

INTERNATIONAL CONCIL FOR HARMONISATION OF TECHNICAL
REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

Estimands and Sensitivity Analysis in Clinical Trials

E9(R1)

Current *Step 2* version
dated 16 June 2017

At Step 2 of the ICH Process, a consensus draft text or guideline, agreed by the appropriate ICH Expert Working Group, is transmitted by the ICH Assembly to the regulatory authorities of the ICH regions for internal and external consultation, according to national or regional procedures.

ICH E9(R1) Technical Document
Estimands and Sensitivity Analysis in Clinical Trials
STEP 2 DRAFT GUIDELINE

TABLE OF CONTENTS

1. Purpose and Scope	1
2. A Framework to Align Planning, Design, Conduct, Analysis and Interpretation	3
3. Estimands.....	4
3.1. Description.....	4
3.2. Strategies for Addressing Intercurrent Events.....	5
3.3. Construction of Estimands	7
3.3.1. <i>General Considerations</i>	7
3.3.2. <i>Considerations of Therapeutic and Experimental Context</i>	8
4. Impact on Trial Design and Conduct	10
5. Impact on Trial Analysis	11
5.1. Main Estimation	11
5.2. Sensitivity Analysis.....	12
5.2.1. <i>Role of Sensitivity Analysis</i>	12
5.2.2. <i>Choice of Sensitivity Analysis</i>	13
5.3. Supplementary Analysis	13
6. Documenting Estimands and Sensitivity Analysis.....	14
7. A Generic Example	15
7.1 One Intercurrent Event	16
7.2 Two Intercurrent Events	19
Glossary	22

1 **1. Purpose and Scope**

2 To properly inform the choices that are made by patients and prescribing physicians, clear
3 descriptions of the effects of a medicine should be available. These descriptions are
4 complicated by the different ways in which each individual patient responds to treatment.
5 Some subjects will tolerate a medicine and adhere to its administration schedule, others will
6 not. Some subjects will require changes in dose of concomitant medication or administration
7 of additional medication, others will not. Multiple ways to quantify treatment effects can be
8 envisaged based on how to take into account, for example, tolerability, adherence and
9 whether or not additional medication is required. Without a precise understanding of the
10 treatment effect that is being described, there is a risk that its magnitude and meaningfulness
11 will be misunderstood.

12
13 Confirmatory clinical trials, usually randomised controlled trials, are conducted to quantify
14 the effects of a treatment and to provide evidence of efficacy and safety to support regulatory
15 decision making. Randomised trials are expected to be free from baseline confounding but,
16 in trials as in clinical practice, certain events will occur that complicate the description and
17 interpretation of treatment effects. In this addendum, these are denoted as intercurrent events
18 (see Glossary) and include, among others, use of an alternative treatment (e.g. a rescue
19 medication, a medication prohibited by the protocol or a subsequent line of therapy),
20 discontinuation of treatment, treatment switching and terminal events such as, in some
21 circumstances, death.

22
23 Choosing and defining efficacy and safety variables as well as standards for data collection
24 and methods for statistical analysis without first addressing the occurrence of intercurrent
25 events will lead to ambiguity about the treatment effect to be estimated and potential
26 misalignment with trial objectives. The correct order is the reverse. Having clarity in the
27 trial objectives and accounting explicitly for intercurrent events when describing the
28 treatment effect of interest at the planning stage should inform choices about trial design, data
29 collection and statistical analysis.

30
31 This addendum presents a structured framework to link trial objectives to a suitable trial
32 design and tools for estimation and hypothesis testing. This framework introduces the
33 concept of an estimand (see Glossary), translating the trial objective into a precise definition
34 of the treatment effect that is to be estimated (Section A.3). It aims to facilitate the dialogue
35 between disciplines involved in clinical trial planning, conduct, analysis and interpretation, as
36 well as between sponsor and regulator, regarding the treatment effects of interest that a
37 clinical trial should address. The statistical analysis, aligned to the estimand, will be
38 associated with assumptions and data limitations, the impact of which can be investigated
39 through sensitivity analysis (see Glossary). This addendum clarifies the definition and the
40 role of sensitivity analysis. References to the original ICH E9 are made using x.y.
41 References within this addendum are made using A.x.y.

42
43 This addendum clarifies and extends ICH E9 in a number of respects.

44
45 Firstly, ICH E9 introduced the intention-to-treat (ITT) principle in connection with the effect
46 of a treatment policy, i.e. the effect of treatment initially assigned at baseline, regardless of
47 adherence to the planned course of treatment, indicating that preservation of randomisation
48 provides a secure foundation for statistical tests. It remains undisputed that randomisation is
49 a cornerstone of controlled clinical trials and that analysis should aim at exploiting the

50 advantages of randomisation to the greatest extent possible. However, the question remains
51 whether understanding the effect of a treatment policy always targets the treatment effect of
52 greatest relevance to regulatory and clinical decision making. The framework outlined in this
53 addendum gives a basis for discussing other treatment effects and some points to consider for
54 the design and analysis of trials to give estimates of these treatment effects that are reliable
55 for decision making.

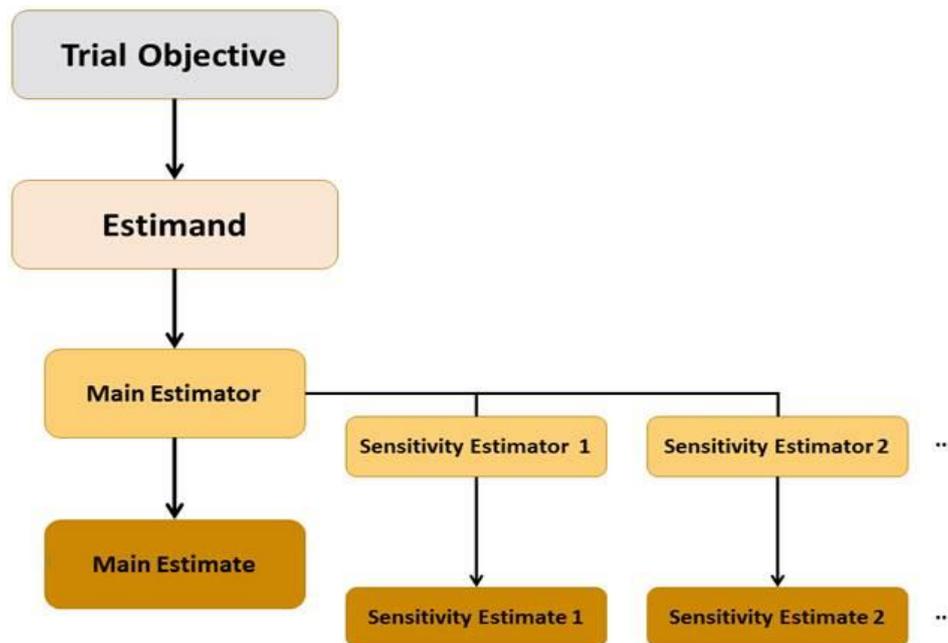
56
57 Secondly, issues considered generally under data handling and missing data (see Glossary)
58 are re-visited. On one hand, intercurrent events such as discontinuation or switching of
59 treatment, or use of rescue medication, may in some circumstances render the later
60 measurements of the variable irrelevant or difficult to interpret even when it can be collected.
61 In the case of death, measurements after a subject dies do not exist. On the other hand, ICH
62 E9 noted the difficulty of fulfilling the ITT principle when clinical trial subjects
63 discontinuing treatment were lost to follow up. This addendum invites consideration of the
64 important distinction between non-adherence with, or withdrawal from, randomised treatment
65 and discontinuation from the trial; also between measurements that exist but have not been
66 collected, and measurements that do not, or cannot, exist. Having clarity in the estimand
67 gives a basis for planning which data need to be collected and hence which data, when not
68 collected, present a missing data problem to be addressed. In turn methods to address the
69 problem presented by missing data can be selected to align with the chosen estimand.

70
71 Thirdly, the concept of analysis sets is considered in the proposed framework. Section 5.2
72 strongly recommends that analysis of superiority trials be based on the full analysis set,
73 defined to be as close as possible to including all randomised subjects. However, trials often
74 include repeated measurements on the same subject. Elimination of some planned
75 measurements on some subjects, perhaps because the measurement is considered irrelevant or
76 difficult to interpret, can have similar consequences to excluding subjects altogether from the
77 full analysis set, i.e. that the initial randomisation is not fully preserved. In addition, a
78 meaningful value of the outcome variable might not exist, as when the subject has died.
79 Section 5.2 does not directly address these issues. Clarity is introduced by carefully defining
80 the treatment effect of interest in a way that determines the population of subjects to be
81 included in the estimation of that treatment effect and the observations from each subject to
82 be included in the analysis considering the occurrence of intercurrent events. The meaning
83 and role of the per-protocol analysis is also re-visited in this addendum; in particular whether
84 the need to explore the impact of protocol violations and deviations can be addressed in a
85 way that is less biased and more interpretable than naïve analysis of the per protocol set.

86
87 Finally, the concept of robustness is given expanded discussion under the heading of
88 sensitivity analysis. In particular, a distinction is made between the sensitivity of inference to
89 the particular assumptions of a particular analysis and the sensitivity to the choice of analytic
90 approach more broadly. With precise specification of an agreed estimand and a statistical
91 analysis that is both aligned to the estimand and pre-specified to a level of detail that it can be
92 replicated precisely by a third party, regulatory interest can focus on sensitivity to deviations
93 from assumptions and limitations in the data in respect of a particular analysis.

94 **2. A Framework to Align Planning, Design, Conduct, Analysis and Interpretation**

95 To promote coherence and clarity, trial planning should proceed in sequence (Figure 1).
96 Clear trial objectives should be translated into key scientific questions of interest by defining
97 suitable estimands. An estimand defines the target of estimation for a particular trial
98 objective (i.e. “what is to be estimated”) through specification of: the population, the
99 variable, the handling of intercurrent events, and the population-level summary for the
100 variable (Section A.3). A suitable method of estimation (i.e. the analytic approach, referred
101 to as the main estimator) can then be selected. The main estimator will be underpinned by
102 certain assumptions. To explore the robustness of inferences from the main estimator to
103 deviations from its underlying assumptions, a sensitivity analysis should be conducted, in
104 form of one or more analyses, targeting the same estimand (Section A.5).
105



106
107 **Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis,**
108 **for a given trial objective**
109

110 This framework enables proper trial planning that clearly distinguishes between the target of
111 estimation (trial objective, estimand), the method of estimation (estimator, resulting in an
112 estimate, see Glossary), and a sensitivity analysis. This will assist sponsors in planning trials,
113 regulators in their reviews, and will enhance the interactions between these parties when
114 discussing the suitability of clinical trial designs, and the interpretation of clinical trial results,
115 to support drug licensing.

116
117 In general, it is important to proceed sequentially, and not for the choice of an estimator to
118 determine the estimand, and hence the scientific question that is being addressed.

119
120 The specification of appropriate estimands (See A.3.3) will usually be the main determinant
121 for aspects of trial design, conduct (Section A.4) and analysis (Section A.5).

122 3. Estimands

123 3.1. Description

124 A central question for drug development and licensing is to quantify treatment effects: how
125 the outcome of treatment compares to what would have happened to the same subjects under
126 different treatment conditions (e.g. had they not received the treatment or had they received a
127 different treatment). Intercurrent events need to be considered in the description of a
128 treatment effect on a variable of interest because both the value of the variable and the
129 occurrence of the event may depend on treatment. The definition of a treatment effect,
130 specified through an estimand, should consider whether values of the variable after an
131 intercurrent event are relevant, as well as how to account for the (possibly treatment-related)
132 occurrence or non-occurrence of the event itself.

133

134 More formally, an estimand defines in detail what needs to be estimated to address a specific
135 scientific question of interest. A description of an estimand includes four attributes:

136 A. the population, that is, the patients targeted by the scientific question;

137 B. the variable (or endpoint), to be obtained for each patient, that is required to address
138 the scientific question;

139 C. the specification of how to account for intercurrent events to reflect the scientific
140 question of interest.

141 D. the population-level summary for the variable which provides, as required, a basis for
142 a comparison between treatment conditions

143 Together these attributes describe the estimand, defining the treatment effect of interest.

144

145 In most cases, the target population is reflected by the patients that are eligible to be included
146 in the clinical trial based on the inclusion/exclusion criteria in the protocol. In some cases, a
147 stratum of those patients may be of interest, defined in terms of a potential intercurrent event;
148 for example, the stratum of subjects who would adhere to treatment.

149

150 The variable typically consists of measurements taken (e.g., blood pressure measurement),
151 functions thereof (e.g., change from baseline to one year in HbA1c), or quantities related to
152 clinical outcomes (e.g., time of death, times of hospitalisations, number of relapses). The
153 variable may also incorporate intercurrent events such as discontinuation of treatment, for
154 example when using measurements taken prior to discontinuation (e.g., area under the curve
155 of HbA1c until discontinuation; the number of weeks blood pressure is controlled while on
156 treatment), or composites (e.g., treatment failure defined as non-response or treatment
157 discontinuation).

158

159 It is necessary to specify how to account for potential intercurrent events in a way that
160 reflects the scientific question of interest. Intercurrent events can present in multiple forms
161 and can affect the interpretation of the variable. For example, if a subject dies before a
162 planned measurement of blood pressure, the blood pressure will not be observed. If a subject
163 takes rescue medication in addition to treatment, the blood pressure may be observed, but will
164 reflect the combined effect of the treatment and the rescue medication. If a subject
165 discontinues treatment because of toxicity, the blood pressure may be observed but will
166 reflect the lack of effect of the treatment when it is not taken. The set of intercurrent events
167 for consideration will depend on the specific therapeutic setting and trial objective. Taking
168 use of rescue medication as an example, two different specifications include the combined
169 effect of treatment and any intercurrent event (in this case use of rescue medication) and the
170 effect of the treatment in the, potentially hypothetical, absence of the intercurrent event.

171 Section A.3.2 describes different strategies for addressing intercurrent events in constructing
172 an estimand that is best aligned with the corresponding scientific question of interest.

173

174 The fourth attribute is the population-level summary measure for the variable, e.g. the mean
175 change from baseline to one year in HbA1c, or the proportion of subjects meeting specified
176 criteria for response. In case of treatment comparisons, the summary measure becomes e.g.
177 the difference in mean change from baseline to one year in HbA1c, or the difference or ratio
178 in the proportion of subjects meeting specified criteria, under two different treatment
179 conditions.

180

181 **3.2. Strategies for Addressing Intercurrent Events**

182 The estimand attributes A through D introduced in Section A.3.1 are inter-related and should
183 not be considered independently. The description of an estimand will not be complete
184 without reflecting how potential intercurrent events are reflected in the scientific question of
185 interest. At least five strategies may be considered. The strategies can be used alone or in
186 combination to address multiple different intercurrent events. Together with the other
187 estimand attributes, the choices made on how to address intercurrent events describe the
188 treatment effect that is targeted. Section A.7 provides illustrations of the use of these five
189 strategies for constructing estimands accounting for one or more intercurrent events.

190

191 The relevance of each strategy will depend on the therapeutic and experimental context. In
192 addition it might or might not be possible, in each experimental situation, to derive an
193 estimate for a particular estimand constructed using these strategies that is considered reliable
194 for decision-making. These considerations are addressed in Sections A.3.3, A.3.4, A.4 and
195 A.5. The labels that are presented below are for ease of reference only; an adequate
196 description of the chosen strategy must be used when constructing an estimand.

197

198 **Treatment policy strategy**

199 The occurrence of the intercurrent event is irrelevant: the value for the variable of interest is
200 used regardless of whether or not the intercurrent event occurs.

201

202 For example, when specifying how to account for rescue medication as an intercurrent event,
203 occurrence of the intercurrent event is ignored and the observations on the variable of interest
204 are used. If applied across all types of intercurrent events, this reflects the comparison
205 described in the ICH E9 Glossary (under Intention to Treat Principle) as the effect of a
206 treatment policy.

207

208 In general, this strategy cannot be implemented when values for the variable after the
209 intercurrent event do not exist for all subjects. For example, an estimand based on this
210 strategy cannot be constructed with respect to a variable that cannot be measured due to
211 death.

212

213 **Composite strategy**

214 The occurrence of the intercurrent event is taken to be a component of the variable, i.e. the
215 intercurrent event is integrated with one or more other measures of clinical outcome as the
216 variable of interest.

217

218 There are multiple different approaches that can be considered under this label. The
219 requirement to use a rescue medication may provide meaningful information on the effect of

220 a treatment and hence may be incorporated into a variable, with appropriate summary
221 measure, that describes a meaningful treatment effect. For example, the variable might be
222 defined as a composite of no use of rescue medication and a favourable clinical outcome.
223 Alternatively, for a numerical variable, experiencing an intercurrent event might be ascribed
224 an extreme unfavourable value and a suitable summary measure selected. A different
225 approach would be to employ area-under-the-curve, reflecting the planned duration of follow-
226 up but based on the values for the variable prior to the intercurrent event.
227 Sometimes an event being considered as intercurrent is itself the most meaningful variable
228 that can be measured for quantifying the treatment effect of interest. This can be the case
229 with death: the fact that a subject has died may be much more meaningful than observations
230 before death, and observations after death will not exist. For example, in a trial with a
231 primary focus on myocardial infarction, it may not always be possible to ascertain whether a
232 subject who died had, or would have had, a myocardial infarction, but if the variable is
233 defined to be a composite of death or myocardial infarction, this may be completely
234 ascertained.

235

236 **Hypothetical strategy**

237 A scenario is envisaged in which the intercurrent event would not occur: the value to reflect
238 that scientific question of interest is that which the variable would have taken in the
239 hypothetical scenario defined.

240

241 For example, when rescue medication must be made available for ethical reasons, a treatment
242 effect of interest might concern the outcomes if rescue medication had not been available.
243 Analogously, another active treatment might be administered upon failure and subsequent
244 discontinuation of treatment (including treatment switching where the experimental treatment
245 is given to subjects previously randomised to the control arm), but the treatment effect of
246 interest might concern the outcome if the subsequent active treatment had not been
247 administered. In these examples the non-availability of rescue medication and the absence of
248 the other active treatment reflect different hypothetical conditions.

249

250 Care is required to precisely describe the hypothetical conditions reflecting the scientific
251 question of interest in the context of the specific trial. For example, the hypothetical
252 condition might usefully address both the use of a rescue medication and adherence to
253 treatment as intercurrent events in order for an estimand to be precisely described.

254

255 **Principal stratum strategy**

256 The target population might be taken to be the principal stratum (see Glossary) in which an
257 intercurrent event would not occur. For example, the target population of interest might be
258 taken to be the stratum of patients in which failure to adhere to treatment would not occur. In
259 other words, a principal stratum is a subset of the broader population who would not
260 experience the intercurrent event. The scientific question of interest relates to the treatment
261 effect only within that stratum.

262

263 Effects in principal strata should be clearly distinguished from any type of subgroup or per-
264 protocol analyses where membership is based on the trial data. Principal stratification (see
265 Glossary) is defined by a patient's potential intercurrent events on both treatments: for
266 example, patients who would adhere to either treatment. It is not possible in general to
267 identify these subjects directly, either in advance of the trial since the occurrence of the
268 intercurrent event cannot be predicted, or based on the data from a randomised controlled
269 trial because each patient will be observed on one treatment only. Membership in a principal

270 stratum must then be inferred, usually imperfectly, from covariates. In contrast, estimation of
271 a treatment effect from any analysis where membership is based on intercurrent events on the
272 assigned treatments is liable to confounding because different subjects will experience
273 different intercurrent events on different treatments.

274

275 **While on treatment strategy**

276 Response to treatment prior to the occurrence of the intercurrent event is of interest. If a
277 variable is measured repeatedly, its values up to the time of the intercurrent event may be
278 considered to account for the intercurrent event, rather than the value at the same fixed
279 timepoint for all subjects.

280

281 For example, subjects with a terminal illness may discontinue a purely symptomatic
282 treatment because they die, yet the success of the treatment can be measured based on the
283 effect on symptoms before death. Alternatively, subjects might discontinue treatment, and in
284 some circumstances it will be of interest to assess the risk of an adverse drug reaction during
285 the period of adherence.

286

287 Altogether, five different strategies are considered in this section. It is important to be
288 precise when describing the preferred strategy for handling each intercurrent event. Consider
289 adherence to treatment; it is of utmost importance to distinguish between treatment effects of
290 interest based on (i) the hypothetical scenario of “if all subjects would adhere” from (ii) the
291 stratum of subjects who “would be able to adhere if administered the experimental treatment”
292 and (iii) the effect during adherence.

293

294 **3.3. Construction of Estimands**

295 **3.3.1. General Considerations**

296 As stated above, in order to unambiguously describe the treatment effect of interest, and to
297 promote the relevance of the treatment effect described to subjects and physicians,
298 intercurrent events need to be considered explicitly in the construction of the estimand. The
299 construction of the estimand should address each intercurrent event that may occur in the
300 clinical trial and that will affect the interpretation of the results of the trial. The description of
301 intercurrent events at the planning stage might in theory reflect very specific details of
302 treatment and follow-up, such as a specific time window for observing a variable. Such
303 specific criteria are not expected to affect interpretation of trial results. It may be impractical
304 to foresee every relevant kind of intercurrent event. Trial reporting should then discuss not
305 only the way unforeseen intercurrent events were handled in the analysis but also the effect
306 on what the chosen analysis estimates. Within the construction of an estimand, different
307 strategies (Section A.3.2, Section A.7) might be selected to address different intercurrent
308 events.

309

310 The construction of the estimand(s) in any given clinical trial is a multi-disciplinary
311 undertaking including clinicians, statisticians and other disciplines involved in clinical trial
312 design and conduct. It should be the subject of discussion in a sponsor’s interactions with
313 regulators about the objectives and designs for prospective clinical trials. The construction of
314 an estimand should be consequent to the trial objectives and should inform choices relating to
315 data collection and analytic approaches. Avoiding or over-simplifying this process risks
316 misalignment between trial objectives, trial design, data collection and statistical analysis.

317

318 An iterative process may be required. The construction of an estimand should be justified
319 considering what is of clinical relevance in the particular therapeutic setting, including the
320 disease under study and the goal of treatment, and the particular experimental setting (Section
321 A.3.3.2). In addition, the adequacy of trial design and statistical methods need to be
322 considered to ensure that an estimate which is reliable for inference can be derived. In
323 particular, the crucial advantage of randomisation in clinical trials should be acknowledged
324 and exploited to the extent possible. Some estimands, in particular those that are estimated
325 using the observed data, can be robustly estimated making few assumptions, whereas other
326 estimands require more specific assumptions that may be more difficult to justify and that
327 may be more sensitive to plausible changes in those assumptions (see Section A.5.1). Where
328 significant issues exist to develop an appropriate trial design or to derive a reliable estimate
329 for a particular estimand, an alternative estimand, trial design and analytic approach would
330 need to be considered.
331

332 *3.3.2. Considerations of Therapeutic and Experimental Context*

333 As indicated above, aspects of the disease setting and the aim of treatment will influence the
334 construction of the estimand. In terms of therapeutic context this might include, respectively,
335 the availability of alternative treatment options and the possibility to monitor individual
336 response to treatment, and whether the treatment is aimed at providing symptom control,
337 modifying the course of the disease or prevention of disease. For example, the goal of a
338 treatment may be control of clinical signs or symptoms in a disease area where multiple
339 alternative treatments exist, with the possibility to tailor the choice of treatment for a patient
340 based on observed response. The use of an alternative treatment (a rescue medication, a
341 medication prohibited by the protocol or a subsequent line of therapy) will likely need to be
342 considered as an intercurrent event. The specification of how to account for intercurrent
343 events to reflect the scientific question of interest might be based on understanding the
344 treatment effect if the alternative treatment was not available, or in the stratum of subjects
345 who can adhere to treatment without needing an alternative. In some circumstances, answers
346 to these questions might be more relevant than e.g. the quantification of the effects of a
347 treatment policy that does not distinguish whether or not a patient has taken an alternative
348 treatment. Such considerations might be of even greater relevance for the intercurrent event
349 of subjects assigned to the control arm switching to treatment. An estimand might be
350 constructed using one of these strategies, providing it is agreed that a robust estimate can be
351 obtained. In other situations, it might be necessary to understand the treatment effect in the
352 context of a treatment policy that exists in clinical practice. For example, the aim of a
353 treatment may be to prevent or delay an adverse clinical outcome (e.g. death). If the
354 treatment is proposed for use in treatment-naïve subjects as part of a treatment policy where
355 subsequent lines of treatment are established, the effect of the treatment policy could be of
356 greater interest. When constructing estimands based on the treatment policy strategy,
357 inference can be complemented by defining an additional estimand and analysis pertaining to
358 the intercurrent event itself; for example, contrasting both the treatment effect on a symptom
359 score and the amount of rescue medication used under each treatment condition.
360

361 Estimands based on the treatment policy strategy might also be more generally acceptable to
362 support regulatory decision making, specifically in settings where estimands based on
363 alternative strategies might be considered of greater clinical interest, but main and sensitivity
364 estimators cannot be identified that are agreed to support a reliable estimate or robust
365 inference. An estimand based on the treatment policy strategy might offer the possibility to
366 obtain a reliable estimate of a treatment effect that is still relevant. In this situation, it is

367 recommended to retain those estimands that are considered to be of greater clinical relevance
368 and to present the resulting estimates along with a discussion of the limitations, in terms of
369 trial design or statistical analysis, for that specific approach.
370

371 One example for a composite strategy is to replace a continuous variable with a binary
372 variable, in which patients are considered as responders versus non-responders based on a
373 predefined threshold of change in score in the absence of the intercurrent event. This
374 dichotomisation of continuous scores would thus result in a change of the estimand. The
375 clinical relevance and interpretation of the estimand will depend on whether clinically
376 interpretable responder criteria and an appropriate population-level summary (e.g., difference
377 in proportions, odds ratio) are available.
378

379 Using the hypothetical strategy, some conditions are likely to be more acceptable for
380 regulatory decision making than others. The hypothetical conditions described must
381 therefore be justified for the quantification of an interpretable treatment effect that is relevant
382 to the use of the medicine in clinical practice. As noted, the question of what the values for
383 the variable of interest would have been if rescue medication had not been available may be
384 an important one, targeting an effect of the treatment under certain conditions rather than a
385 particular treatment policy that includes the use of the rescue medication. In contrast, the
386 question of what the values for the variable of interest would have been under the
387 hypothetical condition that subjects who discontinued treatment because of adverse drug
388 reaction had in fact continued with treatment, might not be justified as being of scientific or
389 regulatory interest. A scientific question of interest based on the effect if all subjects had
390 adhered to treatment is not well-defined without a thorough discussion of the hypothetical
391 conditions under which it is supposed that they would have adhered. Furthermore, the
392 inability to tolerate a treatment in a trial as well as in clinical practice may constitute, in itself,
393 evidence of an inability to achieve a favourable outcome. If the intercurrent event for which
394 a strategy needs to be selected depends not only on, for example, lack of adherence, but also
395 on the reason for the lack of adherence (e.g. due to toxicity), these have to be defined and
396 recorded accurately in the clinical trial.
397

398 The experimental situation should also be considered. If patient management (e.g. dose
399 adjustment for intolerance, rescue treatment for inadequate response) under a clinical trial
400 protocol is justified to be different to that which is anticipated in clinical practice, this might
401 be reflected in the construction of the estimand. In particular, the choice of the control arm
402 might influence the manner in which rescue or other concomitant medications are permitted
403 in the trial.
404

405 Use of a treatment other than the one assigned will commonly be considered as an
406 intercurrent event. The alternative treatments can be diverse, including rescue medications,
407 medications that are prohibited by the protocol or use of a subsequent line of therapy.
408 Moreover, even rescue medications might be understood in different ways; including use
409 instead of, or in addition to, a chronic treatment on which the subject is experiencing
410 inadequate effect, as an alternative where a subject is not tolerating their assigned treatment,
411 or as a short-term acute treatment to manage a temporary flare in disease symptoms. These
412 examples illustrate the importance of considering the handling of the specific intercurrent
413 event in the context of the particular experimental situation.
414

415 The choice of estimands for studies with objectives to demonstrate non-inferiority or
416 equivalence requires careful reflection. In Section 3.3.2 it is stated that such trials are not

417 conservative in nature and the importance of minimising the number of protocol violations
418 and deviations, non-adherence and withdrawals is indicated. In Section 5.2.1, it is described
419 that the result of the full analysis set (FAS) is generally not conservative and that its role in
420 such trials should be considered very seriously. Estimands that are constructed with one or
421 more intercurrent events accounted for using the treatment policy strategy present similar
422 issues for non-inferiority and equivalence trials as those related to the FAS. Responses in
423 both treatment groups will appear more similar following discontinuation of randomised
424 treatment or use of another medication for reasons that are unrelated to the similarity of the
425 initially randomised treatments. Estimands could be constructed to directly address those
426 intercurrent events which can lead to the attenuation of differences between treatment arms
427 (e.g. use of rescue medications and violations from the target population). In this situation,
428 the estimand might target a measure of treatment effect with high sensitivity to detect
429 differences between treatments, if they exist.
430

431 **4. Impact on Trial Design and Conduct**

432 The design of a trial needs to be aligned to the choice of the estimand or estimands that
433 reflect the primary trial objectives and which will form the basis to establish whether those
434 objectives have been met. Specifically, clear definitions for the estimands on which
435 quantification of treatments effects will be based should inform the choices that are made in
436 relation to trial design. If interest lies, for example, in understanding the effect of treatment
437 regardless of whether a particular intercurrent event occurs, a trial in which the variable is
438 collected for all subjects regardless of that event is appropriate. Alternatively, if the
439 estimands that are required to support regulatory decision making do not require the
440 collection of the variable after an intercurrent event, then the benefits of collecting such data
441 for other estimands should be weighed against any complications and potential drawbacks of
442 the collection.
443

444 Efforts should be made to collect all data that are relevant to support a statistical analysis
445 aligned to the estimands of interest including important additional estimands. The occurrence
446 of intercurrent events such as non-adherence, discontinuation of treatment, treatment
447 switching, or use of rescue medication, does not imply that the variable cannot be measured
448 thereafter, unlike for terminal events such as death. Not collecting any data needed to assess
449 an estimand results in a missing data problem for subsequent statistical inference. The
450 validity of statistical analyses may rest upon untestable assumptions and, depending on the
451 proportion of missing data; this may undermine the robustness of the results (Section A.5). A
452 prospective plan to collect informative reasons for why data intended for collection are
453 missing may help to distinguish intercurrent events of interest from residual missing data and
454 thus potentially improve the primary analysis. This may also lead to a more appropriate
455 choice of sensitivity analysis. For example, perhaps a generic “loss to follow up” should
456 correctly be recorded as “treatment discontinuation due to lack of efficacy”. Where that has
457 been defined as an intercurrent event of interest, this can be reflected through the chosen
458 strategy to account for that intercurrent event and not as a missing data problem. Measures
459 taken to retain subjects can be implemented, but care should be taken to retain the external
460 validity of the trial to clinical practice. For example, selection of the trial population or use
461 of titration schemes or concomitant medications to mitigate the impact of toxicity might not
462 be suitable if those same measures would not be implemented in clinical practice.
463

464 Certain estimands may necessitate, or may benefit from, non-standard trial designs such as
465 run-in or enrichment designs, randomised withdrawal designs, or titration designs. Such

466 alternative designs, however, may require special consideration regarding their
467 implementation and subsequent statistical inference. For example, it might be of interest to
468 try to identify the stratum of subjects who can tolerate a treatment, using a run-in period, in
469 advance of randomising those subjects between treatment and control. Dialogue between
470 regulators and sponsors would need to consider whether the proposed run-in period is
471 appropriate to identify the target population, and whether the choices made for the subsequent
472 trial design (e.g. washout period, randomisation) supports the estimation of the target
473 treatment effect and associated inference. These considerations might limit the use of these
474 trial designs, and use of that particular strategy, in practice.

475
476 A precise description of the treatment effects of interest, through specification of strategies to
477 handle intercurrent events, should inform sample size calculations. Where all subjects
478 contribute information to the analysis, and where the impact of intercurrent events and their
479 handling is reflected in the effect size that is targeted and the expected variance, it is not
480 usually necessary to inflate the calculated sample size by the expected proportion of subject
481 withdrawals.

482
483 Section 7.2 addresses issues related to summarising data across clinical trials. The need to
484 have consistent definitions for the variables of interest is highlighted and this can be extended
485 to the construction of estimands. Hence in situations when pooling data from across a
486 clinical trial programme is envisaged at the planning stage, a suitable estimand should be
487 constructed, included in the trial protocols, and reflected in the choices made for the designs
488 of the contributing trials. Similar considerations apply to the design of a meta-analysis or the
489 use of external control groups for the interpretation of single-arm trials. A naïve comparison
490 between data sources, or integration of data from multiple trials without consideration and
491 specification of the estimand that is addressed in each data presentation or statistical analysis,
492 could be misleading and can be considered as a source of bias.

493
494 More generally, a trial is likely to have multiple objectives translated into multiple estimands.
495 A trial design that is suitable for one estimand might not be suitable for other estimands of
496 potential importance. Trials with multiple objectives and endpoints might give rise to
497 concerns over multiple testing and in principle these concerns apply equally to the inclusion
498 of multiple estimands. The same approaches employed to address those concerns, in
499 particular the nomination of one or more as primary and others as secondary, can equally be
500 applied to estimands.

501

502 **5. Impact on Trial Analysis**

503 **5.1. Main Estimation**

504 An estimand for the effect of treatment relative to a control should reflect the outcomes in a
505 group of subjects on the treatment to those in a similar group of subjects on the control, so
506 that the effect of treatment can be isolated from any differences between the groups of
507 subjects on which the comparison is based. For a given estimand an aligned analytic
508 approach, or estimator, should be implemented that is able to provide an estimate on which
509 reliable interpretation can be based. An important consideration for whether a robust
510 estimate will be available is the extent of assumptions that need to be made. Assumptions
511 should be stated explicitly together with the main and sensitivity estimators. Assumptions
512 should be justifiable and implausible assumptions should be avoided. The robustness of the

513 results to the underlying assumptions should be assessed through sensitivity analysis aligned
514 to the estimand (Section A.5.2).

515

516 In particular, if there is complete follow-up of subjects regardless of whether or not the
517 intercurrent event occurs, an estimand based on the treatment policy strategy can be estimated
518 with only minimal assumptions. Estimation for an estimand employing this strategy will
519 require stronger and untestable assumptions if measurements are not collected following
520 intercurrent events. Using a composite strategy it may be possible to perform an analysis
521 without need for imputation or modelling of response after an intercurrent event, and the
522 associated assumptions even when the original variable was not completely ascertained. In
523 contrast, the estimation of estimands constructed using a strategy that requires a hypothetical
524 scenario to address an intercurrent event entails careful specification of the hypothetical
525 conditions and will necessarily rely on modelling assumptions that are untestable and need to
526 be investigated through sensitivity analyses. In a randomised trial, estimation of a treatment
527 effect within a principal stratum of the population will be confounded unless the subjects
528 within that stratum can be identified before randomisation. Otherwise, estimation will rely
529 on assumptions, in particular that all relevant confounders have been measured and accounted
530 for. For example, for the stratum of subjects who would be able to adhere to the treatment it
531 is inappropriate to simply compare the observed adherers on the treatment to adherers on
532 control. These will be systematically different subjects, confounding estimation of the
533 treatment effect. In this case it is essential to account for all important confounders, rather
534 than a small, preconceived set of covariates, though it is difficult to provide assurance against
535 misspecification of the model. For the labelled while-on-treatment strategy, estimation of a
536 treatment effect will require stronger assumptions when the occurrence and timing of an
537 intercurrent event is related to treatment.

538

539 Even after defining estimands that address intercurrent events in an appropriate manner, and
540 making efforts to collect the data required for estimation (Section A.4), some data may still
541 be missing. This missing data is distinguished from systematic failure or avoidance in
542 collecting information that are required for estimation. For example, if an estimand based on
543 the treatment policy strategy is constructed, all efforts should be made to retain subjects in the
544 trial and adhere to the schedule of assessments even after discontinuation of assigned therapy.
545 Where those efforts are not successful it becomes necessary to make assumptions about the
546 missing observations, either to predict or impute individual observations or to justify
547 statistical methods based on observed data only. Handling of missing data should be based
548 on plausible assumptions and, where possible, guided by the strategies employed in the
549 description of the estimand. Predictions for a given subject may be based on observed data
550 from that subject (covariates and post-baseline values) and from other similar subjects.
551 Criteria to identify similar subjects might include whether or not the intercurrent event has
552 been assessed (e.g., for subjects who discontinue treatment without further data collected, a
553 prediction model may use data from other subjects who discontinued treatment but for whom
554 data collection has continued rather than from subjects who remained on treatment).
555 Reasonable deviations from the assumptions of these techniques are an important aspect of
556 sensitivity analysis.

557 **5.2. Sensitivity Analysis**

558 **5.2.1. Role of Sensitivity Analysis**

559 Inferences based on a particular estimand should be robust to limitations in the data and
560 deviations from the assumptions used in the statistical model for the main estimator. This
561 robustness is evaluated through a sensitivity analysis.

562
563 The statistical assumptions that underpin the main estimator should be documented. One or
564 more analyses, focused on the same estimand, should then be pre-specified to investigate
565 these assumptions with the objective of verifying that the estimate derived from the main
566 estimator is robust to departures from its assumptions. Distinct from this sensitivity analysis,
567 each other analysis that is planned, presented or requested in order to more fully investigate
568 and understand the trial data can be termed supplementary analysis (see Glossary). Each
569 supplementary analysis may refer to a different estimand, or a different estimator to the same
570 estimand. Where the primary estimand(s) of interest is agreed between sponsor and
571 regulator, and the main estimator is pre-specified unambiguously, supplementary analyses
572 should generally be given lower priority than a sensitivity analysis.
573

574 **5.2.2. Choice of Sensitivity Analysis**

575 When planning and conducting a sensitivity analysis, it is recommended not to alter many
576 aspects of the main analysis simultaneously, or else it could be challenging to identify which
577 assumptions, if any, are responsible for any potential differences seen. A more transparent
578 and useful approach is to investigate the impact of changing only one assumption at a time.
579 In addition, a distinction between testable and untestable assumptions may be useful when
580 assessing the interpretation and relevance of different analyses.
581

582 Missing data require particular attention in a sensitivity analysis because the assumptions
583 underlying any method may be hard to justify fully and may be impossible to test. Missing
584 data must be defined and considered in respect of a particular estimand. For example, data
585 that were intended to be collected after discontinuation of trial medication to inform an
586 estimand based on the treatment policy strategy are missing if uncollected; however, the same
587 data points might be irrelevant for another strategy, and thus, for the purpose of that second
588 estimand, are not missing if uncollected. Fortunately, relevant types of deviation from
589 assumptions can often be characterized simply. For example, in an analysis of means for
590 continuous outcomes, the original analysis may be biased to the extent that missing and non-
591 missing data for each treatment group differ in their means, and especially when these
592 differences themselves differ across treatment groups. A plausible range of assumed values
593 for these differences should be studied and the robustness of the conclusions assessed. In
594 significance testing, for example, values of the differences for which the treatment effect is or
595 is not statistically significant at a pre-specified level can be plotted in the context of a tipping
596 point analysis. A similar approach can be considered to ascertain values of the differences
597 for which the treatment effect does or does not retain a specific degree of clinical relevance.
598 Similar techniques can be applied to other data structures. For example, proportions of
599 successes or hazards for time-to-event data can be assumed to be different between missing
600 and non-missing data, differentially across treatment groups.
601

602 **5.3. Supplementary Analysis**

603 Interpretation of trial results should focus on the main estimator for each agreed estimand if
604 the corresponding estimate is verified to be robust through the sensitivity analysis.
605

606 Supplementary analyses targeting different estimands play a secondary role for interpretation
607 of trial results, though can provide additional insights. For example, an analysis based on the
608 proportion of responders might be helpful for interpretation of a treatment effect that is
609 quantified by difference in mean changes on a continuous scale. Alternatively, different

610 definitions for a responder might be examined to investigate whether the result is robust to
611 that definition. The need for, and utility of, supplementary analyses should be determined for
612 each trial.

613
614 Section 5.2.3 indicates that it is usually appropriate to plan for analyses based on both the
615 FAS and the per-protocol set (PPS) so that differences between them can be the subject of
616 explicit discussion and interpretation. Consistent results from analyses based on the FAS and
617 the PPS is indicated as increasing confidence in the trial results. Also in Section 5.2.2 it is
618 described that results based on a PPS might be subject to severe bias. In respect of the
619 framework presented in this addendum, an analysis based on the subset of subjects who
620 adhere to the clinical trial protocol having been assigned to a particular treatment group can
621 be conducted, but does not in itself unambiguously define a treatment effect of interest. As
622 noted above, analysis of the per-protocol data set does not achieve the goal of estimating the
623 effect in adherent subjects because it does not compare similar subjects on different
624 treatments. The role of such an analysis is therefore limited to investigating whether the
625 extent of protocol violations and deviations compromises confidence in the trial results.
626 Some protocol violations and deviations might be addressed as intercurrent events. Where a
627 majority of intercurrent events are handled through the construction of the estimands, the
628 number of remaining protocol violations and deviations will be low and analysis of the PPS
629 might not add additional insights.

630

631 **6. Documenting Estimands and Sensitivity Analysis**

632 Estimands should be defined and explicitly specified in the clinical trial protocol. Having
633 specified those types of intercurrent events that can be foreseen and that would affect the
634 interpretation of the results of the trial, a trial protocol should pre-specify a primary estimand
635 that corresponds to the primary trial objective. Furthermore, the protocol and the analysis
636 plan should pre-specify the main estimator that is aligned with the primary estimand and
637 leads to the primary analysis, together with a suitable sensitivity analysis to explore the
638 robustness under deviations from its assumptions. Estimands for secondary trial objectives
639 (e.g. related to secondary variables) that are likely to support regulatory decisions should be
640 described properly, each with a corresponding main estimator and a suitable sensitivity
641 analysis. Additional trial objectives may be considered for exploratory purposes, leading to
642 additional estimands.

643

644 While it is to the benefit of the sponsor to have clarity on what is being estimated, it is not a
645 regulatory requirement to document in detail an estimand for each exploratory question,
646 especially if these are minor variations on primary or secondary estimands in terms of
647 handling intercurrent events. However, where different scientific questions of interest call for
648 materially different estimands, it is recommended that these should be fully documented.

649

650 The choice of the primary estimand will usually be the main determinant for aspects of trial
651 design and conduct. Following usual practices, these aspects should be well documented in
652 the trial protocol. If additional estimands are of key interest, these considerations may be
653 extended to support these as needed and should be documented as well. Beyond these
654 aspects, the conventional considerations for trial design, conduct and analysis remain the
655 same. For example, where there is more than one estimand giving rise to potential issues of
656 multiple testing, the usual considerations for controlling type I error apply and should be
657 described accordingly (Section A.4).

658 Results from the main, sensitivity and supplementary analyses should be reported
659 systematically in the clinical trial report, specifying whether each analysis was pre-specified,
660 introduced while the trial was still blinded, or performed post hoc. Addressing intercurrent
661 events that were not foreseen at the design stage, or identified during the conduct of the trial
662 should then discuss not only the way intercurrent events were handled in the analysis but the
663 effect on what the chosen analysis estimates and the interpretation of the trial results.

664 7. A Generic Example

665 In the following, a generic example for a continuous variable is used to illustrate the
666 framework proposed in this addendum. It should not be construed as a regulatory
667 recommendation and should be adapted to the needs of a given clinical trial setting (in
668 particular, but not limited to, when using binary or time to event variables).

669
670 A new investigational treatment (Drug X) is considered for subjects with a specific chronic,
671 non-life-threatening disease. Response to treatment is monitored monthly using a continuous
672 measurement. The full effect of Drug X is expected to be seen at four to six months after
673 treatment start. The main scientific question concerns the comparison of Drug X to placebo
674 at month 6, and is best addressed by a randomised clinical trial. Use of placebo in the clinical
675 trial is considered ethical but only if provision is made for subjects to discontinue their
676 treatment and switch to rescue medication due to lack of efficacy. Switch to rescue
677 medication is an intercurrent event, after which it is still possible to collect the variable
678 measurements. This is also the case after other intercurrent events such as discontinuation of
679 treatment due to an adverse event, but not for intercurrent events such as death (considered
680 very unlikely in this setting).

681
682 In the unrealistic case where no intercurrent events are expected to occur, the definition of an
683 appropriate estimand is uncontroversial in terms of the following four attributes:

- 684 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
685 targeted patient population for approval;
- 686 B. Variable: change from baseline to month six in the designated measurement;
- 687 C. Intercurrent event: no intercurrent events to be taken into account;
- 688 D. Population-level summary: difference in variable means between treatment
689 conditions.

690
691 The estimand is then the difference in means between treatment conditions in the change
692 from baseline to month six in the designated measurement in the targeted patient population.

693
694 A design that targets this estimand is a randomised parallel group design where all
695 measurements are collected throughout the trial. Failure to do so would result in missing
696 data. As long as all measurements are collected, an analysis of variance model with
697 treatment group as a factor is one example for a statistical analysis for this estimand. In case
698 of missing measurements, data need to be predicted based on plausible assumptions that
699 account for the uncertainty due to missing data. For example, missing data may be imputed
700 based on similar subjects who remained in the trial. Similarity may be established based on
701 the same baseline covariates, the same randomised treatment arm, the same measurement
702 history and information on the intercurrent event. Sensitivity analyses should be pre-
703 specified in the trial protocol to assess, for example, the assumptions of the imputation
704 method. Inference can be complemented by including additional supplementary analyses,

705 possibly targeting different estimands, such as contrasting the proportion and timing of rescue
706 switchers between the treatment groups.

707
708 Attribute C is labelled as “Intercurrent event” for brevity, referring to the specification of
709 how to account for potential intercurrent events to reflect the scientific question of interest.
710

711 **7.1 One Intercurrent Event**

712 In practice, intercurrent events are expected to occur. For ease of exposition, consider
713 initially the case that only the intercurrent event “switch to rescue medication due to lack of
714 efficacy” is expected to occur. In the following, alternative estimands corresponding to
715 different scientific questions are described, together with high level considerations on trial
716 design, conduct and analysis.

717

718 **Treatment-policy strategy**

719 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
720 targeted patient population for approval;

721 B. Variable: change from baseline to month six in the designated measurement;

722 C. Intercurrent event: regardless of whether or not switching to rescue medication had
723 occurred;

724 D. Population-level summary: difference in variable means between treatment
725 conditions.

726

727 In this specific example the estimand described by the treatment-policy strategy is the effect
728 of “Drug X + rescue medication as needed” versus “placebo + rescue medication as needed”
729 on the variable measurement. Thus, dependent on the proportion of rescue medication
730 switchers in both treatment arms, this estimand captures a mixture of the effects of treatment
731 and rescue medication. Also, this estimand does not capture that switching to rescue
732 medication is driven by the unfavourable event of “lack of efficacy”.

733

734 The estimand is then the difference in means between treatment conditions in the change
735 from baseline to month six in the designated measurement in the targeted patient population,
736 regardless of whether or not switching to rescue medication had occurred.

737

738 A similar sentence can be constructed for each of the examples below, also integrating the
739 specification for how the intercurrent events are handled.

740

741 A design that targets this estimand is a randomised parallel group design where all
742 measurements regardless of switching to rescue medication are collected throughout the trial.

743

744 As long as all measurements are collected, an analysis of variance model with treatment
745 group as a factor is one example for a statistical analysis for this estimand. In case of missing
746 measurements, data need to be predicted based on plausible assumptions that account for the
747 uncertainty due to missing data. For example, missing data may be imputed based on similar
748 subjects who remained in the trial. Similarity may be established based on the same baseline
749 covariates, the same randomised treatment arm, the same measurement history and
750 information on the intercurrent event. Sensitivity analyses should be pre-specified in the trial
751 protocol to assess, for example, the assumptions of the imputation method. Inference can be
752 complemented by including additional supplementary analyses, possibly targeting different
753 estimands, such as contrasting the proportion and timing of rescue switchers between the

754 treatment groups. Another estimand of interest could be constructed to address a scientific
755 question on the use of rescue medication.

756

757 **Composite strategy**

758 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
759 targeted patient population for approval;

760 B. Variable: binary response variable indicating a successful response at month six if the
761 change from baseline to month six in the designated measurement is above a pre-
762 specified threshold, and no switching to rescue medication occurred;

763 C. Intercurrent event: the intercurrent event is captured through the variable definition;

764 D. Population-level summary: difference in response proportions between treatment
765 conditions.

766

767 The estimand described by the composite strategy no longer assesses the treatment effect
768 only in terms of the variable measurements at month six. Rather, the treatment effect is
769 established based on a composite variable which combines a clinically meaningful
770 dichotomous change in the variable measurement with the intercurrent event of “switching to
771 rescue”. As switching to rescue medication is based on lack of efficacy, this estimand
772 acknowledges that intake of rescue medication is an unfavourable outcome.

773

774 A design that targets this estimand is a randomised parallel group design. There would be no
775 need to collect measurements after switching to rescue medication, unless there is interest in
776 alternative trial objectives that would require such data (e.g. to collect safety information
777 even after the intercurrent event). In this example, data that could have been collected after
778 the use of rescue medication is not regarded as missing as they are not of interest for
779 estimating the targeted estimand.

780

781 As long as all measurements to establish the response status are collected, a logistic
782 regression is one example for a statistical analysis for this estimand. In case of missing data,
783 i.e. prior to the assessment point without an intercurrent event having occurred, the response
784 status needs to be imputed based on plausible assumptions that account for the uncertainty
785 due to missing data. For example, missing data may be imputed based on similar subjects
786 who remained in the trial. Similarity may be established based on the same baseline
787 covariates, the same randomised treatment and the same measurement history. Sensitivity
788 analyses should be pre-specified in the trial protocol to assess, for example, the assumptions
789 of the imputation method. Inference can be complemented by including additional
790 supplementary analyses targeting the separate components of this composite estimand, such
791 as changing the threshold in the variable definition, leading to a different estimand.

792

793 **Hypothetical strategy**

794 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
795 targeted patient population for approval;

796 B. Variable: change from baseline to month six in the designated measurement;

797 C. Intercurrent event: had rescue medication not been made available to subjects prior to
798 month six;

799 D. Population-level summary: difference in variable means between treatment
800 conditions.

801

802 The estimand described by the hypothetical strategy addresses the treatment effect in an
803 alternative, hypothetical setting where rescue medication was not available to subjects.

804 Conducting a clinical trial to target this scientific question directly may not be ethically
805 justifiable.

806
807 A design that targets the hypothetical estimand is a randomised parallel group design. There
808 would be no need to collect measurements after switching to rescue medication, unless there
809 is interest in alternative trial objectives that would require such data (e.g. to collect safety
810 information even after the intercurrent event). In this example, data that could have been
811 collected after the use of rescue medication is not regarded as missing as they are not of
812 interest for estimating the targeted estimand.

813
814 A statistical analysis for this estimand will rest on assumptions about the measurements that
815 would have been observed under the hypothetical setting where rescue medication was not
816 available to subjects. Generally, the assumptions needed for such predictions cannot be
817 verified based on the observed data so that a sensitivity analysis will be necessary to assess
818 the robustness of conclusions. A discussion on the plausibility of the assumptions will be
819 warranted to give sufficient credibility to these assumptions, and as a consequence the
820 estimation of the treatment effect. Inference can be complemented by including additional
821 supplementary analyses, possibly targeting different estimands, such as contrasting the
822 proportion and timing of rescue switchers between the treatment groups.

823
824 **Principal stratum strategy**

- 825 A. Population: defined through subjects who would not require rescue medication over a
826 period of six months regardless of treatment assignment, within the targeted
827 population defined by inclusion/exclusion criteria;
828 B. Variable: change from baseline to month six in the designated measurement;
829 C. Intercurrent event: the intercurrent event is captured through the population definition;
830 D. Population-level summary: difference in variable means between treatment
831 conditions.

832
833 The estimand described by the principal stratum strategy assesses the effect of the initially
834 randomised treatments in the stratum of the population who would not require rescue
835 medication over a period of six months regardless of which treatment arm they were
836 randomised to.

837
838 One complication with this estimand is that, in practice, it is difficult to identify the members
839 of this population in advance. Thus, in practice one may have to employ non-standard
840 designs to target patients that would not require rescue medication over a period of six
841 months, such as enrichment designs as well as run-in and randomised withdrawal designs.

842
843 A statistical analysis for this estimand is straightforward as long as only subjects who would
844 not require rescue medication over a period of six months had been randomised, and they
845 were followed for the entire trial duration. As noted above, however, it is generally difficult
846 to identify the members of this population in advance. If the targeted population cannot be
847 identified, then a suitable analysis cannot be achieved by restricting the analysis to those
848 subjects who did not switch to rescue medication: this could exclude systematically different
849 subjects on the different assigned treatments, so that the treatment effect would be
850 confounded with patient characteristics that affect the subjects' propensity to switch to rescue
851 medication. An appropriate analysis needs to account for this confounding. In addition, an
852 assessment of the robustness of conclusions to the assumptions made is necessary using
853 appropriate sensitivity analyses. Inference can be complemented by including additional

854 supplementary analyses, possibly targeting different estimands, such as contrasting the
855 proportion and timing of rescue switchers between the treatment conditions.

856

857 **While on treatment strategy**

858 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
859 targeted patient population for approval;

860 B. Variable: average of the designated measurements while on randomised treatment;

861 C. Intercurrent event: the intercurrent event is captured through the variable definition;

862 D. Population-level summary: difference in variable means between treatment
863 conditions.

864

865 This estimand assesses the average treatment effect on the variable measurement. The
866 variable chosen here averages the outcomes while being on treatment, i.e. before switch to
867 rescue medication.

868

869 A design that targets this estimand is a randomised parallel group design. There would be no
870 need to collect measurements after switching to rescue medication, unless there is interest in
871 alternative trial objectives that would require such data (e.g. an alternative estimand that
872 requires those data, or to collect safety information even after the intercurrent event). In this
873 example, data that could have been collected after the use of rescue medication are not
874 regarded as missing as they are not of interest for estimating the targeted estimand.

875

876 As long as all measurements while on the randomised treatments are collected, an analysis of
877 variance model with treatment group as a factor is an appropriate statistical analysis for this
878 estimand. In case of intermittent missing measurements, data need to be interpolated based
879 on plausible assumptions that account for the uncertainty due to missing data. Sensitivity
880 analyses should be pre-specified in the trial protocol to assess, for example, the assumptions
881 of the interpolation method. Inference can be complemented by including additional
882 supplementary analyses, possibly targeting different estimands, such as considering
883 alternative choices for the variable definition by focussing on the last measurement while
884 being on treatment, leading to different estimands.

885

886 **7.2. Two Intercurrent Events**

887 The generic example is now extended to situations where two types of intercurrent events
888 may occur, namely “switch to rescue medication” and “discontinuation of treatment due to an
889 adverse event”. The definition of a clinically meaningful estimand needs to encompass all
890 intercurrent events that are likely to occur and are clinically relevant in a given clinical trial
891 setting, to the extent that the description of the treatment effect being targeted cannot be fully
892 understood without inclusion of the intercurrent event in the estimand. The same holds for
893 choices made about the design, conduct and statistical analysis. Considering the five
894 strategies discussed above, all possible combinations of strategies for two types of
895 intercurrent events can be considered, although not all combinations will be clinically
896 relevant. For ease of exposition, only two different estimand strategies are described in the
897 following, together with high level considerations on trial design, conduct and analysis.

898

899 **Treatment-policy strategy to account for both intercurrent events**

900 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
901 targeted patient population for approval;

902 B. Variable: change from baseline to month six in the designated measurement;

- 903 C. Intercurrent events: regardless of switching to rescue medication and regardless of
904 treatment discontinuation due to an adverse event;
905 D. Population-level summary: difference in variable means between treatment
906 conditions.

907
908 This estimand targets the treatment-policy effect of treatment initiation on the variable
909 measurement. This estimand accounts neither for rescue medication initiation nor for
910 treatment discontinuation due to an adverse event. In particular, it does not capture that
911 switching to rescue medication and adverse events are unfavourable outcomes.

912
913 A design that targets this estimand is a randomised parallel group design where all
914 measurements regardless of switching to rescue medication and treatment discontinuation due
915 to adverse events are collected throughout the trial.

916
917 As long as all measurements are collected, an analysis of variance model with treatment
918 group as a factor is an appropriate statistical analysis for this estimand. In case of missing
919 measurements, data need to be predicted based on plausible assumptions that account for the
920 uncertainty due to missing data. For example, missing data may be imputed based on similar
921 subjects who remained in the trial. Similarity may be established based on the same baseline
922 covariates, the same randomised treatment arm, the same measurement history and
923 information on the intercurrent events. Sensitivity analyses should be pre-specified in the
924 trial protocol to assess, for example, the assumptions of the imputation method. Inference
925 can be complemented by including additional supplementary analyses, possibly targeting
926 different estimands, such as contrasting the proportion and timing of rescue switchers and
927 treatment discontinuations due to adverse events between the treatment groups.

928
929 **Combination of Hypothetical strategy and Treatment-policy strategy to account for the**
930 **two intercurrent events**

- 931 A. Population: defined through appropriate inclusion/exclusion criteria to reflect the
932 targeted patient population for approval;
933 B. Variable: change from baseline to month six in the designated measurement;
934 C. Intercurrent events: had rescue medication not been made available to subjects prior
935 to month six and regardless of study treatment discontinuation due to an adverse
936 event;
937 D. Population-level summary: difference in variable means between treatment
938 conditions.

939
940 This estimand combines two different strategies to account for the two types of intercurrent
941 events. It employs a hypothetical strategy to address switching to rescue medication and a
942 treatment-policy strategy to address treatment discontinuation due to an adverse event. Such
943 an estimand may be of interest and easily interpretable in settings where the pharmacological
944 effect is targeted but withholding rescue medication is not ethical and where subjects remain
945 untreated after treatment discontinuation due to an adverse event.

946
947 A design that targets this estimand is a randomised parallel group design where all
948 measurements regardless of treatment discontinuation due to an adverse event are collected
949 throughout the trial. There would be no need to collect measurements after switching to
950 rescue medication, unless there is interest in alternative trial objectives that would require
951 such data. In this example, data that could have been collected after the use of rescue
952 medication are not regarded as missing.

953 A statistical analysis for this estimand needs to account for both intercurrent events:

- 954 • Switching to rescue medication: Interest lies in the effect had rescue medication not
955 been made available to subjects prior to month six. As measurements under this
956 scenario cannot be directly observed, assumptions about the measurements that
957 would have been observed under this hypothetical setting need to be made.
- 958 • Study treatment discontinuation due to an adverse event: Interest lies in the effect
959 regardless of this intercurrent event. Thus, all measurements regardless of this
960 intercurrent event need to be included in the analysis. In case of missing
961 measurements, data need to be predicted based on plausible assumptions while
962 accounting for the added uncertainty due to missing data. For example, missing data
963 may be imputed based on similar subjects who remained in the trial. Similarity may
964 be established based on the same baseline covariates, the same randomised treatment
965 arm, the same measurement history and information on the intercurrent event, e.g.
966 timing.

967
968 Once the individual predictions are made in line with the observed intercurrent events and the
969 estimand of interest, a statistical analysis using, for example, an analysis of variance model
970 based on all randomised subjects is appropriate. In case of missing measurements, data need
971 to be predicted based on plausible assumptions that account for the uncertainty due to missing
972 data. For example, missing data may be imputed based on similar subjects who remained in
973 the trial. Similarity may be established based on the same baseline covariates, the same
974 randomised treatment arm, the same measurement history and information on the intercurrent
975 events. Sensitivity analyses should be pre-specified in the trial protocol to assess, for
976 example, the assumptions of the imputation method. Inference can be complemented by
977 including additional supplementary analyses, possibly targeting different estimands, such as
978 contrasting the proportion and timing of rescue switchers and treatment discontinuations due
979 to adverse events between the treatment groups.

980 **Glossary**

981 **Estimand:**

982 Is the target of estimation to address the scientific question of interest posed by the trial
983 objective. Attributes of an estimand include the population of interest, the variable (or
984 endpoint) of interest, the specification of how intercurrent events are reflected in the scientific
985 question of interest, and the population-level summary for the variable.

986

987 **Estimate:**

988 Is the numerical value computed by an estimator based on the observed clinical trial data.

989

990 **Estimator:**

991 Is the analytic approach to compute an estimate from observed clinical trial data.

992

993 **Intercurrent Events:**

994 Events that occur after treatment initiation and either preclude observation of the variable or
995 affect its interpretation.

996

997 **Missing Data:**

998 Data that would be meaningful for the analysis of a given estimand but were not collected.

999 They should be distinguished from data that do not exist or data that are not considered
1000 meaningful because of an intercurrent event.

1001

1002 **Principal Stratification:**

1003 Is the classification of subjects according to the potential occurrence of an intercurrent event
1004 on all treatments. With two treatments, there are four principal strata with respect to a given
1005 intercurrent event: subjects who would not experience the event on either treatment, subjects
1006 who would experience the event on treatment A but not B, subjects who would experience
1007 the event on treatment B but not A, and subjects who would experience the event on both
1008 treatments.

1009

1010 **Principal Stratum:**

1011 Is used in this document to refer to any of the strata (or combination of strata) defined by
1012 principal stratification.

1013

1014 **Sensitivity Analysis:**

1015 Is a series of analyses targeting the same estimand, with differing assumptions to explore the
1016 robustness of inferences from the main estimator to deviations from its underlying modelling
1017 assumptions and limitations in the data.

1018

1019 **Supplementary Analysis:**

1020 Is a general description for analyses that are conducted in addition to the main and sensitivity
1021 analysis to provide additional insights into the understanding of the treatment effect. The
1022 term describes a broader class of analyses than sensitivity analyses.