Tell me what you want, what you really really want: estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies

Luijken, K.; Eekelen, R. van; Gardarsdottir, H.; Groenwold, R.H.H.; Geloven, N. van

Received: 20 December 2022 | Revised: 2 March 2023 | Accepted: 18 March 2023

DOI: 10.1002/pds.5620

ORIGINAL ARTICLE

WILEY

Tell me what you want, what you really really want: Estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies

Kim Luijken[1] | Rik van Eekelen[2] | Helga Gardarsdottir[3,4,5] | Rolf H. H. Groenwold[6,7] | Nan van Geloven[6]

[1]Department of Epidemiology, Utrecht University Medical Center, University Utrecht, Utrecht, The Netherlands

[2]Centre for Reproductive Medicine, Amsterdam University Medical Center, Amsterdam, The Netherlands

[3]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

[4]Department of Clinical Pharmacy, Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[5]Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland

[6]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[7]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

Correspondence
Kim Luijken, Universitair Medisch Centrum Utrecht, Utrecht 3508 GA, The Netherlands.
Email: k.luijken-2@umcutrecht.nl

Abstract

Purpose: Ideally, the objectives of a pharmacoepidemiologic comparative effectiveness or safety study should dictate its design and data analysis. This paper discusses how defining an estimand is instrumental to this process.

Methods: We applied the ICH-E9 (*Statistical Principles for Clinical Trials*) R1 addendum on estimands – which originally focused on randomized trials – to three examples of observational pharmacoepidemiologic comparative effectiveness and safety studies. Five key elements specify the estimand: the population, contrasted treatments, endpoint, intercurrent events, and population-level summary measure.

Results: Different estimands were defined for case studies representing three types of pharmacological treatments: (1) single-dose treatments using a case study about the effect of influenza vaccination versus no vaccination on mortality risk in an adult population of ≥60 years of age; (2) sustained-treatments using a case study about the effect of dipeptidyl peptidase 4 inhibitor versus glucagon-like peptide-1 agonist on hypoglycemia risk in treatment of uncontrolled diabetes; and (3) as needed treatments using a case study on the effect of nitroglycerin spray as-needed versus no nitroglycerin on syncope risk in treatment of stabile angina pectoris.

Conclusions: The case studies illustrated that a seemingly clear research question can still be open to multiple interpretations. Defining an estimand ensures that the study targets a treatment effect that aligns with the treatment decision the study aims to inform. Estimand definitions further help to inform choices regarding study design and data-analysis and clarify how to interpret study findings.

KEYWORDS
causal inference, comparative effectiveness and safety research, estimand, ICH-E9 (R1) addendum

Key Points

- Guidance for defining estimands focuses on randomized studies. We outline five elements that specify the estimand in observational pharmacological comparative effectiveness or safety studies.

- The five elements of an estimand are the population, contrasted treatments, endpoint, intercurrent events, and population-level summary measure.
- Three typical types of pharmacologic treatments are distinguished: (1) point treatments, administered in a single dose or for a (very) short duration; (2) sustained treatments, used to treat chronic (possibly lifelong) disease episodes; and (3) as-needed treatments, prescribed as needed based on disease symptoms.
- Defining an estimand clarifies the exact research question, helps to inform choices regarding study design and data-analysis, and clarifies how to interpret study findings.

Plain Language Summary

Ideally, the objectives of a pharmacoepidemiologic comparative effectiveness or safety study dictate its design and data analysis. Although a research question can seem clear, it may actually still leave room for multiple interpretations. By articulating a formal causal quantity that is the target of the analysis, denoted an *estimand*, the research question becomes more specific, and thus results of a study will have a clearer meaning.Guidance for defining estimands in randomized studies is given by the ICH-E9 (*Statistical Principles for Clinical Trials*) R1 addendum on estimands. The current work applied the ICH-E9(R1) addendum to three case studies of observational pharmacoepidemiologic comparative effectiveness and safety studies. Five key elements specify the estimand: the population, contrasted treatments, endpoint, intercurrent events, and population-level summary measure.The case studies illustrated that multiple estimands can be defined for a particular research question. Choosing an estimand clarifies the exact research question, helps to inform choices regarding study design and data-analysis, and clarifies how to interpret study findings.

1 | BACKGROUND

Ideally, medical decisions about the initiation of or changes in pharmacological treatment are supported by scientific evidence. To provide the information that is needed to inform such decisions, studies of causal effects of pharmacological treatments need to be correctly designed and conducted.[1–3] Different study designs and data-analytical approaches have been proposed to estimate causal effects of pharmacological treatments.[4–21]

Because study design and data-analytical decisions unavoidably impact the meaning of estimates,[22] the choice of design and analysis should follow from the research question and not the other way around.[23] Clearly defining the target of a study upfront is needed to ensure that the research question is fully aligned with clinical study objectives and operational decisions about study design and applied data-analytical approaches should follow from the defined target.[12,24,25] In the context of randomized controlled trial research, study targets are increasingly defined in terms of estimands, as is comprehensively described in the ICH-E9 (*Statistical Principles for Clinical Trials*) R1 addendum on estimands.[24] This addendum focuses on how to translate a conceptual clinical question of interest into a formal causal quantity that is the target of the analysis (i.e., the estimand – Box 1).

Although an estimand seems to naturally follow from the clinical objectives and research question of a study, defining it is not trivial. A research question can seem clear, while still leaving room for multiple interpretations. In other words, several estimands can be consistent with a conceptual research question, each with a different interpretation. The current work describes estimands that can be defined for observational studies on comparative effectiveness and safety of pharmacological treatments, taking into account that these differ between typical pharmacological treatments used in a single dose, for a sustained period, or as needed based on symptoms.

BOX 1 Definition of an estimand, estimator, and estimate according to the Glossary of the ICH-E9(R1) addendum on estimands.

Term	Definition
Estimand	"A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared" (Ref. 24, p. 19)
Estimator	"A method of analysis to compute an estimate of the estimand using clinical trial data" (Ref. 24, p. 19)
Estimate	"A numerical value computed by an estimator" (Ref. 24, p. 19)

2 | METHODS

We outline the five key elements of an estimand in studies on comparative effectiveness and safety of pharmacological treatments. Subsequently, three types of pharmacological treatments that are commonly studied in comparative effectiveness and safety research are distinguished based on treatment indication and treatment duration. For each type of treatment, we highlight decisions that need to be made when defining the estimand in a hypothetical case study. Each case study is confined to a discussion of the elements of the estimand that are most relevant or challenging to the type of treatment, but we emphasize that decisions on all five elements must be made in any study.

The five key elements of an estimand are based on the five estimand attributes in the ICH-E9(R1) addendum on estimands.[24] For each element, we describe what should be specified in an estimand for pharmacoepidemiologic comparative effectiveness and safety studies. Our refinements to the attributes of the ICH-E9 (R1) addendum on estimands are informed by the concept of 'sufficiently well-defined interventions',[26–28] considerations on estimands in observational studies of causal inference,[29–31] and the target trial emulation framework.[12,13]

2.1 | Five key elements of an estimand

To formulate the estimand of a quantitative study precisely, five elements need to be specified[12,13,24,26–29]:

1. The population of interest, describing whom the results apply to and when.
2. The contrasted treatments, describing which treatments are compared, how they are administered, and for how long. In this work, we use the term treatment regimen to indicate the duration, dosage, and route of administration of a pharmacological treatment.
3. The endpoint, describing which measure is a clinically relevant outcome and at which time point.
4. Intercurrent events, describing which events may occur after baseline and before endpoint assessment and how these events are handled preferably. Specifically, this element should outline all events that alter the course of a treatment or affect the interpretation or measurement of the endpoint, such as discontinuation of assigned treatment, use of an additional treatment and terminal events like death.
5. The population-level summary measure, describing how the defined treatment regimens are analytically compared in a statistical effect measure.

These five elements combined define the estimand and clarify which medical decision can be informed by studies that estimate it. To illustrate this, we provide in each case study below an example of a medical decision that can be informed by each estimand. Of note, medical decisions encompass many other factors than the evidence from a single scientific study and depend on the stakeholder making the decision, for example, a physician, patient, or policy maker.

2.2 | Three different types of pharmacological treatment regimens

Considerations for defining an estimand are context-specific.[31,32] We distinguish three different types of pharmacological treatments:

1. Point treatments: pharmacological treatments that are typically administered in a single dose or for a (very) short duration;
2. Sustained treatments: pharmacological treatments that are typically used to treat longer (and possibly lifelong) disease episodes;
3. As-needed treatments: pharmacological treatments that are typically prescribed as needed based on disease symptoms.

Pharmacological treatments that are typically used in a single dose or for a short duration aim to cure an acute condition or have a preventative effect. Examples are antibiotic treatments, vaccines, or bolus thrombolytics. The point treatment considered in the case study below is influenza vaccination.

Pharmacological treatments that are used for a sustained period aim to control symptoms or prevent worsening of an episodic or chronic condition. Examples are cholesterol lowering drugs, antidiabetics, or antacids. The sustained treatment considered in the case study below is medication for uncontrolled diabetes.

Pharmacological treatments that are typically prescribed as needed based on disease symptoms aim to control acute symptoms of a (chronic) condition. As such, it can be thought of as a point-treatment given over a sustained period. Examples are treatments with NSAIDs, nitroglycerin spray, or antihistamines. The as-needed treatment considered in the case study below is nitroglycerin spray for adults with stable angina pectoris.

3 | RESULTS

3.1 | Point treatment case study

We considered the research question "what is the effect of influenza vaccination on 3-month mortality risk in adults ≥60 years of age compared to not being vaccinated?". While this research question may seem clear, it is not specific and can be mapped to different estimands, each with a different interpretation and thus supporting different treatment decisions.

Four possible estimands were described for this general research question (Table 1), differing in the defined population of interest and population-level summary measure. We specified the treatment contrast as taking the influenza vaccination versus not taking it and the endpoint as 3-month mortality. Intercurrent events were not considered relevant in this example because the course of a vaccine point treatment cannot be altered and measurement of the endpoint mortality is unlikely to be precluded.

When a study is to inform a decision about implementing a population-based vaccination policy, it may aim to provide information on potential maximal mortality reduction in the population due to the influenza vaccination. The corresponding effect of interest is that of taking an influenza vaccination in the entire population of adults ≥60 years of age invited for vaccination (Table 1, Scenario 1). Together with the defined treatment contrast, endpoint, intercurrent events, and population-level summary measure, the estimand can be formulated as follows: "the average difference in 3-month mortality risk if all adults ≥60 years of age who

TABLE 1 Defining estimands in the influenza vaccination case study (point treatment).

	Elements of the estimand						
Scenario	Population (who and at what time)	Contrasted treatments (what, when, and how)	Endpoint (what, when, and how)	Intercurrent events (what and corresponding strategy)	Population-level summary measure	Estimand stated as a research question	Example of a treatment decision to be informed by the estimand
1	All individuals ≥60 years of age registered at a general practice invited for vaccination through a National Influenza Prevention Program in the period October and November	Taking an intramuscular influenza vaccination versus not taking an influenza vaccination	3-Months risk of all-cause mortality	No intercurrent events are considered relevant	Marginal risk difference	What would be the difference in average 3-month mortality risk if all adults ≥60 years of age who were invited to receive the influenza vaccination had taken it, compared to if they had not taken it?	Implementing a population-based vaccination policy, where this study provides information on potential maximal mortality reduction in the population due to the vaccine
2	Similar to Scenario 1, but of interest is the (hypothetical) group of individuals who are similar to the individuals who received the vaccination under current invitation and uptake patterns	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	What would be the difference in average 3-month mortality risk if all adults ≥60 years of age who took the influenza vaccination had instead not taken it?	Discontinuing an already implemented vaccination policy because of insufficient effectiveness
3	Similar to Scenario 1, but of interest is the (hypothetical) group of individuals who are similar to the individuals who did not receive the vaccination under current invitation and uptake patterns	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	What would be the difference in average 3-month mortality risk if all adults ≥60 years of age who did not take the influenza vaccination had instead taken it?	Stimulating uptake of an implemented vaccination policy among individuals who do not take up the invitation for vaccination
4	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	Identical to Scenario 1	Marginal risk ratio	By which percentage would the average 3-month mortality risk be reduced if all adults ≥60 years of age who were invited to receive the influenza vaccination had taken it, relative to if they had not taken it?	Evaluating vaccine effectiveness, where this study provides information by which factor mortality risk is reduced in a fully vaccinated population compared to a fully unvaccinated population

were invited to receive the influenza vaccination had taken it compared to if they had not taken it". This effect is typically referred to as the *average treatment effect*.[29,33] A study should include a representative sample of all individuals who (are planned to) receive a vaccination invitation.

Alternatively, when a study is to inform a decision on discontinuation of an already implemented vaccination policy, the research question can specifically target the population that has taken an influenza vaccination under current invitation and uptake patterns (Table 1, Scenario 2). The estimand then becomes "the average difference in 3-month mortality risk if all adults ≥60 years of age who took the influenza vaccination had instead not taken it" (typically referred to as the *average treatment effect in the treated*). In a study targeting the average treatment effect in the treated, the study sample should be representative of those individuals who took the vaccination, which can be achieved by, for example, reweighting or taking a matched set of unvaccinated controls, rather than a random sample of the full target population.[33]

Conversely, it can be of interest to know the potential gain if uptake of the offered vaccination were improved. In this case a study should target the effect of vaccination for individuals similar to those who did not take the influenza vaccination under current invitation and uptake patterns (the so-called the *average treatment effect in the untreated*, Table 1, Scenario 3): "the average change in 3-month mortality risk after vaccination if all adults ≥60 years of age who did not take the influenza vaccination had instead taken it". The study sample should be representative of unvaccinated individuals, which can again be achieved by, for example, reweighting or taking a matched set of vaccinated individuals.[33]

When the aim is to investigate the strength of the effect of vaccination, a study should provide information on how many times less likely a fully vaccinated population is to die compared to a fully unvaccinated population. The population-level summary measure can be defined, for instance, as a marginal risk ratio (Table 1, Scenario 4). The marginal risk ratio is a measure of *relative* risk and provides information about the strength of the treatment effect relative to a base endpoint risk. *Additive* risk measures express the difference in endpoint risk attributable to the treatment on an absolute scale and are sometimes said to be more suitable to inform clinical decision-making.[34] Because results on the relative scale and absolute scale can differ, it is important to pre-specify that either both scales will be investigated (and thus reported on) or which of the two population-level summary measures is of primary interest in a study.[35]

For point treatments, the moment at which persons are included in the population is generally defined by an event, such as an invitation for vaccination or the diagnosis of an acute condition. Although this provides a natural time point for cohort entry, the time origin of the contrasted treatments could also be set at the time the treatment is given (as was the case in this example). In the analysis, it should be considered how to handle the period after event-based cohort entry and moment of treatment initiation to avoid introduction of immortal time.[9,36,37]

3.2 | Sustained treatment case study

For the research question "what is the effect of using a DPP-4 inhibitor compared to using a GLP_1 agonist on the 1-year risk of severe hypoglycemia in adults with uncontrolled diabetes?", three possible estimands were described (Table 2). The estimands differed regarding the defined contrasted treatments and handling of intercurrent events. For the other elements of the estimand, we considered the population to be adults with uncontrolled diabetes mellitus type 2, the endpoint to be 1-year risk of severe hypoglycemia, and the population-level summary measure to be the marginal risk difference. The considered intercurrent events were discontinuation of treatment, switch to alternative treatment, switch to intermediate-acting insulin (treatment escalation), and death.

To inform the decision which treatment to start, an appropriate treatment contrast of interest would be the effect of *initiating* a DPP-4 inhibitor versus *initiating* a GLP_1 agonist (Table 2, Scenario 1), which is sometimes referred to as the observational analog of an *intention-to-treat* effect.[13,24] The estimand can then be described as "the difference in average 1-year risk of severe hypoglycemia if all adults with uncontrolled diabetes had initiated a DPP-4 inhibitor, compared to if they had initiated a GLP_1 agonist". The results are applicable to a population with similar treatment compliance to the study sample, meaning that descriptive results on treatment compliance are essential for interpretation of the effect. The analysis does not need to consider whether individuals sustain the treatment after baseline and exchangeability of treatment groups is assessed at baseline only.

Alternatively, for an individual with uncontrolled diabetes who intends to sustain the therapy they are starting with, the choice of treatment can better be informed by a comparative treatment effect under perfect compliance, that is, the effect of the protocolled treatment regimen. The contrasted treatments are defined as initiating a DPP-4 inhibitor or GLP_1 agonist, and – once started – complying to the protocolled treatment regimen over the course of a year (Table 2, Scenario 2), often referred to as the *per-protocol* effect.[13,24] The interpretation is: "the difference in average 1-year risk of severe hypoglycemia if all adults with uncontrolled diabetes had initiated and sustained a DPP-4 inhibitor, compared to if they had initiated and sustained a GLP_1 agonist". The effect is defined for a scenario in which the intercurrent events discontinuation of treatment, and switch to the alternative treatment arm would not occur, that is, these intercurrent events are handled under the ICH-E9 *hypothetical* strategy.[24] Estimation of this estimand from observational data can be challenging because the defined treatment regimen may not be observed for some individuals due to intercurrent events and adjustment for time-varying confounding might be necessary.[6,16,38]

The intercurrent event "switching to intermediate-acting insulin" can be considered a treatment failure. Hence, for individuals who consider switch to intermediate acting insulin an undesirable outcome, treatment choice can be informed by an estimand in which this intercurrent event is part of a composite endpoint indicating an unfavorable outcome, that is, hypo- or hyperglycemia (Table 2, Scenario 3).

TABLE 2 Defining estimands in the diabetes case study (sustained treatment).

	Elements of the estimand						
Scenario	Population (who and at what time)	Contrasted treatments (what, when, and how)	Endpoint (what, when, and how)	Intercurrent events (what and corresponding strategy)	Population-level summary measure	Estimand stated as a research question	Example of a treatment decision to be informed by the estimand
1	Adults with uncontrolled diabetes mellitus type 2[a] presenting at medical check-up, with similar treatment compliance to the study sample	Prescription of DPP-4 inhibitor (oral Linagliptin 5 mg daily) versus prescription of GLP_1 agonist (oral Rybelsus 7 mg daily)[b]	1-Year risk of severe hypoglycemia[c] while alive	Considered part of the treatment regimen of interest, that is, handled under the ICH-E9 *treatment policy* strategy: • Discontinuation of treatment • Switch to alternative treatment arm • Switch to intermediate-acting insulin. Death is handled under the ICH-E9 *while on treatment* strategy	Marginal risk difference	What would be the difference in average 1-year risk of severe hypoglycemia if all adults with uncontrolled diabetes had initiated a DPP-4 inhibitor, compared to if they had initiated a GLP_1 agonist?	Advising on treatment initiation in the population of adults with uncontrolled diabetes mellitus type 2 presenting at medical check-up, in a population with similar treatment compliance and add-on treatments to the study sample
2	Adults with uncontrolled diabetes mellitus type 2[a] presenting at medical check-up, for individuals who would perfectly adhere to treatment and not switch treatment	Initiation and 1-year sustained use of DPP-4 inhibitor (oral Linagliptin 5 mg daily) versus initiation and 1-year sustained use of GLP_1 agonist (oral Rybelsus 7 mg daily)[b]	Identical to Scenario 1	Similar to Scenario 1, but because interest is now in a scenario where, possibly counter to fact, intercurrent events would not occur, the following intercurrent events are handled under the ICH-E9 *hypothetical* strategy: • Discontinuation of treatment • Switch to alternative treatment arm	Identical to Scenario 1	What would be the difference in average 1-year risk of severe hypoglycemia if all adults with uncontrolled diabetes had initiated and compliantly used a DPP-4 inhibitor, compared to if they had initiated and compliantly used a GLP_1 agonist?	Making a medical decision about sustained treatment with DPP-4 inhibitor and GLP_1 agonist for the population of adults with uncontrolled diabetes mellitus type 2 presenting at medical check-up
3	Identical to Scenario 2	Identical to Scenario 2	1-Year composite risk of severe hypoglycemia[c] or switch to intermediate-acting insulin injections, while alive	Similar to Scenario 2, but switch to intermediate-acting insulin injections is considered treatment failure, that is, the ICH-E9 *composite* strategy	Identical to Scenario 1	What would be the difference in average 1-year risk of severe hypoglycemia or switch to intermediate-acting insulin injections if all adults with uncontrolled diabetes had initiated and compliantly used a DPP-4 inhibitor, compared to if they had initiated and compliantly used a GLP_1 agonist?	Making a medical decision about sustained treatment with DPP-4 inhibitor and GLP_1 agonist for the population of adults with uncontrolled diabetes mellitus type 2 presenting at medical check-up who consider switch to intermediate acting insulin an undesirable outcome

[a]Adults who use metformin and gliclazide at recommended dosages and whose fasting blood glucose levels are >8 mmol/L and/or HbA_{1c} levels are 53 to 68 mmol/mol and who are not in need of renal function replacement therapy (eGFR >10 mL/min/1,73 m^2).
[b]Oral Rybelsus initial dose 3 mg daily, after 1 month dose 7 mg daily, possibly adjusted to 14 mg daily.
[c]Defined as a composite of (at least) one of the following: a fasting glucose level of <3 mmol/L, requiring assistance of another person to actively administer carbohydrates, glucagon, or take other corrective actions.

TABLE 3 Defining estimands in the nitroglycerin spray case study (as-needed treatment).

	Elements of the estimand						
Scenario	Population (who and at what time)	Contrasted treatments (what, when, and how)	Endpoint (what, when, and how)	Intercurrent events (what and corresponding strategy)	Population-level summary measure	Estimand stated as a research question	Example of a treatment decision to be informed by the estimand
1	Adults with newly diagnosed stable angina pectoris	Prescribing sublingual nitroglycerin spray 0.4 mg/dose to dock an angina attack versus not prescribing nitroglycerin	6-Months risk of syncope while alive	Considered part of the treatment regimen of interest, that is, the ICH-E9 *treatment policy* strategy: • Having a myocardial infarction • Not using nitroglycerin during angina attack • Overuse of nitroglycerin • Switch in maintenance therapy. Death is handled under the ICH-E9 *while on treatment* strategy	Marginal risk difference	What would be the difference in average 6-months risk of syncope if all adults with stable angina pectoris were prescribed nitroglycerin as-needed, compared to if they had not been prescribed nitroglycerin?	This estimand cannot directly inform a decision but can be a first step in safety research, whether nitroglycerin spray is safe to be prescribed for adults with newly diagnosed stable angina pectoris regarding occurrence of syncope, in a population with similar nitroglycerin use as the study sample
2	Identical to Scenario 1	Taking sublingual nitroglycerin spray 0.4 mg/dose to dock an angina attack versus not taking nitroglycerin	Identical to Scenario 1	Similar to Scenario 1, but not using nitroglycerin during angina attack or overuse of nitroglycerin are handled under the ICH-E9 *hypothetical* strategy	Identical to Scenario 1	What would be the difference in average 6-months risk of syncope if all adults with stable angina pectoris used nitroglycerin as prescribed, compared to if they had not used nitroglycerin?	Determining whether use of nitroglycerin spray as prescribed is safe for adults with newly diagnosed stable angina pectoris regarding occurrence of syncope

The composite endpoint allows to jointly investigate the risk of severe hypoglycemia and the risk of diabetes not being controlled. The interpretation of the estimand changes to "the difference in average 1-year risk of severe hypoglycemia or switch to intermediate-acting insulin if all adults with uncontrolled diabetes had initiated and sustained a DPP-4 inhibitor, compared to if they had initiated and sustained a GLP_1 agonist".

In all three scenarios, the intercurrent event death is handled under the ICH-E9 *while on treatment* strategy, meaning that the endpoint is interpreted as 1-year risk of severe hypoglycemia while alive. The interest lies in evaluating adverse events by treatment before death.[39–41] A competing risk approach can be implemented where death is handled as a competing event and cumulative incidence of severe hypoglycemia is calculated.

The contrasted treatments are active treatments in these scenarios. However, in other clinical settings it might be of interest to investigate whether a sustained treatment should be preferred over no treatment. The treatment contrast is then initiating and sustaining treatment versus non-use. This mainly poses challenges for the analysis in deciding which non-users could be considered a suitable comparator group[4] and in defining the moment from which onwards non-users should be compared to treatment initiators.[12,42]

3.3 | As-needed treatments

For the research question "what is the effect of nitroglycerin spray as-needed in adults with stable angina pectoris on risk of syncope?", two possible estimands were described (Table 3). The estimands differed regarding the defined contrasted treatments and handling of intercurrent events. For the other elements of the estimand, we defined the population as adults with newly diagnosed stable angina pectoris, the endpoint as 6-month risk of cardiac arrest, and the population-level summary measure as the marginal risk difference. The considered intercurrent events were having a myocardial infarction, not using nitroglycerin during an angina attack, switch in maintenance therapy for angina pectoris, and death. Note that syncope is an adverse event potentially related to nitroglycerin use, so this research question would be part of a safety evaluation.

The contrasted treatments of interest can be defined as *prescribing* nitroglycerin as-needed versus not *prescribing* nitroglycerin (Table 3, Scenario 1).[13,24] Some individuals who are prescribed nitroglycerin may not take it during an angina attack or may choose the dose and frequency at their own discretion. The results are thus applicable to a population with similar nitroglycerin use to the study sample, meaning that descriptive results on actual nitroglycerin use are essential for interpretation of the effect.

To evaluate the safety of nitroglycerin as-needed, instead of contrasting treatment prescriptions, it may be more important to know the 6-month risk of syncope for individuals who actually used the treatment.[13,24] When interest is in studying the effect of use that is compliant with the intended use of the treatment ('treatment protocol'), one could target a *per-protocol* effect; "the difference in average 6-months risk of syncope if all adults with stable angina pectoris used nitroglycerin as prescribed, compared to if they had not used nitroglycerin" (Table 3, Scenario 2). In alternative estimands, other variants of treatment use could be chosen for instance based on dosage and frequency of use.

Studies on effects of as-needed treatments face the challenge of obtaining observational data suitable to estimate the estimand of interest. Routinely collected data usually contain information on treatment prescription only, and information about actual treatment use during an angina attack is difficult to retrieve.

4 | DISCUSSION

A research question may seem to unambiguously define the target of a pharmacoepidemiologic analysis but can still be open to multiple interpretations. Defining an estimand helps resolving such unclarities by refining the research question. We outlined how the ICH-E9 (R1) addendum on trial estimands can be applied to observational pharmacoepidemiologic comparative effectiveness and safety studies to formulate an estimand. The case studies illustrated that several estimands can be defined for the same research question. By articulating the estimand, the research question becomes more specifically aligned to the treatment decision that is to be informed by the study.

Defining the five elements of an estimand before conducting a study informs choices regarding the study design and data-analytical approaches. A study protocol thus ideally contains a clear definition of the targeted estimand(s).[11] Additionally, reporting the five outlined elements of an estimand is essential for readers' interpretation of study findings and assessment of whether study design and data-analytical approaches are suitable for the clinical problem at hand. Ideally, the methods section of an article starts with a brief statement of the estimand(s). The operationalization of each element can be elaborated on in its respective subsection of the methods section.

Our examples connect statistical principles[13,29,30] and research objectives in a non-mathematical way and with added refinement compared to the PICOT acronym.[43,44] Researchers can use the example case studies in tandem with existing guidance for designing pharmacoepidemiologic studies. For instance, we recommend defining an estimand as a first step when filling in the structured template for planning and reporting on real-world evidence study implementation (STaRT-RWE).[11] Furthermore, an estimand in words is an insightful starting point when using the CERBOT[32]: a tool informing specification of the target trial in comparative effectiveness research.

Often, multiple estimands need to be specified in a single study. For instance, when a study assesses multiple endpoints, such as a primary effectiveness endpoint, secondary effectiveness endpoints and safety endpoints, all five elements need to be clearly outlined for each study endpoint.

A limitation to our study is that we provided only a selection of possible estimands that can be defined for a research question. Particularly, our discussion of contrasted treatments was limited to static treatment regimens and did not address dynamic treatment regimens,

in which the decision about treatment use at a particular time point is based on updated time-varying covariate information.[31,32] The contrasted treatments did not contain an example of a change in treatment and were restricted to initiation of treatment. Other intercurrent events can be defined when investigating sustained treatments, for instance regarding non-adherence (e.g., skipping dosages), use of rescue medication (also called escape medication), change in maintenance treatment, or remission of symptoms in sustained treatment of episodic diseases. Our discussion of population-level summary measures did not address specification of subgroup effects. Regardless, the principles outlined provide a general basis for specifying an estimand in these scenarios and can be applied to define other estimands, which incorporate expert domain knowledge and consider relevant stake holders for the study at hand.

Another limitation is that we did not explicate how an estimand can inform the choice of study design and analytical approach. We refer to existing papers on this topic,[29,33,41,45–47] particularly to explanations about aligning estimands to estimators in randomized trials.[38,48] An increasingly popular methodological approach is the target trial emulation framework.[13,49–51] Target trial emulation provides a structure to estimate a treatment effect of interest using observational data. The trial emulation steps should ideally be preceded by choosing the target of interest, that is, specifying an estimand. In some cases, estimands that are identified as interesting prior to observing the data cannot be estimated with the available data. For example, conditional exchangeability cannot be assumed in many (secondary) databases due to the limited availability of (time-varying) confounders. Reporting the relevant estimands is useful to inform data collection in follow-up research.

Even if a clinical research question seems clearly formulated, there are multiple quantitative estimands into which it can be translated. Thus, without specifying the elements of an estimand for a particular study, the findings of that study are prone to misinterpretation. Choosing the elements of an estimand clarifies the exact study question and how study findings can be interpreted.

AUTHOR CONTRIBUTIONS

Kim Luijken, Rik van Eekelen, Helga Gardarsdottir, Rolf H. H. Groenwold, Nan van Geloven – wrote manuscript, designed research, performed research.

ACKNOWLEDGMENTS

We acknowledge Victoria Beckham, Melanie Brown, Emma Bunton, Melanie Chisholm, and Geri Halliwell for inspiring the title of this work. Rolf H. H. Groenwold was supported by grants from the Leiden University Medical Center.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ETHICS STATEMENT

The authors state that no ethical approval was needed.

REFERENCES

1. Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther*. 2017;102(6):924-933.
2. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther*. 2019;105(4):867-877.
3. Schneeweiss S, Eichler HG, Garcia-Altes A, et al. Real world data in adaptive biomedical innovation: a framework for generating evidence fit for decision-making. *Clin Pharmacol Ther*. 2016;100(6):633-646.
4. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9):915-920.
5. Brookhart MA. Counterpoint: the treatment decision design. *Am J Epidemiol*. 2015;182(10):840-845.
6. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf*. 2017;26(4):459-468.
7. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med*. 2019;170(6):398-406.
8. Wang SV, Schneeweiss S. A framework for visualizing study designs and data observability in electronic health record data. *Clin Epidemiol*. 2022;14:601-608.
9. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol*. 2008;167(4):492-499.
10. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1. 0. *Value Health*. 2017;20(8):1009-1022.
11. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021;372:m4856.
12. Hernán MA, Sauer BC, Hernández-Díaz 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;79:70-75.
13. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758-764.
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
15. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512-522.
16. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
17. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci: Rev J Inst Math Stat*. 2010;25(1):1-21.
18. Abadie A, Imbens GW. Large sample properties of matching estimators for average treatment effects. *Econometrica*. 2006;74(1):235-267.
19. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-1156.
20. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion). *J Am Stat Assoc*. 1996;91(434):444-455.
21. Pottegård A, Morin L, Hallas J, et al. Where to begin? Thirty must-read papers for newcomers to pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2022;31(2):257-259.
22. Wang SV, Sreedhara SK, Bessette LG, Schneeweiss S. Understanding variation in the results of real-world evidence studies that address the same question. *J Clin Epidemiol*. 2022;151:161-170.
23. Kahan BC, Cro S, Li F, Harhay MO. Eliminating ambiguous treatment effects using estimands. *Am J Epidemiol*. 2023;kwad036.

24. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf (Accessed April 29 2022)
25. Phillips A, Clark T. Estimands in practice: bridging the gap between study objectives and statistical analysis. *Pharm Stat*. 2021;20(1):68-76.
26. VanderWeele TJ. On well-defined hypothetical interventions in the potential outcomes framework. *Epidemiology*. 2018;29(4):e24-e25.
27. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes*. 2008;32(3):S8-S14.
28. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20(6):880-883.
29. Goetghebeur E, le Cessie S, De Stavola B, et al. Formulating causal questions and principled statistical answers. *Stat Med*. 2020;39(30):4922-4948.
30. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology*. 2014;25(3):418-426.
31. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
32. Medical Technology and Practice Patterns Institute. Comparative Effectiveness Research Based on Observational Data to Emulate a Target Trial 2022. Available from: http://cerbot.org/ (Accessed August 24 2022)
33. Greifer N, Stuart EA. Choosing the estimand when matching or weighting in observational studies. *arXiv* Preprint arXiv:210610577. 2021.
34. Rothman KJ, Greenland S, Lash TL. Modern epidemiology: Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia 2008.
35. Vandenbroucke JP, Ev E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med*. 2007;147(8):W-163-W-194.
36. Platt R, Hutcheon J, Suissa S. Immortal time bias in epidemiology. *Curr Epidemiol Reports*. 2019;6(1):23-27.
37. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care*. 2012;35(12):2665-2673.
38. Mallinckrodt C, Bell J, Liu G, et al. Aligning estimators with estimands in clinical trials: putting the ICH E9 (R1) guidelines into practice. *Ther Innov Regul Sci*. 2020;54(2):353-364.
39. Siegel J, Grinsted L, Liu F, et al. Censoring and censoring mechanisms in oncology in light of the estimands framework. *arXiv* Preprint arXiv:220301781. 2022.
40. Siegel JM, Weber H-J, Englert S. The role of occlusion: potential extension of the ICH E9 (R1) addendum on Estimands and sensitivity analysis for time-to-event oncology studies. *arXiv* Preprint arXiv:220302182. 2022.
41. Rufibach K. Treatment effect quantification for time-to-event endpoints–Estimands, analysis strategies, and beyond. *Pharm Stat*. 2019;18(2):145-165.
42. Danaei G, Rodríguez LAG, Cantero OF, et al. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol*. 2018;96:12-22.
43. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12-A13.
44. Haynes RB. Clinical epidemiology: how to do clinical practice research: Lippincott Williams & Wilkins 2012.
45. Penning de Vries BB, Groenwold RH. Bias of time-varying exposure effects due to time-varying covariate measurement strategies. *Pharmacoepidemiol Drug Saf*. 2022;31(1):22-27.
46. Li X, Young JG, Toh S. Estimating effects of dynamic treatment strategies in pharmacoepidemiologic studies with time-varying confounding: a primer. *Curr Epidemiol Reports*. 2017;4(4):288-297.
47. Choi J, Dekkers OM, le Cessie S. Tying research question and analytical strategy when variables are affected by medication use. *Pharmacoepidemiol Drug Saf*. 2023.
48. Mitroiu M, Teerenstra S, Oude Rengerink K, et al. Estimation of treatment effects in short-term depression studies. An evaluation based on the ICH E9 (R1) estimands framework. *Pharm Stat*. 2022;21(5):1037-1057.
49. Franklin JM, Glynn RJ, Suissa S, Schneeweiss S. Emulation differences versus biases when calibrating RWE findings against RCTs. *Clin Pharmacol Ther*. 2020;107(4):735-737.
50. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther*. 2020;107(4):817-826.
51. Usman MS, Pitt B, Butler J. Target trial emulations: bridging the gap between clinical trial and real-world data: Wiley online. *Library*. 2021;23:1708-1711.

How to cite this article: Luijken K, van Eekelen R, Gardarsdottir H, Groenwold RHH, van Geloven N. Tell me what you want, what you really really want: Estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies. *Pharmacoepidemiol Drug Saf*. 2023;32(8):863-872. doi:10.1002/pds.5620