document is informative, not prescriptive.

Step 1: Specification of the Protocol of the Target Trial

As discussed above, none of the existing randomized trials of recently developed second-line antihyperglycemic agents can provide adequate information on their comparative effectiveness and safety regarding cardiovascular outcomes. Here, we propose the design of a target trial that would provide the information of interest if it were conducted. This target trial combines the pragmatic design and pharmacological breadth of the GRADE trial with the ability to detect cardiovascular effects of the other randomized trials. It also includes individuals with diabetes across a broad range of cardiovascular profiles.

Because such a target trial would be costly, difficult to perform, and would take many years to complete, it is unlikely to be conducted. As a result, we will need to use observational databases to emulate it. Remember that, in general, observational data can only be used to emulate a pragmatic trial, that is, one in which treatment strategies are compared under the usual conditions in which they are applied in the real world. Therefore, we cannot emulate a placebo-controlled trial with blind assignment and tight monitoring and enforcement of adherence to the study protocol. The necessarily pragmatic nature of the target trial is not a limitation if the goal is comparing the effects of realistic treatment strategies in individuals who participate in decisions about their care.

The specification of the protocol of the target trial involves several key components: eligibility criteria, treatment strategies, treatment assignment, start and end of follow-up, outcomes, causal contrasts, and statistical analysis plan. A description of each of these components follows.

Eligibility criteria. Individuals in the target trial will meet the following criteria:

- Type 2 diabetes mellitus
- Age ≥45 years (guidelines recommend diabetes screening starting at age 45)
- Antihyperglycemic monotherapy with metformin for at least 3 months in the year 2014 or later. Prior use of other glucose-lowering medications, including insulin, is allowed if (1) use was limited to 1 prescription of less than 30 days no later than 6 months before baseline (as defined in "Assignment to Treatment Strategies" section below), or (2) use took place during a hospitalization.
- Suboptimal glycemic control, as defined below (based on ADA 2018 guidelines[12] and the GRADE trial[7])
  - Two A1C measurements between ≥7% and <9% within a 6-month period after the initial 3-month period (a confirmed A1C ≥7% is the definition of primary failure in the GRADE trial)
  - A single A1C measurement between ≥9% and <11% after the initial 3-month period, which is not followed by insulin initiation.
- Not currently pregnant
- No history of the following conditions in the year prior to initiation of second-line therapy: type 1 diabetes, congestive heart failure, pancreatitis, cancer (other than non-melanoma skin cancer), severe liver disease or alcoholism, serious anemia, severe chronic kidney disease (e.g., estimated glomerular filtration rate [eGFR] <30) and end-stage renal disease, or conditions that require glucose-lowering medications.
- Expectation that the individual will remain actively engaged with the healthcare system that will be used to ascertain the outcomes (defined below)
  - This eligibility criterion might be operationalized by restricting eligibility to individuals who have been in regular contact with the healthcare system before baseline (e.g., those who attended regular check-ups or filled any

prescriptions within the 2 previous years) in the hope that they will remain in contact thereafter.
  - Note that individuals whose claims are no longer found in the database at some time after baseline (perhaps because they changed insurers) cannot be simply excluded. Rather, such individuals are considered initially eligible and later lost to follow-up.

History of bariatric surgery need not be an exclusion criterion, because individuals who underwent surgery and remained diabetic would be treated in the same way as other people with diabetes, and individuals who underwent surgery and were no longer diabetic would not meet the eligibility criteria for the trial.

History of myocardial infarction, stroke, or other major cardiovascular events is not an exclusion criterion because including individuals with both high and low or moderate cardiovascular risk will allow investigators to explore whether the treatment effects vary between these groups.

Treatment strategies. Eligible individuals included in the target trial will be assigned to one of the following 4 treatment strategies within 12 months of suboptimal glycemic control:

1. Initiation of a GLP-1 receptor agonist
2. Initiation of an SGLT2 inhibitor
3. Initiation of a DPP-4
4. Initiation of a sulfonylurea

Like in the GRADE trial, thiazolidinediones (TZDs) are not included because of increased risk of heart failure (rosiglitazone, pioglitazone) and bladder cancer (pioglitazone).

A clinically attractive alternative is to define multiple strategies, each of them defined by individual drugs rather than classes of drugs. If implemented, this alternative definition of treatment strategies requires a sufficient number of individuals initiating each of the individual drugs under study (see detail regarding sample-size considerations on p. 14).

Under all strategies, the protocol dictates:

- Treatment to be initiated within 3 months of prescription (e.g., as determined by billing date).
- A1C to be monitored at least once every 6 months.
- Assigned drug can be discontinued and replaced by a drug from another class or insulin if
    - toxicity arises
    - clinical developments (e.g., drop in glomerular filtration rate [GFR]) make treatment switching advisable
        - Importantly, decisions to discontinue based on toxicity and clinical developments need to be operationalized in detail and several alternatives pre-specified before starting the emulation procedure using observational data.
- In case of lack of efficacy, defined as 2 A1C measurements >8% within 6 months (threshold can be 8.5% for older, frailer people), the following is allowed:
    - Add basal insulin
    - In GLP-1 arm, add sulfonylurea or SGLT2 inhibitor
    - In DPP-4 arm, add sulfonylurea or SGLT2 inhibitor
    - In sulfonylurea arm, add GLP-1, SGLT2 inhibitor, or DPP-4.
- The combination DPP-4 and GLP-1 is not allowed.
- No bariatric surgery is allowed during the follow-up.
- The use of concomitant treatments (e.g., antihypertensives, statins, etc.) will be left to the treating physician's discretion.

Assignment to treatment strategies. Eligible participants recruited into the target trial will be randomly assigned to one of the above strategies. Both participants and their treating physicians will be aware of the assigned treatment strategy (e.g., non-blinded design).

Follow-up. For each individual, follow-up starts at the time of treatment prescription (baseline) and ends at death, diagnosis of the outcome, loss to follow-up, or the administrative end of follow-up. Individuals would start to be recruited in 2014, the year during which the studied treatments were all first available.

Loss to follow-up is defined as any event that prevents (1) the ascertainment of the outcome, (2) the ascertainment of adherence to the assigned treatment strategy (for per-protocol analyses; see below), or (3) the measurement of post-baseline prognostic factors.

Outcomes. The following outcomes will be ascertained during the follow-up via contact with the healthcare system:

- 3-point major adverse cardiovascular event (MACE) or similar composite outcome
- Myocardial infarction
- Stroke
- Hospitalization due to congestive heart failure
- Cardiovascular death
- All-cause mortality
- Severe hypoglycemia (based on ER or hospital admissions)
- Microvascular disease (retinopathy, nephropathy, amputation)
- Renal function (e.g., progression to a higher stage of chronic kidney disease based on GFR, appearance of albuminuria)

Causal contrasts. For each of the outcomes listed above, the causal effects of interest are the intention-to-treat effect and the per-protocol effect.

- *Intention-to-treat effect*: the effect of being assigned to the treatment strategies at baseline, regardless of treatment actually received. In trials in which initiation of the

strategies co-occurs with randomization, the intention-to-treat effect is the effect of initiation of the strategy. The numerical value of the intention-to-treat effect from a given trial depends on the particular patterns of deviation from protocol that occur during the conduct of the trial. Two trials of the same treatment strategies, conducted in the same population, could have different intention-to-treat effects if adherence patterns differed, and both would be internally valid effects of assignment to treatment.

- *Per-protocol effect:* the effect of receiving the treatment strategies throughout the follow-up as specified in the study protocol. The numerical value of the per-protocol effect from a given trial does not depend on the particular patterns of deviation from protocol that occur during the conduct of the trial. Two trials of the same treatment strategies, conducted in the same population, could have different intention-to-treat effects if adherence patterns differed, but would have the same per-protocol effect.

Intention-to-treat effect estimates are difficult to interpret for patients and clinicians because the intention-to-treat effect is agnostic to any treatment decisions made after baseline, including discontinuation or initiation of the treatment strategies of interest, use of concomitant therapies, or any other deviations from protocol. Additionally, in pragmatic trials, the intention-to-treat effect will typically be a combination of the effect of the treatment under study and of any other patient and physician's behavioral changes triggered by the assignment itself. Despite this and other shortcomings of the intention-to-treat effect for pragmatic trials,[13] this effect is widely used because the randomized assignment makes us expect no confounding for the effect-of-treatment assignment. Therefore, statistical analyses to estimate intention-to-treat effects are rarely controversial.

The per-protocol effect is a more patient-centered alternative to the intention-to-treat effect, but it is harder to estimate (see below). The per-protocol effect is closer than the intention-to-treat effect to the sort of effect that patients are mostly interested in learning from pragmatic trials,[14] and in fact the per-protocol effect is often the implicit

target of inference for investigators, too. When the intention-to-treat effect is said to be biased toward the null or conservative, the implication is that the intention-to-treat effect is a biased estimate of the per-protocol effect. An added advantage of the per-protocol effect is that its interpretation does not depend on a trial-specific degree of adherence, which makes it a more transportable effect.

Statistical analysis plan. Two separate sets of statistical analyses will be conducted, one to estimate the intention-to-treat effect (intention-to-treat analysis) and another one to estimate the per-protocol effect (per-protocol analysis). Additionally, the statistical analysis plan will include pre-specified subgroup analyses in subsets of patients defined by

- Age and gender
- Cardiovascular risk group

We now outline the data analysis procedures for the target trial. (Reminder: What follows is the description of the data analysis for a randomized trial. In the next section, we describe the modifications required to emulate these analyses using observational data.)

Intention-to-treat analysis

We refer to the analysis aimed at estimating the intention-to-treat effect as an intention-to-treat analysis. In a large pragmatic trial with complete follow-up and no competing events for the outcome, the intention-to-treat analysis is straightforward: a comparison of the observed outcome distribution between groups defined by assignment to a particular treatment strategy, regardless of whether the individuals adhered to their assigned strategy.

In our trial, the primary analysis will consist of constructing, for each outcome, the survival curve (or, equivalently, the cumulative incidence curve) in each treatment group. If baseline covariates need to be adjusted (see below), these curves can be parametrically estimated using a multiplicative hazards model with product terms between time and indicator for treatment strategies.[15,16]

To compare risks at pre-specified time points (e.g., 2 years, 5 years) after baseline, researchers can estimate the intention-to-treat risk ratio or difference as the ratio or difference of the corresponding mortality risks in those assigned to a given treatment strategy and to the reference strategy (say, initiation of a sulfonylurea). The analysis can also include the fit of a Cox proportional hazards model to estimate the intention-to-treat hazard ratios averaged over the follow-up. For both intention-to-treat and per-protocol analysis, researchers need to propose and justify an approach to handle competing events, i.e., death for non-fatal outcomes.

Intention-to-treat analyses do not require adjustment for non-adherence, but they may require adjustment for random baseline imbalances in risk factors[17] and for differential loss to follow-up between treatment groups.[18] The adjustments are defined as follows:

- **Imbalanced prognostic factors at baseline:** Even randomization cannot ensure that the distribution of all prognostic factors is identical between the trial arms. As a result of these chance imbalances, the effect estimate may be far from the true effect. Some authors refer to the difference between a study-specific effect estimate and the true effect as random confounding[19] because, in a particular study, the problem is empirically indistinguishable from the systematic confounding that results from systematic imbalances of prognostic factors in observational studies. Dealing with any imbalances in intention-to-treat analyses, regardless of whether they are random or systematic, requires adjustment for the imbalanced prognostic factors.[17] Researchers will pre-specify important prognostic factors for adjustment or pre-specify a maximum acceptable imbalance which, if exceeded, will trigger adjustment for that factor.

- **Loss to follow-up:** Intention-to-treat analyses require that the outcomes of all trial participants are ascertained. If some participants are lost to follow-up and therefore their outcomes are unknown, a true intention-to-treat analysis is not possible. Censoring individuals when they are lost to follow-up—a pseudo-intention-to-treat analysis—may introduce selection bias

in the effect estimates when the censoring depends on pre- and post-randomization prognostic factors. Therefore, the pseudo-intention-to-treat analysis needs to adjust for those pre- and post-randomization prognostic factors to provide valid estimates of the intention-to-treat effect, that is, the effect of treatment assignment had all individuals being fully followed.

To preserve the marginal (unconditional) interpretation of intention-to-treat effect estimates in the study population, adjustment for baseline covariates is better carried out via standardization or inverse probability (IP) weighting. Because post-randomization prognostic factors may be affected by treatment assignment, adjustment for time-varying covariates typically is better carried out via IP weighting. For each individual, the denominator of the time-varying IP weight at each time is, informally, his or her probability of remaining uncensored through that time conditional on the pre- and post-baseline prognostic factors that predict loss to follow-up. Because these IP weights need to be estimated from the data, the implementation of the method requires that post-randomization factors are measured regularly over follow-up to ensure that data will be available close to the time of censoring.

In summary, the adjusted intention-to-treat analysis will estimate the hazard ratios and survival curves (standardized by baseline variables, if necessary) after adjustment for prognostic factors associated with loss-to-follow-up (via IP weighting).

Per-protocol analysis

We refer to the analysis aimed at estimating the per-protocol effect as a per-protocol analysis. In general, per-protocol analyses require adjustments for non-adherence,[20] in addition to the adjustments required for intention-to-treat analyses. Because adherence is not randomly assigned, per-protocol analyses are effectively observational analyses of randomized trials. Note that naïve per-protocol analyses that do not appropriately adjust for pre- and post-baseline prognostic factors associated with adherence are not generally valid.[13]

We now outline two options for (non-naïve) per-protocol analyses for treatment strategies that, as in our trial, are sustained over time:

- Censoring at deviation from protocol: Individuals assigned to a particular treatment strategy are censored if or when they deviate from the strategy. This censoring may introduce post-baseline selection bias when time-varying prognostic factors for the outcome are also predictors of adherence. Adjustment for these prognostic factors, when measured, may be achieved via IP weighting. For each individual, the denominator of the time-varying IP weight at each time is, informally, the probability of adhering to his or her assigned strategy through that time conditional on the pre- and post-baseline prognostic factors that predict adherence. The data analysis is then analogous to the intention-to-treat analysis, except that individuals are censored if or when they deviated from the study protocol and uncensored individuals are IP weighted.[21]

- G-formula: The parametric g-formula, a generalization of standardization for time-varying treatments and confounders, also estimates the risk that would have been observed if all patients in the study had adhered to a particular treatment initiation strategy and none had been lost to follow-up, under the assumptions of no residual confounding, no measurement error, and no model misspecification.[22,23] The estimation procedure has two steps. First, regression models are used to estimate the joint distribution of the outcome, treatment, and time-varying covariates conditional on previous treatment and covariate history. Second, a Monte Carlo simulation is run to simulate the distribution of the post-baseline outcomes and time-varying covariates separately under each treatment strategy.

A nonparametric bootstrap procedure can be used to obtain percentile-based 95% confidence intervals when using either IP weighting or the g-formula. Alternatives to these approaches exist but would need to be carefully justified. For example, traditional propensity score matching and related techniques,[24] as well as traditional instrumental variable estimation,[25,26] cannot generally handle time-varying treatments in the presence of treatment-confounder feedback (e.g., past A1C levels affect current treatment choice and current treatment choice affects future A1C levels) and should therefore be avoided when treatment-confounder feedback is expected. Marginal structural models with dose-response functions of treatment may be very sensitive to the functional form[27] and are not well suited to the comparison of dynamic strategies.[28]

Regardless of the statistical method used to adjust for incomplete adherence in per-protocol analyses, the implementation of the method requires that longitudinal data will be available to determine whether individuals adhere to their assigned treatments strategies, as well as to adequately adjust for determinants of adherence. Specifically, discontinuation of a treatment or initiation of a treatment not assigned at baseline does not imply that the individual is deviating from protocol[20] as there are protocol-approved reasons for treatment discontinuation (e.g., toxicity), and for initiation of other treatments (e.g., lack of efficacy).

Variations. Several components of the target trial may be modified in ways that differ from the design described in the "Intention-to-Treat" and "Per-Protocol" sections above. Investigators will need to weigh whether the added complexity brought by any variations are worthwhile in terms of added information. For example, consider the following modifications:

- The eligibility criteria might include perceived ability to adhere to all of the treatments under study. Some possible attempts to operationalize this criterion are:
  - Include individuals who are likely to adhere to treatment. Since no direct measure of adherence may exist, a proxy is needed. For example, these individuals may be partly identified using dispensing data as those who have a history of filling prescriptions to other treatments (e.g., metformin).
  - Avoid individuals who received a particular second-line treatment because of absolute or relative contraindications to other treatments.

The A1C threshold to define suboptimal glycemic control might be individualized, as suggested by ADA 2018 guidelines,[29] based on clinical characteristics such as duration of diabetes, life expectancy, important comorbidities, advanced microvascular complications, and established vascular complications.

- The treatment strategies might replace universal initiation of treatment at baseline by initiation of treatment after baseline triggered by reaching certain A1C thresholds or developing certain clinical events.

- Additional subgroup analyses can be conducted in special groups of patients for whom clinical information is limited, such as users of oral or systemic glucocorticoids and users of atypical antipsychotics.

- In addition to intention-to-treat effects and per-protocol effects, there may be interest in estimating quasi-per-protocol effects in which the treatment strategies are defined under some pre-specified level of suboptimal adherence to the protocol.

Step 2: Emulation of the Target Trial

After the causal question has been articulated via specification of the protocol target trial (Step 1), investigators need to describe how to emulate each of the components of the protocol using an observational healthcare database (Step 2).

Bear in mind that the process of specifying and emulating the target trial is an iterative process because it is possible that no single data source will include all information required to emulate the target trial described in the previous section. When some data items are not available, the specification of the protocol of the target trial needs to change in such a way that those data items are no longer necessary for its emulation. For example, if information on an individual's age were not available, then the target trial would need to refrain from defining eligibility based on age. The final version of the protocol of the target trial is likely to be a compromise that differs from the originally proposed target trial but that can be reasonably emulated using the available

observational data. In that case, investigators need to justify why the modified target trial is still of interest and/or the emulation procedure is expected to succeed.

Note that, because the specific operationalization of some aspects of the emulation is heavily dependent on the characteristics of the proposed data source, the specific emulation procedures need to be developed and defended by the investigators. Some guidance on those procedures to emulate the target trial of the previous section follows, together with a description of data requirements to emulate each component of the protocol of the target trial.

Eligibility criteria. Individuals in the emulated trial will meet the eligibility criteria of the target trial. Identifying these individuals requires adequate information on demographic and clinical characteristics, medical treatments, and lab values to assess glycemic control. Additionally, researchers may consider adding medical insurance status that provides access to all treatments under study as an eligibility criterion, or else providing a justification of why the comparison of groups with differential access to medical care is not a source of bias.

Data Requirements for Emulation. To identify individuals in the database who meet the target trial's eligibility criteria, information on the following variables is needed at baseline and during the year before baseline:

- Age
- Type 2 diabetes mellitus diagnosis (Note that there may be no need to develop an algorithm to identify diagnoses of type 2 diabetes if the other data items are available because everyone on metformin with poor glycemic control can be safely assumed to have type 2 diabetes.)
- Use of metformin and other glucose-lowering medications, including insulin
- Hospitalizations and pharmacy claims
- A1C values
- At least one year of pre-baseline data

to ascertain history of type 1 diabetes, congestive heart failure, pancreatitis, cancer (other than non-melanoma skin cancer), severe liver disease or alcoholism, serious anemia, chronic kidney disease, end-stage renal disease, or conditions that require glucose-lowering medications.

- Pregnancy status
- Absolute or relative contraindications to any antihyperglycemic agents

Handling Missing Data. No missing data are allowed for eligibility criteria. Like the target trial, the emulated trial will include only individuals whose baseline eligibility can be characterized using the available data. In sensitivity analyses, researchers may consider alternative sets of eligibility criteria.[30]

Treatment strategies. Eligible individuals will be assigned to one (or none) of the treatment strategies studied in the target trial. The emulation requires sufficient information to determine when the individuals' data is consistent with initiation of, and adherence to, the treatment strategies. Specifically, to assign individuals to treatment strategies and determine whether they follow their assigned strategies, information on the following variables is needed at baseline and during the follow-up:

Data Requirements for Emulation. The data items can be classified in two groups:

- Data items needed to determine initial assignment to a treatment strategy:
    - Prescription and dispensing dates of GLP-1 receptor agonists, SGLT2 inhibitors, DPP-4, sulfonylureas, insulin, and metformin.
- Data items needed to determine adherence to a treatment strategy during the follow-up:
    - Discontinuation dates of GLP-1 receptor agonists, SGLT2 inhibitors, DPP-4, sulfonylureas, insulin, and metformin. To estimate the date of discontinuation,

researchers may use the expected prescription duration or another approach; sensitivity analyses may be needed.

- Date of bariatric surgery
- A1C values at least once every 6 months
- Determination of drug toxicity and clinical developments (e.g., drop in GFR) that make treatment switching advisable

Reminder: Decisions to discontinue based on toxicity and clinical developments need to be operationalized in detail and several alternatives pre-specified before starting the emulation procedure using observational data.

Handling Missing Data. No missing data are allowed for determination of initial assignment to a treatment strategy. The emulated trial will include only individuals whose initial treatment strategy can be characterized using the available data. Imputation of initial treatment assignment is discouraged as the availability of these variables is crucial for the validity of the emulation.

It is possible that, during the follow-up, the data stream of some individuals is interrupted in such a way that their adherence to a treatment strategy can no longer be determined. In per-protocol analyses, this form of missing data for time-varying variables will be considered loss to follow-up and handled as described above for the target trial.

Assignment to treatment strategies. Unlike in the target trial, eligible individuals in the emulated trial are not randomly assigned to one of the above strategies. To emulate the random assignment of strategies at baseline, researchers need to adjust for all important baseline confounders, that is, pre-baseline prognostic factors that predict initiation of each of the treatment strategies. Full adjustment for these variables would largely erase the difference between a pragmatic trial and an observational analysis of a healthcare database. On the other hand, if the observational database does not

contain sufficient information on baseline confounders, successful emulation of the target trial's random assignment is not possible and confounding bias will arise. A list of potential confounders follows.

Data Requirements for Emulation. A fundamental problem of causal inference from observational data is that confounders may be unknown or unavailable. The following factors affect drug choice (GLP-1 vs SGLT2 inhibitors vs DPP-4 vs sulfonylureas) and therefore should be initially considered:

- Type of medical insurance. This variable may be partly controlled through restriction via eligibility criteria. Note that, even in the absence of changes in insurance status, changes in access to various drugs and services may occur over time.

- Ability to adhere to the treatment strategies. This variable may be partly controlled through restriction via eligibility criteria.

- Frailty. This variable may be partly captured by intensity of medical care in the year before baseline. (Note: The eligibility criteria may need to include restrictions based on frailty.)

- Race or ethnicity, socioeconomic status. These factors can be partly adjusted via zip codes or other residential markers.

- Cardiovascular risk factors, including history of cardiovascular events. There seems to be expert agreement that cigarette smoking does not influence treatment choice so this variable may not be necessary for confounding adjustment.

- Body mass index

- Perceived risk of hypoglycemia (history of hypoglycemia, chronic kidney disease, older age, other comorbidities)

- Physician comfort level with the various antihyperglycemic agents that may be partly captured by specialty (endocrinologist vs primary care provider) and past prescription history and practice pattern.

- Any additional prognostic factors that drive prescribing behavior

Some researchers may believe that some of the above variables are unnecessary for a successful emulation of the randomized assignment. In that case, they are encouraged to present evidence that justifies their belief. Also, researchers may propose additional baseline confounders that have not been specified by name in the list.

Handling Missing Data. No missing data are allowed for the proposed baseline confounders. The emulated trial will include only individuals whose baseline confounders can be characterized using the available data. Imputation of baseline confounders is discouraged as the availability of these variables is crucial for the validity of the emulation. In sensitivity analyses, researchers may consider alternative sets of confounders.

Follow-up. For each individual, follow-up starts at the time of treatment prescription (baseline) and ends at death, diagnosis of the outcome (see below), loss to follow-up, or the administrative end of follow-up. An explicit definition of loss to follow-up needs to be provided. For example, in electronic health records, loss to follow-up can be defined as the first time when ascertainment of outcome and determinants of loss to follow-up is compromised (e.g., the individual exits the healthcare system that maintains the database) or, for per-protocol analyses, when ascertainment of adherence and post-baseline confounders is compromised (e.g., loss of pharmacy benefits).

Data Requirements for Emulation. The start of follow-up is determined by the eligibility criteria and assignment to treatment strategies discussed above. To determine the end of follow-up, the following data items are required:

- Date of death

- Date of diagnosis of other outcomes (defined below)

- Date of discontinuation of medical insurance, pharmacy benefits, or any other interruptions of any

data streams that are needed for a successful emulation.

Handling Missing Data. It is possible that, during the follow-up, the data stream of some individuals is interrupted in such a way that the above dates can no longer be determined. This form of missing data will be considered loss to follow-up and handled as described above for the target trial. Depending on the data stream being interrupted (e.g., outcome diagnosis vs prescription data), the loss to follow-up may apply to the per-protocol analyses but not to intention-to-treat analyses.

Outcomes. To determine when an individual develops one of the outcomes of interest, researchers will rely on routine procedures for clinical diagnosis in the healthcare system used for the emulation.

Data Requirements for Emulation. The date of occurrence (if any) of the following variables throughout the follow-up is needed:

- 3-point MACE or similar composite outcome
- Myocardial infarction
- Stroke
- Hospitalization due to congestive heart failure
- Cardiovascular death
- All-cause mortality
- Severe hypoglycemia (based on emergency room or hospital admissions)
- Microvascular disease (retinopathy, nephropathy, amputation)
- Renal function (e.g., progression to a higher stage of chronic kidney disease based on GFR, appearance of albuminuria)

Investigators are expected to provide information that supports the validity of their outcome ascertainment procedures.

Handling Missing Data. It is possible that, during the follow-up, the data stream of some individuals is interrupted in such a

way that the above diagnoses can no longer be determined. This form of missing data will be considered loss to follow-up and handled as described above for the target trial. The definition of loss to follow-up may vary across analyses that consider different outcomes. For example, if all-cause mortality is ascertained through the National Death Index rather than through the healthcare system, then individuals in an intention-to-treat analysis of all-cause mortality would not be considered lost to follow-up when their medical insurance is interrupted, but the same individuals in an analysis of stroke will be considered lost to follow-up when their medical insurance, and therefore their data stream in the database, is interrupted.

Causal contrasts. For each of the outcomes listed above, the causal effects of interest are the observational analogs of the intention-to-treat effect and the per-protocol effect in the emulated trial.

- In the target trial, the Intention-to-treat effect is the effect of being assigned to the treatment strategies at baseline, regardless of treatment actually received. However, when emulating a trial using observational data, the concept of assignment to treatment is not the same as in a true randomized trial. The preferred option is to define assignment as prescription of therapy at baseline, regardless of whether the prescription was actually filled. In the absence of prescription data, an alternative is to define assignment as dispensing of therapy at baseline, that is, to compare initiators vs non-initiators. A comparison of initiators parallels the intention-to-treat effect in a target trial in which assignment and initiation of the treatment strategies always occur together at baseline, regardless of whether individuals continue on the strategies after baseline.

- In the target trial, the per-protocol effect is the effect of receiving the treatment strategies throughout the follow-up as specified in the study protocol. Once assignment is defined, the per-protocol effect in the trial emulated using observational data is analogous to the per-protocol effect in the target trial.

Data Requirements for Emulation. See Treatment Strategies above.

Handling Missing Data. Not applicable.

Statistical analysis plan. Two separate sets of statistical analyses will be conducted, one to estimate the observational analog of the intention-to-treat effect (intention-to-treat analysis) and another one to estimate the observational analog of the per-protocol effect (per-protocol analysis):

- **Intention-to-treat analysis.** The analysis of the emulated trial will be the same as the intention-to-treat analysis described for the target trial, with the exception that adjustment for multiple baseline confounders is expected to adjust for systematic imbalances in prognostic factors at baseline. The adjustment for baseline confounders may be performed via matching (perhaps on the propensity score), stratification or regression, standardization or inverse probability weighting, g-estimation, or doubly robust methods. Note that the confounders may differ depending on whether assignment is defined as prescription or dispensing.

- Per-protocol analysis. The analysis of the emulated trial will be the same as the per-protocol analysis described for the target trial.

The same subgroups as in the target trial will be considered for subgroup analyses.

Data requirements for emulation. Information on two sets of adjustment factors will be needed to emulate the estimation of the intention-to-treat effect and the per-protocol effect:

- Pre- and post-baseline prognostic factors that predict loss to follow-up. These are variables that would also need to be measured and adjusted for in a truly randomized trial to validly estimate the intention-to-treat effect. It may be reasonable to assume that these prognostic factors largely overlap with factors that predict treatment initiation (see above) and adherence to the protocol (see below).

- Pre- and post-baseline prognostic factors that predict adherence to the protocol. These are confounders that would also need to be measured and adjusted for in a truly randomized pragmatic trial to validly estimate the per-protocol effect. These factors include the baseline confounders listed above and their post-baseline, time-varying values. Some potentially key time-varying confounders are:

  - A1C
  - Blood pressure
  - Cardiovascular events (there may be switches to medications with perceived cardiovascular benefit) for analyses in which these events are not the outcome of interest
  - Other comorbidities

Additionally, data on age, gender and cardiovascular risk group is necessary for the subgroup analyses.

Handling Missing Data. It is possible that, during the follow-up, the data stream of some individuals is interrupted in such a way that the values of the above time-varying confounders can no longer be determined. This form of missing data will be considered loss to follow-up and handled as described above for the target trial. In sensitivity analyses, researchers may consider alternative sets of time-varying confounders.

Additional sensitivity analyses

The success of the emulation procedure relies on the assumption that the most important confounders have been adequately measured and adjusted for. Unfortunately, this assumption is not empirically verifiable so it is generally impossible to determine whether the emulation failed because of uncontrolled confounding. However, indirect analytic approaches may alert researchers to possible unmeasured confounding. Some examples of these approaches are suggested below:

- **Benchmarking with existing randomized trials.** As an extension of the previous

approach, the existing placebo-controlled randomized trials may be used as a benchmark for some cardiovascular outcomes in individuals at high cardiovascular risk, and the GRADE pragmatic trial may be used as a benchmark for non-cardiovascular outcomes in individuals at low to moderate cardiovascular risk. Incompatible effect estimates would suggest that the emulation failed.

- **Emulating of a target trial with reversed strategies.**[31] That is, rather than emulating a target trial of initiation of a particular antihyperglycemic agent, researchers may emulate a target trial in which users of the agent are assigned to the strategies of continue using therapy or stop using therapy. Incompatible or surprising effect estimates (e.g., a decreased risk both when initiating therapy in the original target trial and when discontinuing therapy in the reversed target trial) would suggest that at least one of the emulations failed to ensure a fair comparison.

- **Use of negative outcome controls[32] (falsification outcomes):** The treatments under study are not expected to have a causal effect on some non-trial outcomes (e.g., respiratory diseases). However, some confounders for the trial outcomes (e.g., access to medical care) may be confounders for the non-trial outcomes too. Therefore, if confounding adjustment were incomplete, an association between the drugs under study and the non-trial outcomes would be expected to arise. As a result, a lack of association between the studied treatments and the non-trial outcomes (after adjustment) would indirectly support that those confounders were successfully adjusted for the trial outcomes. Investigators are encouraged to suggest and justify the use of particular negative controls.

- **Use of positive outcome controls.** Since some of the drugs under study have well-known effects on certain clinical outcomes, identifying these effects in the observational analysis will support the validity of the target trial emulation. These outcomes may include body weight (e.g., weight loss expected for SGLT2 inhibitors and GLP1, weight gain

expected for sulfonylureas), blood pressure (less need for antihypertensive treatment is expected for some drugs but not others), and specific infections (for SGLT2 inhibitors), as well as the expected differences in A1C reduction among drugs. Investigators are encouraged to suggest and justify the use of particular positive controls.

- **Instrumental variable methods** for the estimation of intention-to-treat analyses in subsets of the population of interest, if a compelling instrument can be proposed and a monotonicity assumption can be reasonably made.[17] Traditional instrumental variable analyses cannot be used to compare treatment strategies sustained over time and therefore cannot be used to estimate the per-protocol effect in the emulated trial.

- **Formal bias analysis,** including sensitivity to unmeasured confounding, selection bias, and measurement error, to estimate bounds of the causal effects under realistic assumptions.[33] When some confounders are not generally available in the proposed data source but are available in a subsample, estimation of bounds of the effect if all confounders had been available in the data source.

Selection of an appropriate data source

Considerations about how to select a database, or set of linked databases, for emulation of the target trial need to consider both the quantity (sample size and length of follow-up) and richness of the available data.

- **Sample size.** The selected data source needs to include a sufficient number of eligible individuals to allow for precise estimation of absolute risks at pre-specified times after baseline in subgroups defined by age, gender, and cardiovascular risk. A formal sample-size calculation can be conducted to calculate the minimum number of eligible individuals to achieve a particular precision.[34] For intention-to-treat analyses, the calculation is identical to the ones usually proposed for the design of randomized trials (which require informed estimates informed estimates about incidence of the outcome, losses to follow-

up, etc.) with the exception that there also must be accommodation for the magnitude of confounding (which may be estimated by inspection of the literature). For per-protocol analyses, additional information about adherence patterns is also needed. Given the potential complexity of these sample-size calculations, researchers may want to approach them via simulations informed by the characteristics of their proposed data source.

- **Length of follow-up.** Emulation of this target trial requires a data source with information on the required data items after 2014 (when the studied treatments became widely available). Because determination of eligibility requires a full year before baseline, data from 2013 can be used for this purpose.

- **Data richness.** The selected data source needs to include longitudinal information on the multiple data domains described above. These domains include drug prescription and dispensing, clinical events and comorbidities, death and causes of death, laboratory values (e.g., A1C) and other clinical parameters (e.g., blood pressure, body mass index), insurance status, and others.

Not all available health databases contain sufficient information (in terms of either quantity or richness of data) to tackle the target trial emulation. Specifically, databases of administrative insurance claims, though widely used in large-scale health research, seem ill-suited for the purposes of emulating this target trial. While these databases often contain adequate information on prescriptions and clinical diagnoses, they lack information on the clinical parameters that are needed to characterize eligibility (e.g., A1C levels) and adjust for confounding (e.g., blood pressure, body mass index) as well as on dispensing of treatments. Additionally, many U.S. commercial databases of insurance claims have high rates of loss to follow-up as people often switch membership across health plans.

Given all of the above, integrated systems that combine insurance claims, electronic medical records, and pharmacy records appear to be the best option for emulation of a target trial on antihyperglycemic therapy and cardiovascular risk.

Disclaimer

References

1.   Centers for Disease Control and Prevention. National Diabetes Statistics Report. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services; 2017.

2.   American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care* 2018;doi.org/10.2337/dci18-0007.

3.   Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med* 2015;373:2117-28.

4.   Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med* 2017;377:644-57.

5.   Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 2016;375:311-22.

6.   Marso SP, Bain SC, Consoli A, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med* 2016;375:1834-44.

7.   Nathan DM, Buse JB, Kahn SE, et al. Rationale and design of the glycemia reduction approaches in diabetes: a comparative effectiveness study (GRADE). *Diabetes Care* 2013;36:2254-61.

8.   Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology* 2016;183:758-64.

9.   Emilsson L, García-Albéniz X, Logan RW, Caniglia EC, Kalager M, Hernán MA. Examining bias in studies of statin treatment and survival in patients with cancer. *JAMA Oncology* 2018;4:63-70.

10.   Hernán MA, Robins JM. Observational studies analyzed like randomized experiments: Best of both worlds. *Epidemiology* 2008;19:789-92.

11.   Cain LE, Logan R, Robins JM, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med* 2011;154:509-15.

12.   American Diabetes Association. 8. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2018. *Diabetes Care* 2018;41:S73-S85.

13.   Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9:48-55.

14.   Murray EJ, Caniglia EC, Swanson SA, Hernandez-Diaz S, Hernan MA. Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. *J Clin Epidemiol* 2018;103:10-21.

15.   D'Agostino RB, Lee M-L, Belanger AJ. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine* 1990;9:1501-15.

16.   Thompson WA, Jr. On the treatment of grouped observations in life studies. *Biometrics* 1977;33:463-70.

17.   Hernán MA, Robins JM. Causal Inference. Boca Raton: Chapman & Hall/CRC; 2019 (forthcoming).

18.   Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine* 2012;367:1355-60.

19.   Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur J Epidemiol* 2015;30:1101-10.

20.   Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *New England Journal of Medicine* 2017;377:1391-8.

21.   Murray EJ, Hernan MA. Improved adherence adjustment in the Coronary Drug Project. *Trials* 2018;19:158.

22.   Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period — Application to the healthy worker survivor effect [published errata appear in Mathl Modelling 1987;14:917-21]. *Mathematical Modelling* 1986;7:1393-512.

23.   Lodi S, Sharma S, Lundgren JD, et al. The per-protocol effect of immediate vs. deferred ART initiation in the START randomized trial. *AIDS* 2016;(in press).

24.   D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265-81.

25.   Angrist J, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996;91:444-55.

26.   Martens. Instrumental variables: applications and limitations. *Epidemiology* 2006;(in press).

27.   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.

28.   Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology* 2010;21:528-39.

29.   American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2018. *Diabetes Care* 2018;41:S55-S64.

30.   Schneeweiss S, Patrick AR, Sturmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Medical Care* 2007;45:S131-42.

31.   Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;19:766-79.

32.   Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383-8.

33.   Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;43:1969-85.

34.   Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. *Epidemiology* 2018;29:599-603.