Analytical Report

DIA

Therapeutic Innovation & Regulatory Science
1-15

DOI: 10.1177/2168479019841316
tirs.sagepub.com

Defining Efficacy Estimands in Clinical Trials: Examples Illustrating ICH E9(R1) Guidelines

Bohdana Ratitch, PhD[1], Niti Goel, MD, FACR[2,3], Craig Mallinckrodt, PhD[4], James Bell, PhD[5], Jonathan W. Bartlett, PhD[6], Geert Molenberghs[7], Pritibha Singh, MSc, MBA[8], Ilya Lipkovich, PhD[9], and Michael O'Kelly, PhD[10]

Abstract

This paper provides examples of defining estimands in real-world scenarios following ICH E9(R1) guidelines. Detailed discussions on choosing the estimands and estimators can be found in our companion papers. Three scenarios of increasing complexity are illustrated. The first example is a proof-of-concept trial in major depressive disorder where the estimand is chosen to support the sponsor decision on whether to continue development. The second and third examples are confirmatory trials in severe asthma and rheumatoid arthritis respectively. We discuss the intercurrent events expected during each trial and how they can be handled so as to be consistent with the study objectives. The estimands discussed in these examples are not the only acceptable choices for their respective scenarios. The intent is to illustrate the key concepts rather than focus on specific choices. Emphasis is placed on following a study development process where estimands link the study objectives with data collection and analysis in a coherent manner, thereby avoiding disconnect between objectives, estimands, and analyses.

Keywords

estimands, clinical trials, intercurrent events

Introduction

This paper provides detailed examples of estimand definitions in several clinical trial settings to facilitate practical implementation of guidelines outlined in the ICH E9(R1) draft addendum on "Estimands and Sensitivity Analysis in Clinical Trials."[1] This paper follows the framework introduced in ICH E9(R1) that is further discussed in our companion papers on estimands[2] and estimators.[3]

Three progressively more complex examples are illustrated in this paper. The first is a proof-of-concept (PoC) trial in major depressive disorder (MDD), where the key decision maker is the sponsor who must decide whether the drug has sufficient potential benefit to continue in development. The second example is a confirmatory trial of an add-on maintenance treatment for patients with severe asthma. The third example is a confirmatory trial in rheumatoid arthritis (RA). The last two are complex examples where estimands need to address the interests of multiple stakeholders, for example, sponsors, regulators, patients, prescribing physicians, and payers, although our focus is on estimands for regulatory decision making.

In most studies (including our examples), a variety of post-randomization events that mark a change in the course of treatment, for example, initiation of rescue therapy or premature discontinuation of the randomized treatment, can be anticipated. Such events may influence the estimation and interpretation of treatment effects. These events are referred to as intercurrent events (ICEs) in ICH E9(R1), and the guideline stipulates that handling of these events needs to be described as part of the estimand definition. The examples in this paper describe the strategies that can be chosen for the ICEs in the corresponding clinical contexts. For each example, we provide full specifications of estimands following a template of the estimand definition suggested in ICH E9(R1).

The estimands described in these examples are not the only acceptable choices for the respective clinical settings. Other

[1] Eli Lilly, Montreal, Quebec, Canada
[2] Kezar Life Sciences, South San Francisco, CA, USA
[3] Department of Medicine, Duke University School of Medicine, Durham, NC, USA
[4] Biogen, Cambridge, MA, USA
[5] Elderbrook Solutions GmbH, High Wycombe, United Kingdom
[6] Department of Mathematical Sciences, University of Bath, Bath, United Kingdom
[7] I-BioStat, Universiteit Hasselt and KU Leuven, Belgium
[8] Novartis, Basel, Switzerland
[9] Eli Lilly, Indianapolis, IN, USA
[10] IQVIA, Dublin, Ireland

Submitted 22-Oct-2018; accepted 26-Feb-2019

Corresponding Author:
Michael O'Kelly, PhD, IQVIA, Estuary House, East Point Business Park, Dublin 3, Ireland.
Email: Michael.OKelly@iqvia.com

estimands may be of interest to some decision makers. This paper intends to illustrate the process and key concepts rather than focus on specific choices. More details about the assumptions behind and implications of different strategies for dealing with ICEs can be found in the companion paper on estimands.[2] For each example, we also briefly mention which analysis methods, that is, estimators, can be used so that they are well aligned with the defined estimands. It should be noted that for a given estimand several suitable estimators may exist (eg, for estimating a difference between population means); justification of the specific choices of analysis methods is beyond the scope of this manuscript, and more details about estimation methods can be found in the second companion paper.[3]

The next section provides a brief overview of the estimand definition elements and strategies for the ICEs outlined in ICH E9(R1),[1] as well a trial design process[2] wherein the estimand defines the scientific question of interest and links the study objectives with data collection and analysis in a coherent manner. The third section presents the first example, MDD, and provides a high-level view of the key considerations in a PoC study that drive the choice of the estimand in this context. The fourth and fifth sections are devoted to the examples in severe asthma and rheumatoid arthritis, respectively. Each of these examples begins with a description of the clinical context so that ICEs, likely to occur in these contexts, can be identified. We then briefly summarize one historical confirmatory clinical trial for the corresponding indication. The estimands were not explicitly defined in publicly available sources for these trials, so we reconstruct the implied primary estimands based on the description of the trial designs and primary analyses. This is *not* the process that should be followed at the time of a prospective trial design; we use it mainly to fix ideas and provide some background about the therapeutic areas and as an illustration of how some approaches typically utilized in these therapeutic areas can be expressed in a formal estimand definition. We provide a discussion of how the primary estimands used in these historical trials focus on the needs of the intended decision makers. Subsequently, we suggest some alternative estimands that may also be appropriate in the same clinical context but would stem from a different objective.

Overview of estimand definition elements and trial design process

This section briefly summarizes the key elements required for an estimand definition. The reader is referred to ICH E9(R1)[1] and the companion paper[2] for an in-depth introduction.

An estimand describes the quantity to be estimated to address a specific study objective. ICH E9(R1) outlined 4 elements that together comprise the estimand definition:

A. the population, for example, the patients targeted by the scientific question;
B. the variable (or endpoint) to be measured for each patient to address the scientific question;
C. how to account for ICEs to reflect the scientific question of interest; and
D. the population-level summary for the variable that provides a basis for a comparison between treatment conditions.

Elements A, B, and D have typically been specified in study protocols, albeit not as part of a unified estimand definition. Element C is a new requirement. Intercurrent events are those events that occur after randomization and alter the course of the randomized treatment during the intended study treatment period. Examples of such events include premature discontinuation of the randomized treatment, initiation of rescue therapy, or switch to an alternative therapy. Some ICEs may also represent a change in the subject state, for example, death.

Postrandomization events can undermine randomization and compromise the evaluation of causal effects of the randomized treatment. Postrandomization treatment changes may confound the effect of the originally randomized treatment. Although the causal link between the *assignment* to a randomized treatment and the outcome may still exist, the outcome will reflect the pharmacologic effect of the entire realized treatment history, which may be different from that of the randomized treatment.

ICH E9(R1) suggested 5 strategies that can be used to handle ICEs:

- Treatment policy
- Composite
- While-on-treatment
- Hypothetical
- Principal stratification

The strategy choice is driven by the treatment regimen that is targeted for evaluation and depends on the clinical context. When the treatment policy strategy is used for an ICE that marks the start of a new treatment, the new treatment becomes part of the evaluated treatment regimen in addition to the randomized treatment. Note that a "new treatment" may represent a period of no treatment, for example, when the originally randomized treatment is discontinued and no alternative therapy is administered. The other 4 strategies are used when ICEs mark the start of new treatment elements that should not be included in the evaluated treatment regimen as they introduce confounding in the estimate of treatment effect, which would make it difficult or impossible to derive useful conclusions about causal effects of the experimental treatment in view of a specific study objective.

The estimand should be defined early in the trial design process: after identifying the decision maker(s) and their objectives, and before estimating the required sample size, planning assessment schedule, and choosing analysis methods. Once the estimand is defined, the estimators (statistical analysis methods) should be chosen to align with the estimand. The strategies specified in the estimand to handle ICEs determine which data

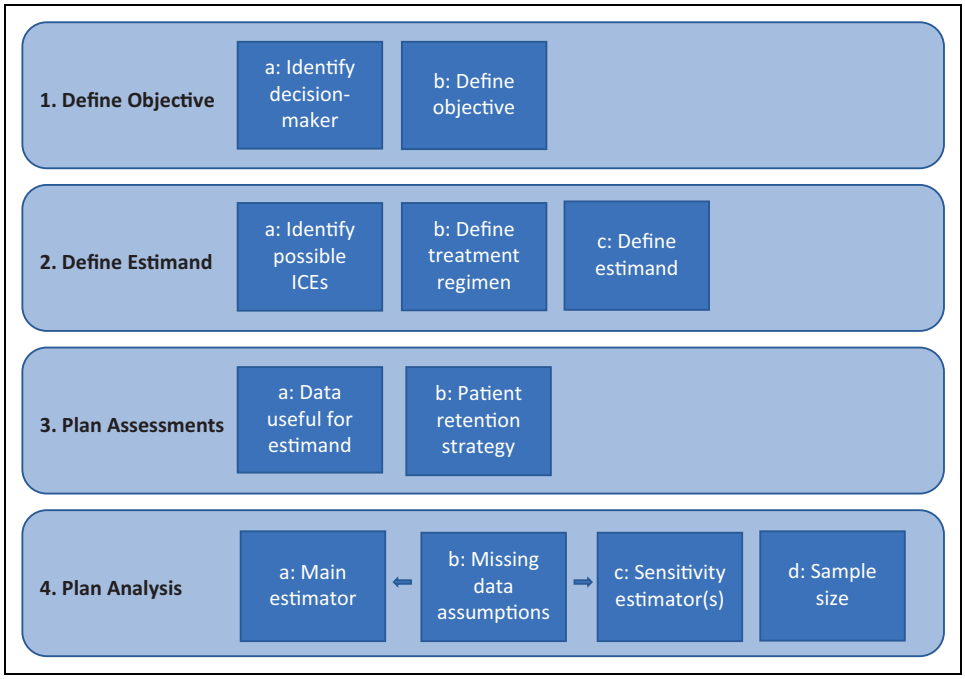

Figure 1. Study development process chart.

are useful for the estimand and, therefore, influence when and how data should be collected.

For convenience, the study development process chart from our companion paper[2] is reproduced in Figure 1. The examples presented herein focus on step 2 of this process, defining an estimand, with brief comments on steps 3 and 4.

Proof-of-Concept Trial in Major Depressive Disorder

Background

Major depressive disorder (MDD) is a common psychiatric condition with a lifetime incidence of approximately 15%.[4] The disorder ranges from mild to severe and is associated with significant potential morbidity and mortality, contributing to suicide and adverse impact on concomitant medical illnesses, interpersonal relationships, and work. The treatment objectives are to reduce or resolve signs and symptoms of the disease, restore psychosocial and occupational function, and reduce the likelihood of relapse or recurrence.[5] Guidelines support pharmacologic therapy for the treatment of depression in addition to psychotherapy. Antidepressant medications include selective serotonin reuptake inhibitors (SSRIs), serotonin/norepinephrine reuptake inhibitors (SNRIs), atypical antidepressants, serotonin-dopamine activity modulators (SDAMs), tricyclic antidepressants (TCAs), and monoamine oxidase inhibitors (MAOIs).[6] Antidepressants in established classes (eg, SSRIs, SNRIs) typically demonstrate initial benefits after 3 to 4 weeks. The current standard design for short-term efficacy trials in MDD are randomized, double-blind, placebo-controlled, parallel designs of 6 to 8 weeks' duration.[7] Expectations are that patients of all severities will be evaluated, but that evaluation of patients with treatment-resistant depression (usually defined as having failed 2 or more pharmacologic therapies) will be performed separately.

There are several aspects to consider for PoC trials in MDD: high rates of placebo response and premature discontinuation of the randomized treatment, which limit the ability to distinguish an effective drug from placebo.[8,9]

Another consideration for PoC trials (not specific to MDD) is that at this early stage of development, the optimum dose, dosing regimen, and/or formulation may not be known. Suboptimal dosing in the PoC trial could reduce treatment effects. However, knowledge gained from the PoC study could result in improved dosing and improved outcomes in subsequent trials.

Trial Description

The example trial was a randomized, double-blind, placebo-controlled, parallel-group phase 2, PoC trial in MDD with an 8-week treatment period in an adult outpatient population.[10] The primary efficacy outcome was the GRID-Hamilton Depression Rating Scale (GRID-HAMD) 17-item total score[11] at the end of the 8-week double-blind treatment period. Efficacy assessments were planned at baseline, and each post-baseline visit at weeks 1-8.

At this early stage of development in this indication, treatment regimens involving other drugs are not relevant. Moreover, with many drugs already on the market for MDD, new drugs are likely to be used in difficult-to-treat patients who have not responded to or been intolerant of other drugs, making

Table 1. Anticipated Treatment Regimens[a] in a PoC Trial of MDD

Scenario	Treatment Regimen Over 8 wk
1	Z
2	Z→O
3	Z→P(i)

[a]Treatment regimen: Z = randomized treatment; O = no treatment; P = post discontinuation of randomized treatment; (i) = standard of care treatment for MDD, not pre-specified by study protocol.

assessments including rescue therapies less relevant. Therefore, no rescue therapy was to be made available concomitantly with the randomized treatment. If patients prematurely discontinued the randomized treatment (which would be considered an ICE), they were offered a standard-of-care therapy, which may include no pharmacologic treatment depending on patient's symptoms. The objective was to evaluate the experimental treatment with respect to its pharmacologic efficacy and to assess the tolerability and adherence separately.

Primary Estimand

We now follow the study development process presented in the second section for the primary objective of evaluating efficacy.

1a. Identify decision maker: The key decision maker is the sponsor.

1b. Define objective: The main objective is to determine whether to continue development of this investigational product (IP) into phase 3 by evaluating the superiority of the pharmacologic effect of the experimental drug to placebo in treating the symptoms of MDD.

2a. Identify possible ICEs: Actual treatment regimens that may occur in this trial in either a planned manner (as per the study treatment discontinuation guidelines mentioned above) or unplanned are summarized in Table 1. Scenario 1 represents ideal adherence to the randomized treatment through 8 weeks of the double-blind period, without any ICEs. (Note: by ideal adherence, we do not mean 100% compliance, but rather continuing with the randomized treatment for the planned duration of 8 weeks.) Treatment changes, that is, switch to no treatment or to standard of care, represented by scenarios 2 and 3, respectively, are ICEs that may occur at any time during the 8-week double-blind period. Concomitant use of other antidepressant medications is prohibited. Deaths are not expected during this short-term study in the enrolled subject population.

2b. Define treatment regimen under evaluation: The treatment regimen intended for evaluation to address the objective stated in 1b is the randomized treatment administered as directed for the planned duration of 8 weeks. The motivation for this choice is to assess the full efficacy potential, that is, the pharmacologic effect, of the experimental treatment if all subjects adhere to it.

2c. Define estimand: The estimand is defined as follows, specifying the 4 elements as outlined in the second section per ICH E9(R1):

A. The treatment effect is to be estimated for the population of adult patients with MDD as defined by the protocol inclusion/exclusion criteria (in the sense of delineating the population).
B. Efficacy is to be measured using the primary endpoint of the change from baseline to week 8 of the double-blind study period in GRID-HAMD 17-item total score.
C. All ICEs leading to changes in treatment, such as premature discontinuation of the randomized treatment with or without a switch to alternative therapies, will be handled by a hypothetical strategy to estimate what the outcome would have been at the designated time point if all subjects adhered to their randomized treatment through that time point.
D. The mean of changes from baseline to week 8 of the double-blind study period in GRID-HAMD 17-item total score will be estimated for each treatment group, and the experimental treatment will be compared to placebo using difference in group means.

3a. Data useful for estimand: The GRID-HAMD 17-item data necessary for this estimand are those observed while adhering to the initial randomized treatment. Observations after discontinuation of the randomized treatment, regardless of initiation of subsequent therapies, are not useful for evaluation of the pharmacologic effect of an 8-week study treatment regimen. Therefore, the data after such ICEs do not need to be collected for the purposes of this estimand. For subjects with ICEs, week 8 data will not be available and the corresponding outcomes will need to be estimated via statistical modeling in a manner that is consistent with the hypothetical scenario stated in element C of the estimand definition (see more details in step 4a below).

Although data post ICEs are not required for the primary estimand, it may be useful to collect these data for estimation of supportive estimands that can inform subsequent trials.

3b. Patient retention strategy: Retention strategies can focus on trial conduct features to minimize missing data. These features go beyond our current scope and have been discussed elsewhere.[12]

4a. Main estimator: An estimator aligned with the estimand is a likelihood-based repeated measures approach, such as mixed model for repeated measures (MMRM).[3,13] The MMRM model should be applied to all available data (changes from baseline in GRID-HAMD-17 total score) collected from all randomized subjects at scheduled assessments during adherence to randomized treatment, that is, through week 8 or the latest time point prior to an ICE. The model should typically include an unstructured modeling of time and within-subject correlations. The treatment contrast between the experimental treatment and placebo at week 8 is the estimate of the primary estimand (see additional details in our companion paper on estimators[3]).

4b. Missing data assumption: In this trial, some intermittently missing data may be expected because of subjects

occasionally missing a study visit while continuing with the randomized treatment. Additionally, data post ICEs as described in the element C of the estimand definition are not usable for the primary estimand even if collected and will be treated as missing. For both types of missing data, the primary analysis MMRM model assumes that subjects with missing data would have efficacy outcomes like those in similar subjects in their treatment group who continue their randomized treatment through the time point at which data are missing, up to week 8. This type of assumption is referred to as Missing at Random (MAR).[12] This assumption is aligned with the estimand and the hypothetical strategy in element C.

4c. Sensitivity estimators: Sensitivity analysis needs to be conducted to assess the robustness of conclusions from the primary analysis to missing data assumptions. The key assumption in the primary analysis is that missing data arise from an MAR mechanism. This assumption can be stress-tested via a delta-adjustment tipping-point sensitivity analysis (see Ratitch et al[14] and Mallinckrodt et al[3] for additional details).

4d. Sample size: Sample size required for the primary estimand defined above is based on the treatment effect size expected under ideal adherence through Week 8 as well as the anticipated rate of ICEs. Subsequent trials may need to allow for additional margins for sensitivity analyses or to adjust sample size for other estimands.

Confirmatory Trial in Severe Asthma

Background

Asthma is a heterogeneous chronic inflammatory respiratory disease and impacts more than 300 million people worldwide. Characterized by symptoms of wheezing, shortness of breath, chest tightness and/or cough, and accompanied by variable expiratory airflow limitation, asthma ranges from mild to severe disease,[15] associated with compromised quality of life and reduced survival.[16] Goals of asthma management include achieving symptom control, maintaining normal levels of activity, and minimizing future exacerbations to avoid long-term morbidity and mortality.[16] Early treatment increases the likelihood of improved asthma control and less additional asthma medication use.[17] In addition to addressing modifiable risk factors and nonpharmacologic approaches, subjects often step up pharmacologic therapy with increasing doses and potency of inhaled corticosteroids (ICSs), leukotriene receptor antagonists (LTRAs), theophylline, and long-acting $beta_2$-agonists (LABAs) based on continued symptomatology, receiving short-acting $beta_2$-agonists (SABAs) as needed. Those with continued symptoms may receive additional therapy with oral corticosteroids (OCSs) and/or anti-immunoglobulin E (anti-IgE) or interleukin 5 (IL-5) inhibitors.[16]

In clinical trials of new add-on treatments for subjects with severe asthma uncontrolled with high-dosage ICSs and LABAs, a placebo-controlled add-on design (standard therapy plus experimental drug vs standard therapy plus placebo) with a provision for a short-term rescue medication is the preferred approach.[18] Marketing approval of new medicines is typically based on the primary efficacy measure of clinically significant asthma exacerbations rate.[18] Clinically significant exacerbations of asthma are usually defined as a requirement for systemic corticosteroids or an increase of the maintenance dose of oral corticosteroids for at least 3 days and/or a need for an emergency visit, hospitalization or death due to asthma. A clinical trial of 1-year duration is required by regulators to assess annual exacerbation rate.[18]

In asthma trials of add-on therapies, the standard asthma controller background therapy consists of an ICS/LABA formulation. The pre-study dosage and regimen is continued throughout the study treatment period. Other allowed asthma controllers (eg, long-acting muscarinic antagonists [LAMAs], LTRAs, OCSs) that the subject may have been taking at least 30 days prior to enrolment are usually allowed during the study, but typically, prior exposure to biologic therapies would not be permitted or would require sufficient washout. SABAs via a metered dose device are also typically permitted as needed for worsening asthma symptoms, that is, for occasional short-term rescue use. However, a regularly scheduled or prophylactic (eg, prior to planned exercise) use of SABAs in absence of asthma symptoms is typically discouraged. Other changes to treatment are also typically discouraged or disallowed during the study treatment period, for example, changes to the subject's background controller regimen and use of LABAs as a reliever. Asthma exacerbations are normally treated with oral or other systemic corticosteroids according to standard practice, and the protocols typically outline the exacerbation treatment guidelines. Each study may include a list of other specific nonasthma excluded medications.

Trial Description

For illustration purposes, we consider the SIROCCO trial[19] of benralizumab. In the remainder of this section, we summarize the main features of this trial and the primary estimand used as the basis for marketing approval of benralizumab in as much as we can infer it from the study publication[19] and publicly available regulatory marketing application review documents.[20-22] Then we discuss another estimand that may be of interest for supportive purposes.

The SIROCCO trial was a randomized, placebo-controlled phase 3 study of benralizumab for subjects with severe asthma uncontrolled with high-dosage ICSs and LABAs. Subjects were to receive injections of the study drug as add-on to a stable prestudy standard-of-care therapy at clinical centers every 4 weeks for the duration of 48 weeks. Planned assessment times included the randomization visit (week 0) and visits at 4-week intervals during the treatment period (weeks 4, 8, 12, . . . , 48). The primary endpoint was the annual asthma exacerbation rate, evaluated over 48 weeks. The primary objective was to demonstrate efficacy in an enriched population of subjects with blood eosinophil counts of at least 300 cells/μL at baseline. The objective was to assess the effect of benralizumab as an add-

Table 2. Anticipated Treatment Regimens[a] in a Trial of Severe Asthma

Scenario	Treatment Regimen Over 48 wk
1	Z
2	Z→O
3	Z→C(i)
4	Z→P(i)
5	Z→P(ii)
6	Z→C(ii)
7	Z→P(iii)

[a]Treatment regimen: Z = randomized treatment as add-on to pre-study ICS/LABA regimen; O = background pre-study ICS/LABA regimen only; C = concomitantly with the randomized treatment; P = post discontinuation of randomized treatment; (i) = SABAs for worsening asthma symptoms as rescue and protocol-specified treatment for exacerbation events; (ii) = changes to the subject's background controller regimen, regular or prophylactic use of SABAs, treatment with short-acting anticholinergics or with oral or injectable corticosteroids outside of managing an asthma exacerbation event, use of LABAs as a reliever; (iii) = alternative treatment for asthma, not prespecified by study protocol.
Note: In addition to the above treatments, exacerbation events were managed per the study-defined treatment protocol.

on treatment. Therefore, subjects continued taking their background asthma controller treatments with a stable prestudy dosage and regimen during the study treatment period. The allowed rescue therapy, discouraged/disallowed medications, and management of exacerbation events in this study were similar to the typical setting of an add-on treatment trial for severe asthma subjects described above.

Primary Estimand

We now follow the steps of the clinical trial design process chart from the second section while reconstructing what the primary estimand was in the SIROCCO study.

1a. Identify decision maker: The key decision makers were the regulatory agencies.

1b. Define objective: The primary objective was to determine whether to grant marketing authorization approval by evaluating effectiveness of the experimental drug, benralizumab, compared to placebo as an add-on maintenance treatment in patients with severe asthma.

2a. Identify possible ICEs: Actual treatment regimens that may have been anticipated in the SIROCCO trial, occurring in either planned or unplanned manner, are summarized in Table 2. All scenarios in this table, except the one in the first row, represent ICEs that occur at the time point when the treatment changes from the randomized treatment (Z) to either the prestudy background therapy only (O) or a different treatment. Using the notation in Table 2, C(i) represents a protocol-allowed rescue therapy for short-term management of worsening asthma symptoms, whereas treatment changes (ii) and (iii) were discouraged or not allowed. Any of the listed treatment changes could occur at any point in the trial; their handling in the primary estimand did not depend on the timing.

Typical study treatment completion rates in similar studies range between 80% and 85%, with higher rates observed in more recent confirmatory studies. The treatment completion rates in the SIROCCO trial were 89%, 87%, and 90% for the 3 treatment groups, respectively.[19] A total of 8% took a disallowed concomitant medication, the most common of which was regularly scheduled SABAs.[20] Withdrawals and important protocol deviations were fairly balanced across treatment groups.[22]

2b. Define treatment regimen under evaluation: The treatment regimen under evaluation was the randomized treatment taken for up to 48 weeks as add-on to the subject's background prestudy ICS/LABA regimen and including protocol-defined rescue therapy and treatment of exacerbation events, as well as including any other asthma treatments that may be administered in the course of the 48-week study period as per the investigator and subject decision.

In other words, as discussed in the companion paper on estimands,[2] the effect of being assigned to one of the 3 study treatment groups was evaluated.

2c. Define estimand: The primary and secondary estimands were not explicitly defined for the SIROCCO trials. Based on the reported methods and results,[19] we infer that the primary efficacy estimand was:

A. The treatment effect was to be estimated for the population of adult and adolescent patients with severe asthma uncontrolled with high-dosage ICSs and LABAs as defined by the protocol inclusion/exclusion criteria (in the sense of delineating the population) who had blood eosinophil counts at entry of at least 300 cells/μL.
B. Efficacy was to be measured using the primary endpoint of the number of asthma exacerbations experienced by a subject over the 48-week double-blind study period.
C. All types of ICEs, including use of SABAs for worsening asthma symptoms as rescue, treatment of exacerbation events as specified in the study protocol, a premature discontinuation of the randomized treatment, and any modifications of asthma treatment including those that were discouraged/disallowed by study protocol but might have occurred as per the investigator and subject decision, were handled using the treatment policy strategy, that is, included in the treatment regimen under evaluation.
D. The annual rate of asthma exacerbation events was to be calculated for each randomized treatment group based on the data collected over the 48-week postrandomization period, and each of the experimental treatment groups was to be compared to the placebo group using the event rate ratio.

The ICEs corresponding to the use of SABAs for worsening asthma symptoms as rescue and treatment of exacerbation

events are protocol-defined treatments that are part of the standard-of-care recommended for ongoing disease management in this patient population. Based on their mechanism of action, and on considerable prior clinical experience, these therapies are not expected to produce lasting, disease-modifying effects. Apparently, a treatment policy strategy was applied for all ICEs. That is, no ICEs were a break from the treatment regimen under evaluation. Using the treatment policy approach for all ICEs seems disconnected with the trial design in that the protocol explicitly stipulated medications that were discouraged or not allowed, but the analytic approach disregarded this fact. These ICEs may have been thought as likely to occur in the general clinical practice in a small percentage of this patient population. Therefore, the overall treatment effect estimated in the presence of these ICEs was not expected to be significantly biased and was considered clinically relevant for evaluation of benralizumab effectiveness for marketing approval.

3a. Data useful for estimand: Because the treatment policy strategy was used for all types of ICEs, usable data for this estimand were the exacerbation-related data over the 48-week postrandomization period regardless of adherence to the randomized treatment. Subjects who switched to an alternative asthma treatment after they discontinued from the randomized treatment were expected to complete the remaining study visits. Subjects who had postrandomization treatment changes discouraged or disallowed by the protocol were not withdrawn from the study and continued to be followed as planned.

3b. Patient retention strategy: It was expected that the study withdrawal rate would be considerably lower than the treatment discontinuation rate because of the post-treatment discontinuation data collection effort.[21] However, the study completion rates, 90%, 89%, and 90% for placebo, benralizumab 30 mg Q4W, and benralizumab 30 mg Q8W treatment groups respectively, were only slightly higher than the treatment completion rates 89%, 87%, and 90%, respectively, for the 3 treatment groups.[19] The completion rate in the SIROCCO trial was greater than similar historical trials (80%-85%), which could indicate a general trend in recent years of regulatory and sponsors emphasizing improved subject retention.

4a. Main estimator: The SIROCCO trial analyzed the rates of exacerbation events with a negative binomial model for recurrent event data[23] with the logarithm of the subject's follow-up time used as an offset variable in the model to adjust for different follow-up times. The response variable was the number of exacerbation events reported by a subject over the double-blind treatment period. The model included covariates of treatment group, region, number of exacerbations in previous year, and the use of maintenance oral corticosteroids.

4b. Missing data assumption: Although using the treatment policy strategy for all types of ICEs generally reduces the amount of missing data as compared to other strategies, missing data may nevertheless occur (as it did in the SIROCCO trial) as a result of subjects withdrawing from the study. To account for a shorter duration of follow-up of subjects who withdrew, the primary analysis negative binomial model should include an offset term for the logarithm of follow-up duration. The negative binomial model makes the assumption that missing data were MAR, that is, that subjects who withdrew from the study would, taking into account their exacerbation rate up to the time of withdrawal, have a similar exacerbation rate postwithdrawal to the exacerbation rate of subjects in the same treatment group who remained in the study (and who have similar values of baseline characteristics included in the model).

4c. Sensitivity estimators: Sensitivity analysis focusing on the assumptions about missing data can be performed by varying assumptions about the rates of exacerbations after early study withdrawal, for example, in the SIROCCO trial it was assumed that subjects with missing data from the experimental arms had a greater exacerbation rate postwithdrawal than those who withdrew from the placebo arm.[21] A range of such assumptions can be considered to find a tipping point.[24]

Another option for handling post study withdrawal missing data for the main or sensitivity estimator could be to impute these missing data based on the model estimated from subjects who discontinued the randomized treatment but remained in the study and have available data, as they can be considered as a reference group that would fit the estimand by providing clinically plausibly poor imputed data. However, this approach may have been impractical as the number of subjects in such a reference group was small ($\sim$1% of subjects).

4d. Sample size: Some ICEs marked treatment changes that may lower the risk of exacerbation events. When data after such ICEs are used for the estimation of the overall treatment effect, the estimated treatment difference may be attenuated, and this should be taken into account in the sample size calculations at the trial design stage. The impact of loss of information due to missing data on power should also be accounted for.

Supportive Estimand

The primary estimand provided a pragmatic assessment of effectiveness, which could be considered as estimating a lower limit of the experimental drug's efficacy. A supportive estimand could indicate an upper limit of efficacy, thus enabling a decision maker to evaluate a spectrum of evidence. This can be achieved by estimating the treatment effect corresponding to adherence to randomized treatment, allowing only for the protocol-defined rescue therapy and treatment of exacerbation events. The benefit of the randomized treatment while taken as directed can subsequently be interpreted considering separate analyses of safety, tolerability, and adherence.[25] This supportive estimand would also be consistent with the stipulations of allowed and not allowed medications in the protocol, that is, using the treatment policy strategy only for ICEs of the type "Z→C(i)" in Table 2. This approach with respect to the concomitant medications would also follow the recommendations in the EMA guidelines on the clinical investigation of medicinal products for the treatment of asthma,[18] which suggests that

"concomitant and rescue therapy should be simplified where possible and documented to avoid compromising the interpretation of the data."

We now follow the trial design process chart again to define a supportive estimand.

Steps *1a (Identify decision maker*), *1b (Define objective)*, and *2a (Identify possible ICEs)* are similar to those discussed for the primary estimand, with an exception being that the objective here is to provide a supportive estimand for a broader evaluation of efficacy.

2b. Define treatment regimen under evaluation: The treatment regimen under evaluation is the randomized treatment taken as directed for up to 48 weeks, including use of SABAs for worsening asthma symptoms as rescue and treatment of exacerbation events as specified in the study protocol.

2c. Define estimand: The elements A and D for the primary endpoint of the number of exacerbation events would remain the same as in the case of the primary estimand, while the elements B and C would describe a combination of the treatment policy and while-on-treatment strategies:

B. Efficacy is to be measured using the primary endpoint of the number of asthma exacerbations reported by a subject while he or she receives the randomized treatment as directed, possibly with occasional uses of SABAs and management of exacerbation events as permitted by the protocol, up to 48 weeks of the double-blind study period.

C1. For subjects who require SABAs as rescue treatment for worsening asthma symptoms or treatment of exacerbation events as specified in the study protocol, outcomes observed during the period of these additional treatments while continuing the randomized treatment as directed are included as they are consistent with the treatment regimen under evaluation. Therefore, the treatment policy strategy is used with respect to these types of ICEs.

C2. For subjects who initiate any other changes to their treatment, including any asthma treatments or changes to the background controller therapy that are discouraged/disallowed by study protocol or a premature discontinuation of the randomized treatment, outcomes after the ICEs that mark the start of such treatment changes are irrelevant for evaluation of the treatment regimen of interest. A while-on-treatment strategy will be used with respect to these ICEs.

3a. Data useful for estimand: Data useful for this estimand are the observations collected while subjects adhere to the randomized treatment and take additional treatments only for rescue or management of exacerbation events as permitted by the protocol. All data after discontinuation of the randomized treatment and data after ICEs that mark the start of additional treatments that are discouraged or disallowed by the protocol would not be useful for this estimand and would be excluded from analyses.

Step *3b (Patient retention strategy)* would not require any changes compared to what was discussed in the context of the primary estimand.

4a. Main estimator: The main estimator for analysis of exacerbation rates would remain the same as for the primary estimand, the only difference being the data that would be included in the analysis as discussed in 3a. The offset term for the logarithm of the subject's follow-up duration in the negative binomial model accounts for the varying length of time during which the exacerbations are counted for each subject. Note that this estimand does not target the treatment effect over the full intended study period of 48 weeks for all subjects. It aims at assessing the treatment effect during the period of adherence only, regardless of the duration. However, group-level estimates may still be reported in terms of the annualized exacerbation rates for ease of interpretation and comparison with the primary estimand. In this case, the annualized exacerbation rate estimates from the negative binomial model are based on the MAR assumption, but, unlike in the primary estimand, here the extrapolation is based exclusively on the data that reflects the event rates during adherence to treatment.

4b. Missing data assumption: For the endpoint of the number of exacerbation events in the context of an estimand with while-on-treatment strategy, the only subjects with missing data are those who do not have adequate follow-up data to support determination of whether they did or did not have exacerbation events while receiving the treatment regimen of interest—for these subjects the exacerbation rate is assumed to remain constant. The while-on-treatment assumption would lead to far less missing data.

4c. Sensitivity estimators: Sensitivity estimators would be similar to the slope-adjustment (including tipping point) methods described for the primary estimand, except that they should only be applied to periods of time that are on-treatment. In other words, the analytical methods could be the same as in the primary estimand's sensitivity analysis, but the data included in the analysis would differ, aligning with the target of the supportive estimand.

Step *4d (Sample size)* would involve similar considerations as discussed in the context of the primary estimand.

Confirmatory Trial in Rheumatoid Arthritis

Background

Rheumatoid arthritis is a systemic inflammatory autoimmune disease impacting approximately 0.5% to 1% of the population. Severity ranges from mild to severe disease associated with progressive joint destruction, compromised quality of life, and reduced survival.[26,27] Remission is the optimal treatment goal because it is correlated with the prevention of structural damage and maintenance of function.[28,29] Early and aggressive treatment increases the likelihood of disease control[30,27]; however, remission rates are low despite significant advances in the

treatment during the past two decades. Patients often receive 1 or more conventional disease-modifying antirheumatic drugs (cDMARDs), with methotrexate (MTX) considered the gold standard. Often, cDMARDs are used in combination with low-dose oral or intra-articular corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs). In addition, biological agents that antagonize critical inflammatory mediators, T cells or B cells, are used with or without concomitant cDMARDs.

As an example for defining estimands, we consider a confirmatory trial of a biologic agent for MTX inadequate responder (MTX-IR) subjects. In such studies, the primary and key secondary endpoints are measures of symptom improvement and physical function measured at or after 12 weeks of treatment (with earlier time points favored in recent placebo-controlled trials in order to limit the time of exposure to placebo). Per the FDA draft guidance for marketing approval of drug products for treatment of RA,[31] the demonstration of efficacy should include clinical response and physical function using measures such as the ACR 20% response criteria (ACR20) and the Health Assessment Questionnaire-Disability Index (HAQ-DI), respectively. ACR20 is defined as at least 20% improvement in the number of tender joints and swollen joints and at least 20% improvement in 3 of the remaining 5 ACR core-set measures (subject pain, subject and physician global assessments of disease, physical functioning assessment, and acute-phase reactant). It is a binary (yes/no) endpoint. The HAQ-DI is a patient-reported outcome questionnaire that measures disease-associated disability. Although the HAQ-DI is an ordinal outcome ranging from 0 to 3, it is commonly analyzed as a continuous endpoint in terms of change from baseline.

Most new biologics for MTX-IR patients are tested in combination with stable background MTX therapy, determined during the prerandomization period. In some trials, the biologic may be tested both in combination with MTX and as monotherapy, but placebo is typically administered in combination with MTX. After randomization, several changes in treatment may be anticipated—some planned and some unplanned. Most trials in MTX-IR patients have a planned assessment of minimal required response to treatment, for example, $\geq$20% improvement from baseline in both tender and swollen joint counts at a specific time point. Subjects not meeting the minimal improvement are offered rescue therapy for ethical reasons. Rescue may involve adjustments to background therapy, for example, an increase of MTX dose or change in route of administration; addition of other cDMARDs such as sulfasalazine or hydroxychloroquine; increase in NSAID or prednisone dose, change in NSAID, or new NSAID or prednisone start; or intra-articular corticosteroid administration, or any combination of the preceding. Additionally, RA studies often include as rescue an "escape" (or "step-up") therapy with the IP, where subjects randomized to placebo are switched to the active experimental drug and subjects randomized to a lower dose of active drug are switched to a higher dose.

These escape treatment switches are typically implemented in a blinded manner and triggered by a protocol-defined requirement for rescue, such as the minimal required response mentioned above. Escape therapy may confound and complicate evaluations of some estimands at time points after its initiation but allows for longer duration of exposures to IP, for example, when the placebo-treated subject is switched to IP, to supplement the safety database. If escape therapy to a higher dose of IP occurs, it may also help to answer whether dose titration is a viable option to implement in clinical practice or not. However, historically in RA trials, the treatment effect evaluated as per the primary efficacy objectives typically excluded the confounding effects of escape-type rescue. Switching to other nonstudy biologic agents typically is not part of the protocol-allowed rescue but may occur as a result of physician's and patient's decision. All the above-mentioned postrandomization treatment changes constitute ICEs and require careful consideration in the estimand definition.

Trial description

We will now discuss a hypothetical trial with design elements that resemble the historical phase 3 study of golimumab, GO-FORWARD.[32] Our example trial is a 24-week double-blinded placebo-controlled phase 3 trial in MTX-IR subjects evaluating an IP that is an injectable biologic agent. The 4 treatment arms are (1) placebo injections plus MTX capsules, (2) IP injections at high dose as monotherapy (ie, with placebo capsules instead of MTX), (3) IP injections at low dose plus MTX capsules, or (4) IP injections at high dose plus MTX capsules. Injections are to be administered every 4 weeks.

The primary efficacy evaluation is based on co-primary endpoints: ACR20 at week 14 and change from baseline to week 24 in HAQ-DI score. Although HAQ-DI can be measured at earlier time points, function typically follows symptomatic improvement and may continue to increase over time.[33]

Estimand for an RA Study Design 1

We first consider a study design and a primary efficacy objective mimicking the GO-FORWARD study as inferred from the description of methods in published material.[32] We define an estimand in that context following the trial design process chart presented in the second section.

1a. Identify decision maker: The key decision makers are the regulatory agencies.

1b. Define objective: The primary objective was to determine whether to grant marketing authorization approval by evaluating effectiveness of the experimental drug compared to placebo at specified time points in MTX-IR patients when taken as an add-on treatment without any modifications of therapy postrandomization.

2a. Identify possible ICEs: Table 3 presents several scenarios of treatment sequences that may occur in subjects in this type of RA trial. Ideally, all subjects would stay on the randomized treatment (Z) through week 24. As in the GO-FORWARD study, the need for rescue is to be assessed at week 16 based on predefined criteria, and the rescue offered

Table 3. Anticipated Treatment Regimens[a] in a Trial of RA in MTX-IR Subjects.

	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Treatment Regimen Over 24 wk
1	Z	Z	Z	Z	Z	Z	Z
2	Z	Z	O	O	O	O	Z→O
3	Z	Z	Z	E	E	E	Z→E
4	Z	P(i)	P(i)	P(i)	P(i)	P(i)	Z→P(i)
5	Z	C(i)	C(i)	C(i)	C(i)	C(i)	Z→C(i)
6	Z	P(ii)	P(ii)	P(ii)	P(ii)	P(ii)	Z→P(ii)
7	Z	Z	Z	E	P(ii)	P(ii)	Z→E→P(ii)
8	Z	Z	Z	E+C(i)	E+C(i)	E+C(i)	Z→ E+C(i)

[a]Treatment regimen: Z = randomized treatment as add-on to pre-study MTX regimen; O = background pre-study MTX treatment only; P = post discontinuation of randomized treatment; C = concomitantly with the randomized treatment; E = escape therapy; (i) = increased dose of MTX above the baseline dose for treatment of RA, new conventional disease-modifying anti-rheumatic drugs (DMARDs) or systemic immunosuppressive agents, treatment with oral corticosteroids for RA (new or dose above the baseline dose), or intravenous or intramuscular administration of corticosteroids for RA; (ii) = alternative biologic agents for RA, not pre-specified by study protocol.

as part of the study is the escape therapy discussed above. All the treatment modifications implemented as part of this escape therapy were performed in a double-blind manner. Escape therapy is denoted by "E" in Table 3.

In addition to the planned rescue, it is anticipated that the study investigators may occasionally modify the treatment per clinical judgment. For example, investigators could (i) increase dose of MTX above the baseline dose for treatment of RA, or initiate new cDMARDs or systemic immunosuppressive agents, or modify treatment with oral, intravenous, or intramuscular corticosteroids for RA (new or dose above the baseline dose); (ii) initiate nonstudy biologic agents for RA. In our example, treatment modifications (i) may be initiated either after a permanent discontinuation from the study treatment (denoted by "P" in Table 3) or concomitantly with the randomized treatment (denoted by "C" in Table 3) as modifications in the background therapy. Initiation of a new nonstudy biologic would normally occur after the permanent discontinuation of the study treatment in this patient population. Some subjects may discontinue the randomized treatment and not initiate any new treatment before the time point of interest, that is, remain on their background therapy only, as denoted by "O" in Table 3. Other subjects may initiate the escape therapy, but then also initiate other changes as, for example, shown in row 7 of Table 3. We do not consider an exhaustive list of possibilities as to an exact timing of the treatment modifications—the important point is that they may occur before the time points at which efficacy needs to be established, for example, week 14 or week 24 in our example, and therefore the changes in treatment will affect the interpretation of treatment effect even if all subjects are fully assessed through week 24. Note that one of the two co-primary efficacy endpoints, HAQ-DI, is measured at week 24—that is, after escape therapy could be initiated.

All treatment sequences in Table 3, except scenario 1, contain ICEs for which a strategy needs to be specified as part of the estimand definition.

2b. Define treatment regimen under evaluation: The treatment regimen under evaluation is the randomized treatment taken for up to 24 weeks without any adjustments to the background therapy but allowing early discontinuation because of reasons other than lack of efficacy.

2c. Define estimand: The primary estimand that mimics the one inferred from the GO-FORWARD study is as follows:

A. The treatment effect will be estimated for the population of adult subjects with active RA despite MTX therapy (MTX-IR) as defined by the protocol inclusion/exclusion criteria (in the sense of delineating the population).
B. Efficacy will be measured using two co-primary endpoints: ACR20 at week 14 and change from baseline in the HAQ-DI score at week 24.
C1. For subjects who prematurely discontinue the randomized treatment for reasons other than lack of efficacy and do not initiate any adjustments to the background therapy, observed outcomes at the designated time points provide evidence compatible with the treatment regimen under evaluation as defined in 2b above. Therefore, the treatment policy strategy will be used for this ICE.
C2. For subjects who initiate a protocol-defined escape therapy, a hypothetical strategy will be used to estimate what the treatment effect would have been at the designated time point if subjects did not receive the escape therapy and continued on their randomized treatment.
C3. Subjects who prematurely discontinue the randomized treatment for lack of efficacy or initiate any treatment adjustments other than the protocol-defined escape therapy will be considered treatment failures at the designated time points after the start of such treatment changes. Therefore, the composite strategy is used for these types of ICEs.
D. The proportion of subjects with an ACR20 response at Week 14 will be estimated for each randomized

treatment group, and each of the experimental treatment groups will be compared to the placebo group using absolute differences of proportions. Median change from baseline to week 24 in HAQ-DI score will be computed for each treatment group, and a hypothesis of no difference between each of the experimental treatment groups and placebo group will be tested based on the composite outcome (as defined in part C3) converted to ranks, with treatment failures assigned the worst rank. An estimate of median treatment differences will be used to quantify the difference between each experimental treatment group and placebo for the HAQ-DI score.

The treatment policy strategy for ICEs described in C1 accounts for imperfect compliance, including early discontinuation of study treatment in the absence of evidence (or perception) of lack of efficacy. A determination of the primary reason for discontinuation, however, could be based on subjective judgements (of patients and/or investigators) and not on formal criteria such as ACR20. It is, therefore, important to provide clear guidance in the protocol for determining the primary reason for discontinuation and close monitoring of these data during the study.

The strategies described in C2 and C3 aim at estimating the effect of randomized treatment without any confounding by the effect of other medications.

A composite strategy used for all other ICEs as described in C3 is based on interpreting all such events as study treatment failures and assuming that continuing with the randomized treatment alone would provide no chance of improvement at a later time.

3a. Data useful for estimand: Usable data that should be collected for this estimand are measurements used in the ACR20 response evaluation and HAQ-DI scores at baseline and week 14 and week 24 (for the two endpoints respectively) for all subjects except those with ICEs described in C2 and C3 above.

In the GO-FORWARD trial, subjects who prematurely discontinued the randomized treatment continued to be evaluated for safety and selected efficacy assessments for 4 months after the last dose of study treatment. In more recent RA trials, a typical regulatory recommendation is to continue study participation (with efficacy and safety evaluations) for the duration of the double-blind period, with possibly limiting the assessments to the essential evaluations, and these additional data used for supportive analyses.

3b. Patient retention strategy: Retention strategies can focus on trial features to reduce discontinuations from the study. Offerring adjustments to background therapy and escape therapy, as planned for this study, tend to help with this objective. Including an option of a simplified schedule of assessment after discontinuation of the randomized treatment can also improve patient retention in the study follow-up. Efforts should be made to minimize missing data.[12]

4a. Main estimator: For analysis of ACR20 response at week 14, standard methods for estimation of proportions and their differences can be used, with the hypothesis test carried out using, for example, a chi-square test. For analysis of changes from baseline to week 24 in HAQ-DI scores, a rank-based method, for example, Wilcoxon rank-sum test, can be used for hypothesis testing. Subjects with ICEs who are considered as treatment failures for this estimand are assigned the worst rank. To obtain an estimate of median treatment differences, for example, a Hodges-Lehmann estimate, the treatment failure outcome attributed to subjects with ICEs needs to be represented by some numerical value that is worse than any observed value. A careful choice of such assigned value should not have material impact on the Hodges-Lehmann estimate, given a study of sufficient size. Other treatment difference measures could also be used in conjunction with a rank-based analysis, for example, win ratio[34] or difference of trimmed means.[35]

The estimators also must accommodate the fact that subjects with ICEs, as described in C2 of the estimand definition, will not have observed outcomes that can be used for the estimand because they do not pertain to the regimen to be assessed, and need to be handled with a hypothetical strategy. In this study, all subjects requiring rescue as per the protocol-defined criteria are expected to initiate the escape rescue, and therefore no reference group with available data can be identified for estimating a statistical model to predict unavailable outcomes for subjects handled with the hypothetical approach of C2. In this case, additional estimation assumptions are necessary. In the GO-FORWARD trial, it was assumed that subjects who require rescue, would not improve or worsen if they remained on their randomized treatment between weeks 16 and 24. This assumption was implemented using a Last Observation Carried Forward (LOCF) single imputation approach. Single imputation can lead to underestimation of variance, but an LOCF-like approach can also be implemented using multiple imputation.[36,37]

4b. Missing data assumption: Data may be missing intermittently if a subject without any ICEs described in C2 and C3 of the estimand definition misses the required assessments at week 14. Missing data would also arise if subjects decide to withdraw from the study overall after discontinuing from the randomized treatment because of reasons other than efficacy. In these cases, it may be reasonable to assume that the missing outcomes would be similar to those of subjects with similar baseline and previous postbaseline values in their treatment group (the MAR assumption). Multiple imputation can be used to impute these missing values. The amount of such missing data should be limited in a well-executed study.

Unobserved outcomes of subjects with ICEs as described in C2 of the estimand definition are assumed to be similar to their outcomes prior to escape initiation.

4c. Sensitivity estimators. To assess sensitivity to missing data, a more extreme assumption is often used, where all subjects with missing/unobserved data as described above are

considered as treatment failures. Delta adjustment/tipping point analyses can also be performed.[3]

4d. Sample size: Sample size requirements should be based on assumptions that incorporate the likely rates of ICEs described in C1 to C3 of the estimand definition above and their impact on the overall treatment effect. Subjects should be encouraged to continue their participation in the study in cases of premature discontinuation of the randomized treatment. In case of discontinuations due to reasons other than lack of efficacy, their data are critical, as they are used for the primary estimand; in other cases, their data are valuable for supportive analyses.

Estimand for RA Study Design 2

To illustrate a broader range of possibilities, we describe a different study design and define an estimand in this new context. There are two key differences between study designs 1 and 2: considerations for the premature discontinuations of the randomized treatment and implementation of a protocol-defined rescue therapy. All premature discontinuations of the randomized treatment will be considered treatment failures regardless of discontinuation reason to avoid relying on subjective judgments.The rescue therapy will now have two components: (1) protocol-defined adjustments to the background therapy will be made for all subjects who meet rescue criteria at week 16; (2) additionally, subjects requiring rescue will be randomized at week 16 to either initiate the escape therapy or not in a blinded manner.

1a. Identify decision maker: The key decision makers are the regulatory agencies.

1b. Define objective: The primary objective is to determine whether to grant marketing authorization approval by evaluating effectiveness of the experimental drug compared to placebo at specified time points in MTX-IR patients when taken as an add-on treatment *allowing for specific adjustments to the background therapy commonly undertaken in clinical practice* (note, the wording in italics replaces the following wording in the RA study design 1 above: "without any modifications of therapy post randomization").

2a. Identify possible ICEs: The details of anticipated ICEs listed in Table 3 need to be refined to split the treatment changes mentioned under "(i)" into adjustments to the background therapy that will be part of the protocol-allowed rescue therapy and those that will not. The need for these allowed adjustments will not be considered a treatment failure. On the contrary, once these adjustments are made, the subject can improve meaningfully above and beyond of what is expected from the effect of background therapy if she or he continues the randomized treatment. For example, the background adjustments may allow the subject to reach the minimal required response in a short term, but continuing with the experimental treatment might provide further benefit later as some subjects take more time to respond than others. The allowed adjustments will be prespecified in the protocol and could include an increase of MTX dose or change in route of administration; addition of other cDMARDs such as sulfasalazine or hydroxychloroquine; new NSAID or change in NSAID dose; modifications of corticosteroids use; or any combination of the preceding. Prespecification enables inferences about a specific treatment regimen and an unambiguous interpretation and comparison with other treatments in the future.

For the primary evaluation of efficacy, the confounding effect of treatment switching on HAQ-DI at week 24 in subjects who initiate the escape therapy must be removed. We can still employ a hypothetical strategy for what would happen if the subject continued with the treatment regimen under evaluation without the escape therapy. However, to implement this hypothetical strategy in a more robust manner so that it does not rely solely on assumptions, subjects requiring rescue will be randomized in a blinded manner to either initiate escape or not, so that data can be collected from some subjects that actually followed the hypothesized scenario (ie, if the subject continued with the treatment regimen under evaluation without the escape therapy). Note that in line with recent regulatory recommendations, all subjects meeting requirements for rescue will initiate protocol-defined changes in their background therapy regardless of whether they are randomized to escape or not.

2b. Define treatment regimen under evaluation: The treatment regimen under evaluation is the randomized treatment taken for up to 24 weeks possibly with protocol-defined adjustments to the background therapy as rescue.

2c. Define estimand: Elements A, B, and D of the estimand definition are similar to those specified for RA study design 1, so we focus on an alternative for element C, that is, handling of ICEs.

For subjects who require rescue and have their background therapy adjusted as allowed per protocol without initiating an escape therapy, observed outcomes at the designated time points provide nonconfounded evidence for the effect of the treatment regimen under evaluation. Therefore, the treatment policy strategy is used with respect to these types of ICEs.

For subjects who initiate a protocol-defined escape therapy, a hypothetical strategy is used to estimate what the treatment effect would be at the designated time point if subjects did not receive the escape therapy and continued on their randomized treatment with protocol-allowed adjustments to the background therapy.

Subjects who prematurely discontinue the randomized treatment for any reason or initiate any treatment adjustments other than the protocol-allowed modifications in background therapy are considered treatment failures at the designated time points after discontinuation. Therefore, the composite strategy is used for these types of ICEs.

Part D of the estimand definition could be modified compared to the previous specification by choosing a binary endpoint for HAQ-DI, where a subject is defined as responder if she or he experiences a clinically meaningful improvement in the HAQ-DI score, defined as 0.22 or greater reduction from

baseline.[38] In this case, subjects considered treatment failures as per C3 would be considered nonresponders on this endpoint, as done for the ACR20. This alternative should be evaluated in terms of its impact on sample size.

Considerations for items *3a (Data useful for estimand)* and *3c (Patient retention strategy)* are similar as in the case of the RA study design 1.

4a. Main estimator: Analysis considerations for this estimand are similar to those discussed for RA study design 1, except for handling of subjects with ICEs described in C2. Rather than an LOCF approach, outcomes under the hypothetical scenario (if the subject requiring rescue continued with the treatment regimen under evaluation without the escape therapy) can be modeled based on data from subjects who actually follow that scenario. Data from rescued subjects who are randomized not to initiate the escape therapy are used to fit a statistical multiple imputation model that is used to estimate hypothetical outcomes for subjects who were randomized to escape. This multiple imputation model should include baseline covariates and postbaseline assessments prior to rescue and can be implemented using reference-based imputation.[3]

4b. Missing data assumption: Considerations for this estimand are similar to those for the estimand of RA study design 1, except for the assumption used with the hypothetical strategy. Subjects with ICEs described in C2 are assumed to have similar efficacy outcomes as subjects in their treatment group who also met conditions for rescue therapy and had their background therapy adjusted in the protocol-defined manner without receiving the escape therapy.

4c. Sensitivity estimators. Similar sensitivity analyses as mentioned for the RA study design 1 can be used.

Considerations for item *4d (Sample size)* are similar as in the case of the RA study design 1.

Discussion

This paper illustrated examples of defining estimands consistent with the concepts outlined in ICH E9(R1) and discussed in our companion papers on estimands[2] and estimators.[3] The 3 example indications illustrated a variety of ICEs that can be anticipated in each setting as well as strategies that can be used to handle them consistently with study objectives and the clinical context. The estimands chosen for these examples are not the only acceptable choices for their respective scenarios. As previously stated, the intent was to illustrate the process and key concepts rather than focus on justification of specific choices.

Emphasis was placed on following the study design process chart in the second section. Following the steps outlined in that process, a development team should arrive at a suitable estimand without the need for iterative revisions of the study design to achieve alignment between objectives and planned inferences. ICH E9(R1) emphasized the importance of defining the estimands before choosing suitable estimators and of defining estimands that are not overburdened by statistical details so that team members from all backgrounds can understand and contribute to the estimand definition. Our examples have illustrated that this is feasible, and they promote a thoughtful consideration of the clinical context and decision-making objectives, which is especially important in complex settings where many types of ICEs can be anticipated. Nevertheless, in some instances, iterative revisions to study design may be required. For example, when sample size requirements are established and statistical power is evaluated for suitable estimators, it may become evident that alternative strategies for handling ICEs are required to ensure the study is feasible in terms of patient recruitment, timelines, and budget. Sample size calculations in this framework need to take account of various factors related to strategies chosen to handle different types of ICEs, which may require considering a range of assumptions and necessitate simulations.

It is advisable to follow the study design process chart for each objective that is envisaged for the trial. For example, in the companion paper,[2] we discussed considerations for safety estimands and estimands related to secondary efficacy parameters such as health-related quality of life, which may require different approaches than the primary efficacy estimand. Using the process for each major objective ensures that the design and data collection are adequate for all trial needs.

It is also advisable to verify the design considerations with respect to different stakeholders. For example, confirmatory trials provide the basis for the marketing approval and are also used by payers for Health Technology Assessment. The latter may be based primarily on considerations of effectiveness of prescribing/buying a therapy regardless of subsequent patterns of adherence (although an assumption of similarity between adherence patterns in the clinical trial and in practice must be made). The regulators, on the other hand, may be more interested in the risk/benefit of taking a specific experimental therapy rather than in the effect of being randomized to it. Patients and physicians may be interested in both aspects, although for these stakeholders the main consideration for initiating a new treatment typically is about what can be expected if it is taken as prescribed. To address various perspectives and priorities, it would be valuable to formulate several supporting estimands for a trial, where the estimands employ different strategies to deal with the same ICEs in alignment with the stakeholders' primary objectives. To aid in interpretation and comparison of the results under different estimands, the use of advanced visualization techniques is advisable.

Acknowledgments

DIA Scientific Working Group on Missing Data and the authors' respective employers.

Declaration of Conflicting Interests

James Bell's contributions came from work that has been sponsored and initiated by Boehringer-Ingelheim Pharma GmbH & Co. KG. Bohdana Ratitch's and Ilya Lipkovich's contributions came primarily during previous employment at IQVIA.

Funding

No financial support of the research, authorship, and/or publication of this article was reported.

ORCID iD

Michael O'Kelly, PhD https://orcid.org/0000-0003-0419-663X

References

1. ICH Harmonised Guideline E9(R1). Estimands and Sensitivity Analysis in Clinical Trials. Step 1 version dated 16 June 2017.
2. Ratitch B, Bell J, Mallinckrodt CH, et al. Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) into Practice. Submitted to *Therapeutic Innovation & Regulatory Science*. 2018.
3. Mallinckrodt CH, Bell J, Ratitch B, et al. Technical and Practical Considerations in Aligning Estimators with Estimands in Clinical Trials. Submitted to *Therapeutic Innovation & Regulatory Science*. 2018.
4. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602.
5. Mitchell J, Trangle M, Degnan B, et al. Institute for Clinical Systems Improvement. Adult Depression in Primary Care; 2013. http://pcptoolkit.beaconhealthoptions.com/wp-content/uploads/2016/02/ICSI_Depression.pdf. Accessed March 27, 2019.
6. Practice guideline for the treatment of patients with major depressive disorder, third edition. https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf. Published 2010. Accessed March 27, 2019.
7. Major Depressive Disorder: Developing drugs for treatment guidance for industry, draft guidance, revision 1. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM611259.pdf. Published June 2018. Accessed March 27, 2019.
8. Potter WZ, Mallinckrodt CH, Detke MJ. Controlling placebo response in drug development: lessons from psychopharmacology. *Pharm Med*. 2014;28:53-65.
9. Mallinckrodt CH, Tamura RN, Tanaka Y. Improving signal detection and reducing placebo response in psychiatric clinical trials. *J Psychiatr Res*. 2011;45:1202-1207.
10. Goldstein DJ, Mallinckrodt CH, Lu Y, Demitrack MA. Duloxetine in the treatment of major depression: A double-blind clinical trial. *J Clin Psychiatry*. 2002;63(3):225-231.
11. Williams JB, Kobak KA, Bech P, et al. The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. *Int Clin Psychopharmacol*. 2008;23(3):120-129.
12. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010.
13. Mallinckrodt CH, Lipkovich I. A practical guide to analyzing longitudinal clinical trial data. Boca Raton, FL: Chapman and Hall/CRC; 2017.
14. Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat*. 2013;12(6):337-47.
15. Global Asthma Network. The Global Asthma Report. http://globalasthmareport.org. Published 2014.
16. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. www.ginasthma.org. Published 2018. Accessed March 27, 2019.
17. Busse WW, Pedersen S, Pauwels RA, et al. The Inhaled Steroid Treatment as Regular Therapy in Early Asthma (START) Study 5-year follow-up: effectiveness of early intervention with budesonide in mild persistent asthma. *J Allergy Clin Immunol*. 2008;121(5):1167-1174.
18. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). Guideline on the clinical investigation of medicinal products for the treatment of asthma. Version 22, October 2015. http://www.ema.europa.eu. Accessed March 27, 2019.
19. Bleecker ER, FitzGerald JM, Chanez P, et al. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial. *Lancet*. 2016;388(10056):2115-2127.
20. Food and Drug Administration, Center for Drug Evaluation and Research. Application Number: 761070Orig1s000, Clinical Review(s); 2017. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761070Orig1s000MedR.pdf. Accessed March 27, 2019.
21. Food and Drug Administration, Center for Drug Evaluation And Research. Application Number: 761070Orig1s000, Statistical Review(s); 2017. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761070Orig1s000StatR.pdf. Accessed March 27, 2019.
22. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). Fasenra Assessment Report; 2017. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/004433/WC500245333.pdf. Accessed March 27, 2019.
23. Keene ON, Jones MRK, Lane PW, Anderson JV. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat*. 2007;6:89-97.
24. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharm Stat*. 2014;13:258-264.
25. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials—broadening the perspective. *Stat Med*. 2017;36(1):5-19.
26. Smolen J, Steiner G. Therapeutic strategies for rheumatoid arthritis. *Nat Rev Drug Discov*. 2003;2:473-488.
27. Colmegna I, Ohata BR, Menard HA. Current understanding of rheumatoid arthritis therapy. *Clin Pharmacol Ther*. 2012;91:607-620.

28. Smolen JS, Breedveld FC, Burmester GR, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis*. 2016;75:3-15.
29. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis*. 2011;70:404-413.
30. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet*. 2009;373(9664):659-672.
31. Food and Drug Administration. Guidance for Industry: *Rheumatoid arthritis: developing drug products for treatment, draft dated* May 2013. https://www.fda.gov/downloads/drugs/guidancecompliancеregulatoryinformation/guidances/ucm354468.pdf. Accessed March 27, 2019.
32. Keystone EC, Genovese MC, Klareskog L, et al. Golimumab, a human antibody to tumour necrosis factor α given by monthly subcutaneous injections, in active rheumatoid arthritis despite methotrexate therapy: the GO-FORWARD Study. *Ann Rheum Dis*. 2009;68:789-796.
33. Radner H, Alasti F, Smolen JS, Aletaha D. Physical function continues to improve when clinical remission is sustained in rheumatoid arthritis patients. *Arthritis Res Ther*. 2015;17:203.
34. Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharm Stat*. 2016;15:238-245.
35. Permutt T, Li F. Trimmed means for symptom trials with dropouts. *Pharm Stat*. 2017;16:20-28.
36. Carpenter J, Roger J, Kenward M. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat*. 2013;23(6):1352-1371.
37. O'Kelly M, Ratitch B. *Clinical Trials With Missing Data: A Guide for Practitioners*. West Sussex, UK: John Wiley & Sons, Inc.; 2014.
38. Strand V, Balbir-Gurman A, Pavelka K, et al. Sustained benefit in rheumatoid arthritis following one course of rituximab: improvements in physical function over 2 years. *Rheumatology*. 2006;45:1505-1513.