EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

26 April 2018
EMA/744060/2017

Overview of comments received on Draft ICH E9(R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (EMA/CHMP/ICH/436221/2017)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	Robert G. Newcombe PhD CStat FFPH HonMRCR, Emeritus Professor of Biostatistics, Cardiff University, UK
2	Prof Nick Freemantle PhD, Comprehensive Clinical Trials Unit, University College London, UK
3	Paul-Ehrlich-Institut, Langen, Germany
4	CBG-MEB
5	Teva Pharmaceuticals
6	Instituto de Evaluación Tecnológica en Salud (IETS), Colombia
7	Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany
8	Pfizer
9	Gilead Sciences International Ltd
10	AstraZeneca
11	Boehringer Ingelheim
12	ACRO (Association of Clinical Research Organizations)
13	German Region of the International Biometric Society (IBS-DR) German Society for Medical Informatics, Biometry and Epidemiology (GMDS) Network of coordinating centres for clinical trials (KKS-Netzwerk)
14	International Society for Pharmacoeconomics and Outcomes Research (ISPOR)
15	AGES – Austrian Agency for Health & Food Safety, Federal Office for Safety in Health Care
16	BfArM: Federal Institute for Drugs and Medical Devices, Bonn, Germany
17	International Society for Clinical Biostatistics: Statistics in Regulatory Affairs

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Stakeholder no.	Name of organisation or individual
	Subcommittee
18	EUCROF - European CRO Federation
19	Mark Lomax, Mundipharma Research Ltd.
20	Novartis
21	PSI/EFSPI/EFPIA

Please note that comments will be sent to the **ICH E9 IWG** for consideration in the context of Step 3 of the ICH process.

1. General comments – overview

Stakeholder no.	General comment (if any)
1	Before my attention was drawn to this document I had not encountered the term 'estimand'. My first thought was, At last, there'll be due emphasis on effect sizes and confidence intervals for them! No such luck. On reading the ICH E9 addendum this seems to relate more to issues of intercurrent events than to effect sizes. Effect sizes are mentioned, but interval estimates are not mentioned at all. While taking intercurrent events into account correctly is a major issue in many types of trials, this does seem to have been a missed opportunity to also reorient the reporting of findings away from p-values with their known deficiencies as a means of drawing inferences, towards a more informative use of effect sizes and confidence intervals. See for example my book Confidence Intervals for Proportions and Related Measures of Effect Size, available at http://www.crcpress.com/product/isbn/9781439812785.
2	The concept behind this addition to GCP is welcome, and I turned to the draft with some anticipation (ICH E9 has proven remarkably effective as a tool to improve the state of science in clinical trials in my view). Unfortunately, I do not believe this document advances our position in a positive way, indeed I fear that it opens the door for sponsors to undertake suboptimal research and hide behind some of the methodology (bits of it especially MAR imputation being wildly inappropriate and potentially biased and misleading) identified. It would be much more helpful if the document concentrated on the situations where data were not available for structural reasons beyond the investigators control. For example we are undertaking a trial of methylphenidate for fatigue in patients under palliation. In that trial we expect up to 25% of patients to die prior to the primary endpoint fatigue scale measure. Our plans to deal with that challenge would not be soundly aided by the document under review (there is too little detail included and too much reliance on making sound judgements and assumptions). There is also the absence of adequate referencing. There is an extensive literature on this area (of mixed quality); it would be helpful if the document cited this. So overall I do not believe that this document as currently developed is helpful nor does it meet the high standard and excellent utility of other ICH publications. It would fit much better as a discussion document issued by EMA and should in my view remain as apocrypha.
3	The Paul-Ehrlich-Institut appreciates the addendum to the guideline. However, with 23 pages in total, 13 pages thereof being the actual guideline text and 7 pages example(s), the addendum is considered rather extensive and it is very repetitive in places. A concise presentation of the topic would foster dissemination and discussion as it will be easier to find the time and willingness for clinicians and statisticians to read the text. A long version as currently presented helps, of course, for a deeper insight and could be published in a scientific journal to supplement a brief and concise guideline.
4	The concept of estimands and the handling of intercurrent events is becoming increasingly important. Therefore, clarification of these concepts and the framework provided in this draft addendum is appreciated and it is anticipated that this addendum will encourage and support discussion on these matters between regulatory bodies

Stakeholder no.	General comment (if any)
	and industry, as well as increase the awareness of these issues in the broader research community. In general, we agree with the content and key messages of the addendum. However, we have provided some comments below on how the addendum could be further enhanced in order to improve the readability, especially for non-statisticians, and to support a successful translation into disease-specific guidelines. In order to make the addendum more accessible to a broader clinical trials audience, it is suggested that clear, detailed and complete examples are included. These examples should use the kind of wording that sponsors would be expected to use in their protocols and refer to realistic and recognizable diseases. Complete examples should be included for each labelled strategy and include at least realistic examples for oncology, a safe and mild treatment for symptomatic relief like nasal decongestion and a disease modifying drug (e.g. anti-rheumatic). It is suggested to include these examples in a separate chapter which can be referenced throughout the guideline. Otherwise, consider introducing the examples in A.7 as early as possible in the text: for example, A.7 could be moved to between A.3.1 (formal definition of estimand) and A.3.2 (five generic estimands). Then the description of the abstract strategies in A.3.2 will be more tangible because the reader will have a more concrete example in mind. It is considered that the addendum would benefit from more disentanglement of the concept of an estimand from the concept of missing data, especially since discussion about the estimand tends to arise in dossiers with missing data as well. In the current version of the addendum, the problem of missing data as a separate issue to the estimand is only touched upon in sections A.4, A.5 and A.7. If in section A.3.2, describing the different strategies, there was a subsection referring to what would constitute missing data (as it is partially done in section A.7 "a generic example"), this would probably make the distinction between these two concepts more clear. The text is quite a long read due to repetitions. Usually the first sentence of the paragraphs are very informative, The text could be made shorter, by removing repetitions and more efficient use of examples. Variable is a rather statistical term; the term ‘endpoint’ may be more accessible to the broader audience.
5	• From line 13 it seems that the focus of the addendum is on pivotal studies. If this is the intention it should be stated more clearly in the Scope. Moreover, the expectations for other phases are not clear. In particular, the objectives of proof-of-concept and pivotal studies may be different which may lead to different choices of estimands and hence comparison between studies may be problematic. • Non-inferiority is mentioned briefly in line 427. Equivalence is not mentioned and there are similar issues for these two types of studies. Moreover the FDA guideline on non-inferiority discusses a statistical approach to select a non-inferiority margin based on previous studies (and this can be based on literature). Since future analysis will be based on the prior definition of estimands, it is not clear if past

Stakeholder no.	General comment (if any)
	studies can be used in order to derive a margin. • Estimands considerations may be more complicated in the context of adaptive designs, for example, when the adaptation is to a sub-population for example. It would be helpful if this issue will be discussed. • Starting at line 100, the definition of estimand includes the term “the variable” presumably to address the outcome of interest in the trial. I believe that a different word other than “variable” would be used, potentially “outcome” or “endpoint”
6	This document is an important contribution as an easy-to-understand guide to the possibilities of statistical analysis for clinical trials. Its relevance is given in the fact that by generating an experimental (or observational) design, the construction of good statistical tests and good information systems, must provide the researcher with a robust basis for comparison and criteria to reference analytically if there are significant differences between study groups (treatment and control). In general terms, this document seems complete, concrete and with a coherent logical sequence. The definitions are clear and the technical specifications include a good explanation (supported by examples).
7	IQWiG appreciates the opportunity to comment on the draft guideline. The new ICH E9 (R1) addendum on estimands and sensitivity analysis represents a mixture of useful clarifications, trivial explanations (neglecting well-known approaches of evidence-based medicine), and a number of critical issues. From a Health Technology Assessment (HTA) point of view, some of the listed estimands are not relevant and cannot be estimated in an unbiased manner. By suggesting these five different strategies without clarifying which of them would be a general or minimum requirement and which of them might only be used as sensitivity analyses or in special situations, the addendum weakens the requirement of robust analyses for regulatory decision-making. In addition, for HTA it is essential that data collection for all endpoints is carried out for all patients up to the end of the study. We see the danger that, for example, the “while on treatment” strategy described in this addendum will be used as a justification for refraining from endpoint data collection after discontinuation of the initial treatment and will thus jeopardise the analyses required for HTA. In addition, this is not consistent with the recently adopted “Guideline on evaluation of anticancer medicinal products in man” (EMA/CHMP/205/95 Rev. 5, compare section “Extended safety data collection”). We recommend revising the addendum, taking the well-known PICOS approach into account, and avoiding estimands that cannot be estimated without a high risk of bias and that contradict the statistical principles for clinical trials of the ICH E9 guideline. IQWiG strongly supports a revision of the addendum, because only two of the described strategies (treatment policy, composite) should be used in general as the main analysis. The other three strategies (hypothetical, principal stratum, while on treatment) are useful only as a possible supplementary analysis for hypothesis

Stakeholder no.	General comment (if any)
	generation or sensitivity analysis in special situations. So far, this does not become clear from the addendum. Therefore, there is the risk that the addendum is considered to be supportive of inappropriate analyses of clinical trial data.
8	Informed consent Given the emphasis on collecting as much data after an intercurrent event, would it be worth mentioning some choices of informed consent that could be used, e.g. to request that patients continue their scheduled visits even after an event such as discontinuation of study drug?
9	The guideline is too abstract and not easy to follow at the beginning until the examples are given later. Principal stratum strategy is a post hoc analysis and is problematic because subjects in the stratum cannot be identified before treatment starts. This estimand is not very meaningful for real-world practice. Treatment effect from early phase studies cannot be extended to late phase studies because treatment effect using principal stratum strategy is likely over-estimated and cannot be used to design a registrational trial. At best it tells us that the drug may work in the subset of patients who will not require rescue medication. Similarly the hypothetical strategy has confounding issues and requires many un-testable assumptions. Therefore usage of neither principal stratum nor hypothetical strategy should be encouraged.
10	AstraZeneca fully supports the General Comments made by EFSPI/EFPIA on the ICH E9 Addendum. However, we would like to stress the importance of 3 of the comments made in particular (reproduced on page 3 for convenience). Addressing these points will be crucial for the successful implementation of the framework described in the Addendum. The first point relates to the importance of Case studies. It is understood that Case studies will not be included in the Addendum but instead will be provided as slides that will be released soon to explain the Addendum. This is supported but there is a concern that these case studies will not be detailed enough and it is understood that there will be no opportunity for industry to comment on these case studies even though they will be released at least one year before the Addendum is completed. A way needs to be found for improvements to be made to these Case studies if necessary. The second point stresses the important of regulators moving quickly to modify disease specific guidance documents to provide advice on estimand strategies that may be acceptable in specific situations. Waiting until the Addendum is finalised to initiate this process will waste valuable time in giving sponsors and regulators advice. The third point relates to drug labels. It is essential that drug labels give clear information on the estimand strategy employed that leads to the estimated treatment effect given in the label. Also if a regulator uses a different strategy to that pre-specified by the sponsor in their decision making this needs to be clearly explained to make the process more transparent. It is understood that labels/SmPCs have region

Stakeholder no.	General comment (if any)
	specific rules and regulations that apply but it is important that the Addendum makes a clear statement stressing the importance of transparently reflecting the estimand strategy used in the label. 1. We acknowledge and agree that guidelines need to outline principles. However, there is a need for more case studies to support this addendum. It is proposed that the generic examples in section A.7.A are supplemented by a companion of case studies. The case studies should be a detailed description of a real situation where different strategies for handling intercurrent events are described and the pros and cons of each approach to address them as it relates to the question of interest are articulated. However, it is understood that the drafting group plans to address the lack of examples in the main text by producing case studies in slides that will accompany the Addendum. There are several issues with this approach. Firstly, it is understood that these case studies will be made available shortly, but the final version of the Addendum is unlikely to be completed until the end of 2019. Also, there is no consultation process planned for commenting on the case studies and hence if they could be improved upon there is no opportunity for this to happen, even though there is probably over a year available to do this. It is therefore strongly encouraged that this process is changed to allow comments on the case studies if the Drafting group does not include extensive examples in the main text of the Addendum. 2. The addendum clearly impacts existing disease specific guidance. Regulatory authorities should produce a prioritised list of disease specific guidance documents to revise in light of the Addendum. Given the Addendum will not be finalised until 2019 work should start on revising disease specific guidance documents as a matter of priority to provide better guidance on estimand strategies that may be acceptable in specific disease areas/indications. What plans are there for review of these guidelines once the addendum is issued? 3. It is very important that how intercurrent events have been handled is transparently described in the drug label/SmPC . Please clarify in the Addendum how estimands will link to drug label/SmPC, in particular when a regulatory authority bases its decision on a different Estimand strategy to that pre defined by the Sponsor.
11	The estimands framework described is generally welcomed due to the increased structure and clarity it provides around the handling of intercurrent events. However, it does add complexity to trial planning and documentation. There are also considerable barriers to implementation due to the change in thinking required compared to existing practice. It is therefore important that the material benefits from the framework are sufficient to outweigh the additional burdens. The emphasis that sensitivity analysis should refer to robustness of assumptions made in the estimation of an estimand, rather than investigation of alternative estimands, is greatly welcomed as a means to reduce complexity of clinical trials and reduce unnecessary analysis. Sensitivity analysis should be recognised as the primary way of responding to concerns regarding estimators being insufficiently robust. Since estimands affect all clinical trials, the estimands framework should be mandatory

Stakeholder no.	General comment (if any)
	for all trials, including single-arm and pragmatic trials. We believe that this is the intention of ICH E9(R1), but it is not clear from how it is written; we request that this is clarified. Applying the estimands framework to all trials simplifies implementation, increases compliance and avoids ambiguity around whether it applies. It is concerning that no attention is paid to choice of estimand in early-phase trials, and the focus of estimand choice/strategy within the document appears to only be from a regulatory perspective on pivotal trials. We therefore request that general guidance on the choice of estimand in early phase trials is provided. In particular it should cover how clinical relevance in phase I/II trials differs from late phase trials, and how suitable early phase estimands can inform both the choice of estimand and expected treatment effect size for later phases. As estimands provide exact definitions of the scientific questions of interest, they apply to all types of analysis, not just those of efficacy. In this context, it is disappointing that there is no discussion of choice of estimands for safety data. In particular, whether estimands for efficacy and safety should be coherent. It is felt that in order to do a fair benefit-risk assessment, intercurrent events handling and patient populations ought to be comparable between efficacy and safety analyses. 'Intercurrent events' is viewed unfavourably as a term because the word 'intercurrent' is neither in common usage (and hence extremely poorly understood *a priori*), nor is it particularly succinct. More importantly, 'intercurrent' is already existing clinical terminology and has a different and distinct meaning from its usage here; it refers to the occurrence of a disease while another disease is also present. This is a serious issue for its usage when clinicians are one of the main target audiences. More easily understood would be 'post-randomisation events' as it is simpler and clearer to understand. The name 'hypothetical' adopted by this document for the estimand strategy is negative in connotation, suggesting clinical irrelevance (when, ironically, it is typically of great clinical relevance), and we would strongly request that a less pejorative name be used. In general, insufficient weight is given to the utmost importance of clinical relevance: We believe that regulatory dialogue on choice of estimand(s), and the subsequent regulatory decision-making, should be based on the one(s) of greatest clinical relevance. Concerns over bias and robustness to assumptions should not change the estimand, but instead be reflected in good trial design, suitable statistical analysis and by sufficient sensitivity analysis. There is considerable, unwarranted, editorial preference expressed towards treatment policy (and to some extent composite) strategies, particularly in sections 3.3 and 5.1. Lines 321-322 state that there might be "some circumstances" where other estimands may be more relevant than treatment policy, and even then only provided 'a robust estimate can be obtained'. Lines 335-338 then state that treatment policy may be more acceptable even where less clinically relevant. Conversely, lines 350-352 state that a hypothetical estimand's conditions must be justified in their relevance in clinical practice, with the entire paragraph (lines 349-364) dedicated to critiquing its clinical relevance. In contrast, nowhere are the clinical shortcomings of treatment policy referred to and it is not requested that its clinical relevance be justified. There appears therefore to be an editorial line that treatment policy should be the default estimand strategy, where deviation should be justified.

Stakeholder no.	General comment (if any)
	This is harmful from several perspectives: Firstly, clinical relevance should primarily determine the estimand, and this will vary between individual circumstances, making defaults unwelcome. Secondly, early phase trials are not run to directly secure regulatory approval, but to inform later trials; treatment policy strategies will rarely be relevant in these settings. Thirdly, this bias is blinding to the flaws of treatment policy, the statistical ones of which are described in other comments. Fourthly, if regulatory opinion changes over time, current preferences for estimand strategy will remain enshrined in the document. ICH E9(R1) should neutrally set out choices without an editorial line; estimands should be based on the trial and compound specific circumstances. In certain therapeutic areas, it is common to seek regulatory approval using biomarker data from placebo-controlled trials to measure efficacy (e.g. HbA1c, FEV1), with the subsequent regulatory requirement to run outcome trials. Where the clinical goal is efficacy, and subsequent treatment policy-based outcome trials are planned, it is generally not clinically relevant to use treatment policy for the pivotal trials and this approach will result in artificial attenuation of treatment effect. Treatment policy strategies appear clinically relevant mainly for hard outcome endpoints or assessment of treatment regiments. We believe that it is good clinical practice to follow-up patients who discontinue until the scheduled end of their trial participation as much as is logistically and ethically possible. We separately note that although hypothetical estimands disregard post-intercurrent event data, such data can be useful in assessing the reasonableness of sensitivity analyses, since it can help demonstrate reasonable future performance of the worse-performing patients who receive no benefit from the randomised treatment (and hence provide information about clinically reasonable penalties for e.g. tipping point analysis). Little consideration is given to the practical statistical issues implied by wider adoption of treatment policy, composite or principle stratum strategies, including increases in statistical complexity and loss of power. Some estimands, such as principle stratum, may not be estimable for certain trial designs. The greater the statistical complexity, the greater the chance that a pre-specified analysis fails outright (e.g. treatment policy using observed post-intercurrent event data for small groups), the more assumptions that are required and the harder it is to assess the suitability of these assumptions. Issues that are currently known about and recognised (e.g. in MMRM) may become hidden behind complex models that are poorly understood (e.g. causal inference for principal stratum), and new issues will be created, perhaps without it being realised. Wider adoption of treatment policy will lead to larger and more expensive pivotal clinical trials due to smaller estimated treatment effects. Additional usage of composite endpoints will also result in lower power from loss of information due to conversion of continuous data into binary, time-to-event or partially-binary-composite data. More complex composites, such as those treating intercurrent events as ‘bad results’, may result in important violations of assumptions of normality. There is no overt consideration given to how estimands fit into time-to-event or recurrent event endpoints; to include this would be welcome as it is less straightforward how the estimand strategies presented fit in these contexts. While the examples in section 7 are welcome, particularly in regards to how to describe estimands, they are still ‘hypothetical’ and hence lack the real-world details

Stakeholder no.	General comment (if any)
	that make choice and implementation of estimand so difficult. The example of two intercurrent events simultaneously complicates matters while being over-simplified, since other types then need to be considered (e.g. discontinuation for other reasons, death, switching after discontinuation to other treatments, patients who discontinue due to an AE then take alternative treatment). All examples remain stuck in the 'classical' estimands setting of continuous endpoints.
12	The Association of Clinical Research Organizations (ACRO) represents the world's leading, global clinical research organizations (CROs). Our member companies provide a wide range of specialized services across the entire spectrum of development for new drugs, biologics and medical devices – from discovery, pre-clinical, proof of concept and first-in-man studies through post-approval and pharmacovigilance research. With more than 130,000 employees engaged in research activities around the world (including 57,000 in Europe), ACRO advances clinical outsourcing to improve the quality, efficiency and safety of biomedical research. Each year, ACRO member companies conduct more than 7,000 clinical trials involving 1.3 million research participants in over 100 countries. On average, each of our member companies works with more than 700 research sponsors annually. ACRO welcomes the opportunity to comment on the planned ICH E9 (R1) revision and has provided specific comments below in order to increase its usefulness to sponsors and other parties involved in clinical trials. ACRO also has a number of general concerns with the proposed guideline. These are as follows: • The stated aim of the guideline is "to facilitate the dialogue between disciplines involved in clinical trial planning, conduct, analysis and interpretation, as well as between sponsor and regulator, regarding the treatment effects of interest that a clinical trial should address". However, we think it would be difficult to use this document as guidance as it is currently written. We recommend that the document should be reduced in size, be less repetitive, and that some of the text is simplified to make it more accessible to a wider audience, given that consideration of estimands will involve multi-disciplinary teams. The current draft document is a difficult read and we recommend that it would benefit from stating the key points once rather than repeating them. In ACRO's view, a document comprising key recommendations illustrated with more practical/real examples would be more suited to achieve the stated aim. • The terminology around estimand/estimate/estimator is hard to follow. Vague language is used too often: 'might' and 'may', for example. While 'must' and 'mandatory' may be too strong, 'should' or 'it is advised' would provide clearer direction. • Throughout the document, the notion of "intercurrent event" seems to be the major driver for this draft guideline, and it seems like the estimand approach has been introduced only or mostly to deal with intercurrent events. If so, ACRO recommends that it would be simpler and clearer to develop a guideline on handling intercurrent events, with specific recommendations for data collection and useful sensitivity analyses (instead of introducing all the complex terminology around estimands).

Stakeholder no.	General comment (if any)
	• It is difficult to see how the estimands approach will enhance the way in which data are currently reported, analysed and understood. For example, (1) it is not clear how the principal stratum strategy differs from a per-protocol or subgroup analysis, (2) if the intake of rescue medication forms an intercurrent event, it is not clear how this is different from presenting those who take rescue medication while evaluable for the primary endpoint separately from those who did not. It would be helpful if more parallels were to be drawn with existing methods or explain why this is different (if it is) from what is currently being done. • There is inconsistent use of the glossary terms. Sometimes the quantity is defined (e.g., line 121), other times the glossary is referred to (e.g., line 70). ACRO recommends cross-referencing the glossary in all cases and within the text of the document to identify terms which are available in the glossary using bold and/or italics. • ACRO recommends re-ordering the contents of the document to make it more easily readable, e.g., starting with section A1 and then providing worked examples to introduce the different types of estimands. • Guidance is not included on how the use of estimands will fit into reporting in public clinical trial databases (e.g., clinicaltrials.gov): it is not clear whether estimands be used in place of endpoints (so that the entry is restricted to a maximum of 10) or up to 10 endpoints could be reported, with multiple estimands created, based on these. ACRO recommends that such guidance be included in the final version of the document. • In the descriptions, line 154 defines C as “The specification of how to account for intercurrent events to reflect the scientific question of interest”. However, ACRO recommends that it would be helpful for each time to define in C, the intercurrent event and how to account for it. In the general example, the “C. Intercurrent event:” seems to switch between describing what the intercurrent event is and describing the handling of the intercurrent event, without specifying the intercurrent event. For example, line 609 (“No intercurrent events to be taken into account”), line 637 (“regardless of whether or not switching to rescue medication had occurred)” and line 696 (“had rescued medication not been made available to subjects prior to month six”), define the intercurrent event; however, line 669 (“the intercurrent event is captured through the variable definition”) and row 719 (“the intercurrent event is captured through eh population definition”) specify the handling of the intercurrent event. ACRO recommends that both the intercurrent event and its handling should be included in the specification, and that this should be made clear in the final guidance document. ACRO thanks the EMA for the opportunity to provide comments on “ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials.” Please contact ACRO (knoonan@acrohealth.org) if we can answer any questions or provide additional details.

Stakeholder no.	General comment (if any)
13	The ICH E9 (R1) addendum stresses the importance of the detailed clarification of the scientific questions of treatment comparisons in the presence of intercurrent events before deciding on the analytical methods. Sensitivity analyses with regard to statistical methods and underlying assumptions for one estimand are distinguished from sensitivity analyses with regard to the choice of the estimand, i.e. the scientific question. This approach is much appreciated. The problem is that the five estimands are listed on the same level without giving explicit advice on preferred estimands in specific situations from the regulatory point of view. It would be helpful to discuss estimands for specific scenarios and reflect on the regulatory point of view resp. discuss the perspective of different stakeholders. There is little discussion on the feasibility applying specific strategies in given scenarios and the methodological challenges coming with it. However, this discussion is needed in order to establish the guideline in practice. Especially the role of the hypothetical estimand and of the principle stratum estimand for regulatory decision making has to be questioned as they rely on untestable assumptions. For this reason the ICH E9 (R1) addendum should provide • examples for the hypothetical strategy, including how to estimate the estimands in this scenarios with low risk of bias and including the extent and type of expected sensitivity analyses, • examples for the case the principle stratum strategy would be the preferred estimand and which methods are available for a robust estimation and advice on expected sensitivity analyses. Otherwise, the addendum can be perceived as providing a comprehensive framework without giving any recommendation for the application. The described estimand framework seems to address efficacy analyses. Considerations on appropriate estimands for safety endpoints should be added.
14	ISPOR is the leading global scientific and educational not-for-profit organization for health economics and outcomes research and their use in decision making to improve health. With over 20,000 individual and chapter members worldwide, our mission is to promote health economics and outcomes research excellence to improve decision making for health globally. We appreciate the opportunity to respond to call for comments on ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials. While this addendum focuses on a specialized area of statistical principles for clinical trials, ISPOR has a vested interest in regulatory data which is used by reimbursement authorities, physicians, and patients for coverage and treatment decision making. From the opening sentence of this addendum: “To properly inform choices …by patients and prescribing physicians, clear descriptions of the effects of a medicine should be available,” it is clear that our ‘constituent’ audiences and data needs overlap. Thus, we

Stakeholder no.	General comment (if any)
	feel it is important to consider reimbursement authorities, and the health technology assessors who inform them, to be consumers of these data and analyses as well. ISPOR's response was formulated in coordination with leaders of several of ISPOR's Councils and Special Interest Group (Statistical Methods, HTA, Institutional, and Health Science Policy) along with input from interested members of these sub-groups. To solicit such input, we asked members to respond to an on-line survey. Recognizing the technical nature of this addendum, as we expected, most responses came from our statistical experts. We received 15 responses in addition to the comments from our sub-group leadership. We felt that this new guidance would have a positive impact on the way trials are conducted, particularly with respect to the transparency and applicability of estimates of treatment effect. However, there were some areas to clarify. One of the most often mentioned areas is the impact that proper estimand specification could have on the ability of efficacy estimates to more closely answer the research question relevant to a real-world setting (or not) depending on how they are defined. Healthcare decision makers often want to know how a treatment will perform outside of a well-controlled setting. There is some concern about how to apply estimands to a pragmatic or real-world trial setting as this was not mentioned in the current guidance. On one hand, if the estimands are defined such that they include the intercurrent events as they happen in real practice (i.e. treatment switching), it would give a better view of how the product may work outside of the randomized controlled trial (RCT) setting. However, the opposite can occur - the estimand could be defined so that it leaves out the intercurrent event and gives a much more narrow view of treatment effect, which is less relevant outside of the RCT setting, especially to payers. To that end, more examples or details regarding the handling of intercurrent events are needed in the guidance. The examples should be structured around categories such as disease area or type of endpoint (time to event, continuous, etc.) to give more clarity. Estimands, by providing a standardized framework for research questions, could increase transparency and usefulness of clinical trial outcome results. While it may increase the time (and cost) of upfront trial planning and the number of analyses needed to report the endpoints, this could be offset by a decrease in the probability of having a study that fails or is uninformative due to inappropriately defined endpoints, and thus in the end could save resources. However, this will require a multidisciplinary approach to the estimand/trial design from the very beginning. We suggest that such a multidisciplinary approach be reflected in this guidance more strongly. The guidance will have an important role in future dialogues between drug developers, regulatory bodies and health technology assessors on requirements for evidence generation. It will be essential to ensure that sufficient support is provided by clinical, regulatory, and HEOR/market access personnel, otherwise there is risk that the development process could be delayed or that the ultimate estimands may not be fit for purpose, especially outside the regulatory arena.
15	We fully support the initiative in general to develop this ICH E9 addendum on the topic of estimands. We agree that too often in clinical trials what is actually measured and in consequence the target of statistical inference and estimation is only implicitly specified. To verify whether the target of statistical inference and estimation is aligned

Stakeholder no.	General comment (if any)
	with the scientific objectives of the trial requires careful consideration of the combination of endpoint, study population, statistical analysis method, and missing data handling approach. Any initiative to shift the specification and alignment of these concepts to the planning stage is welcome. Not only would it make the connection between scientific (clinical) meaning of the results less opaque, but it could also avert a number of possible discrepancies between a trial's results and what a treatment under scrutiny actually does. In general, anything that is prospectively specified increases the credibility of the results, once the data have been obtained. The document is in many instances, likely dependent on the addressee, difficult to follow. The main concept appears well structured, but its application to a given case/example is perceived as challenging, particularly in the absence of a clear normative position taken on what is the preferred strategy from a regulatory perspective (even if not necessarily achievable or feasible in all cases). In this sense, it is hard to recognise whether the addendum is eventually intended as a rather neutral description of options to address the estimand question, or if there is the intention to provide guidance, at least to some extent. In this context, section A.3.3.2 seems to be of high relevance as it contains an attempt to link the introduced estimand strategies to therapeutical and experimental context: These aspects will be of biggest interest for those who want to follow (new?) regulatory guidance. The more agreement on ICH level could be achieved regarding '<u>which strategy would be considered most suitable under which circumstances</u>', the more helpful this addendum will be in practice. One might need to think in broad categories, e.g. 'symptomatic treatment' vs 'disease modifying'; or 'many alternative treatment options available' vs 'no alternatives available'; or 'therapeutic' vs 'prophylactic'; etc.. Even if no agreement on clear-cut recommendations regarding estimand strategy under specific circumstances could eventually be achieved, at least tendencies for what might be preferred could serve as valuable starting point for the user. Good guidance for the (estimand) decision makers is key, as the addendum generally implies a huge number of possibilities to come up with an estimand(s) choice for a particular situation (trial): at least five strategies are introduced, which could also be blended. Furthermore, there is the option or even need to follow different strategies of dealing with intercurrent events within one trial, and the repeated mentioning that there might be more than one estimand per trial in general. It is felt that this opens a too broad field, if no further guidance is provided. Clinical relevance of a strategy might not always mean that a conservative strategy has been adopted when it comes to making a decision about approvability of a certain IP (e.g., see lines 335-338). Lack of (broad) clinical relevance (and this diction is unlikely to allow for a clear black & white distinction) might eventually in presence of appropriate estimand/estimator and meaningful estimate still be managed at the SmPC/label level in the context of benefit/risk assessment during MAA. Put differently: a fundamental question that is perceived as not explicitly addressed in the draft addendum document is whether regulatory decision making prefers

Stakeholder no.	General comment (if any)
	'conservative' over (clinically) most meaningful experiments and analyses. It might hence be worthwhile to consider a section which addresses the balance between striving for optimal estimands and the more pragmatic approach to reflect actual limitations and deficiencies of a certain development program/study via labelling. Throughout the text of the draft, one might find redundant information, which could be shortened and omitted. Another aspect which makes reading sometimes cumbersome is the fact that often explanations, examples and recommendations are mixed (e.g. lines 279 ff: the introduction part of this section is redundant). The draft mostly fails to provide an explicit definition of the estimand. Hints into that direction are scattered around the text. The main section, however, provides only an implicit definition (i.e. via population, endpoint, IE handling and population summary measure). Moreover, the discussion is largely centred around handling of intercurrent events. It becomes unclear whether this is the main purpose of the debate, or whether the concept has more general implications. For example, may an estimand specification also be useful in trials where intercurrent events are unlikely to occur? For the most part the concepts discussed in the document are illustrated by the example of a trial involving an endpoint measured only sometime after treatment. During that time subjects may stop to adhere to their prescribed treatment or require rescue medication. While this is an interesting – and certainly practically important - example for intercurrent events, we are not convinced that it is representative for the estimand debate as a whole. More examples, e.g. involving non-inferiority or bioequivalence would be welcome. From a statistical perspective, more precision in the language would be welcome. It is understood that the main audience of this guidance may be non-statisticians, which warrants a certain degree of simplification. However, we feel that too often a clear reference to the underlying methodological concepts is missing. Important statistical concepts - bias, variance - are not even considered. Rather, quite general terms like 'robustness' and 'reliability' are used. And it remains unclear to which extent this terminology is aligned with corresponding concepts in statistical theory. For example, it appears that "robust" in line 539 refers to a broader concept of "robustness" - that e.g. includes also sensitivity of the estimator to assumptions about missing data - as compared to what is studied in the field of robust statistics, which focuses on an estimator's sensitivity with respect to outliers. Consequently, it is unclear how "robustness" could be quantified or even verified. Importantly, it needs to be clarified how the estimand concept discussed in the guideline relates to the estimand concept discussed in the causal inference literature. In the latter the estimand represents the parameter of the (causal) statistical model targeted by the inference procedure. In addition, while the topic of missing data is amply discussed, issues with competing risks are not. Considering that certain IEs could be seen as competing risks, this appears like a missed opportunity. We suggest to at least mention this relation. The distinction between statistical decision making in terms of hypothesis tests and estimation should receive more attention. Differences in the degree to which the

Stakeholder no.	General comment (if any)
	estimand concept plays a role for these two objectives should be discussed. For example, do point estimates and statistical hypothesis test necessarily have to target the same estimand? Would it be acceptable to combine a hypothesis test that targets an estimand, which is clinically less meaningful (e.g. ITT) but guarantees a conservative analysis with an estimate that targets an estimand of improved clinical relevance, which, however, may only be obtained under additional assumptions (e.g. principal stratum estimand)? Section 1 (Purpose and Scope) might benefit from shortening. It could focus on the problem statement, which is the fact that patients respond to treatments in different ways (intercurrent events) and that this complicates the interpretation of treatment effects in a clinical trial. The aim of a correct choice of an estimand are: >> To inform on trial design, conduct and analysis (main analysis and sensitivity analysis) >> to give a clear definition of the treatment effect measured, in order to bring consistency in inference and decision making within and between regulatory regions >> To inform the choices made by prescribing physicians; suggested alternative wording here: “To provide a clinically meaningful summary of the benefts and risks on which prescribing physicians and patients can make their choices”. Section A.3.2 introduces “five strategies for constructing estimands”, and it is not clear why the title of this section is “Strategies for addressing intercurrent events”, and why these construction strategies would not fall under a heading “Construction of estimands” as chosen for A.3.3; There are four attributes to describe an estimand: Point A, B, and D are attributes that are already currently standardly defined. The most “new” idea about the estimand topic is the consideration of intercurrent events, which can then subsequently be incorporated into the population (A) or variable/endpoint (B) definitions. In this sense, the component C of IEs appears conceptionally different from components A, B, and D. Our suggestion is to reconsider this issue for the wording.
16	In general, the presented structured framework on estimands and sensitivity analyses is very much appreciated. It identifies an important source of ambiguity related to the target of a clinical trial and helps to clarify the handling of distinct events happening after randomisation, as treatment discontinuation, that led to potentially inconsistent requirements and interpretations in the past. It represents an excellent basis to support the identification of relevant estimands in indication specific guidelines and provides a common terminology that is important for harmonization and a targeted discussion with the different stakeholders. Although the full understanding of the document may need intense familiarization with the topic and potentially training for non-methodologists, it is acknowledged that the addendum is written in a comprehensive form that allows for an adequate appreciation of the presented issues. Some more introductory remarks on the development and history of the topic during the last decade, e.g. starting with missing data issues and treatment non-adherence may have supported a general understanding of the issue, but it is also understood that the guideline (which should be limited in length) is not

Stakeholder no.	General comment (if any)
	considered to be a textbook. From discussion with different stakeholders, it appears that the selection of estimands may be highly controversial depending on the indication and setting. Hence it may be worthwhile to further stress, that the addendum is intended to primarily provide a principle framework. Some of the proposed options may finally not be considered in most settings but it is considered important and helpful to identify the potential options to support a well informed decision on the relevant estimand to choose. While there is consensus that ITT (intention to treat) analysis is considered the gold standard for analysis of RCTs, there is quite a lot ambiguity in defining and applying the principle in practice. While the original ICH E9 guideline is not that clear on defining ITT, the addendum defines the ITT principle clearly as addressing 'treatment policy' (i.e. all random. patients according to randomized allocation and using complete follow-up data of all patients). This has some unfortunate consequences: • Considering the ITT principle as addressing 'treatment policy' would mean that only RCTs addressing treatment policy would be conform to the ITT principle. • Trials primarily addressing other relevant estimands (such as estimands using a composite strategy) would then not conform to the ITT principle. However, they are still based on the principle that all randomized patients should be included according to randomized allocation to avoid baseline confounding. • Since primary estimands other than treatment policy are possible according to the addendum, the ITT definition in the addendum (page 7 lines208/209) means stepping away from ITT being the gold standard to analyse RCTs. • If the ITT principle is defined as addressing 'treatment policy', the principle to include all randomized patients according to randomized allocation is still fundamental and needs a new name. • In practise, ITT has been rarely applied in the sense of treatment policy. Patients were often withdrawn from the study after experiencing an intercurrent event and although analysis methods (for example MMRMs) did not address a treatment policy estimand, they were still claimed to be ITT analyses. In summary, defining ITT not as addressing treatment policy but simply as including all patients as allocated would allow a variety of ITT conform trial objectives/estimands. This would allow ITT to remain the gold standard of analysing RCTs while not needing a new term for the fundamental principle to include all randomized patients according to randomized allocation that also concerns all the other estimands/strategies to handle intercurrent events. The problem of having/not having complete follow up, which is needed for robust estimation of treatment policy, is not part of defining ITT. Missing data handling is instead addressed when an analysis aligned to the targeted estimand is chosen, while ITT in our view is simply a definition of the analysis population/set (all randomized as allocated). The document should more clearly state that an estimand is an entity which is

Stakeholder no.	General comment (if any)
	unrelated to a specific clinical trial but a parameter that describes how the new drug would work in a specific well defined setting and compared to the situation in which the drug had not been prescribed. Hence, the estimand should not be derived from a clinical trial setting, but the clinical trial as an experimental setting is to be designed to estimate the estimand. E.g. the intercurrent event should be considered as a matter of clinical practice which is then depicted or estimated in the clinical trial. Since the intercurrent event is taken into account in the definition of an estimand, the clinical trial should represent clinical practice with regard to the occurrence of intercurrent events, which would, e.g., not be possible if the trial is designed such that the intercurrent event of interest would artificially be avoided by design. However, settings where the intercurrent events are artificially introduced by the trial design (e.g. by certain rules for rescue medication) but not measured as an outcome of the trial may be unavoidable (e.g. for ethical reasons). Here, external validity should be discussed further. In general, a distinction should be made between “observed” (or measured) and “designed” intercurrent events. E.g. observed treatment discontinuation as an outcome of the trial captures an aspect of the drug’s consequences to be incorporated in the estimation of its effectiveness, whereas fixed rules for the intake of rescue medication are designed in contrast to rescue medication on demand.
17	The new ICH E9 (R1) addendum on estimands and sensitivity analysis represents a mixture of useful clarifications, trivial explanations (neglecting well-known approaches of evidence-based medicine), and a number of critical issues. I recommend to revise the addendum taking the well-known PICOS approach into account and avoiding estimands which cannot be estimated without a high risk of bias and contradict statistical principles for clinical trials of the ICH E9 guideline. This requires a complete revision of the guideline, because only two of the described strategies (treatment policy, composite) should be used in general as main analysis. The other three strategies (hypothetical, principal stratum, while on treatment) are useful only as possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations. In practice, linking the estimand information to the objective is tricky. For example, the generic examples do not specify what the objective is – please add that to clarify how to link the two The phrasing of an estimand under a given strategy is difficult. Suggest to add examples of how that could be done. The directional description of objective -> estimand -> design in practice is likely much more circular, suggest to reflect that in text, since the graphic illustration suggest it to be linear The addendum targets all confirmatory trials, but endpoints based on survival analysis do not really fit well in the template used. Any indication of how to use it for survival analyses would be appreciated Enrichment designs are often used to address some of key issues in clinical trials; high

Stakeholder no.	General comment (if any)
	placebo response, high withdrawal rates. Enrichment designs are hard to fit into the template provided in the addendum An analysis that aims at testing the robustness like tipping point analyses or placebo mean imputation will not target the same estimand, as the analysis that it is testing the robustness of. This means that a lot of obvious choices for sensitivity analyses will not fulfil the requirement that the sensitivity analysis targets the same estimand. Please clarify whether sensitivity analyses can be targeting other estimands, or at least other assumptions on behaviour after withdrawal. And please provide examples of possible sensitivity analyses that targets the same estimand, to illustrate how close a match is needed on the target estimand. The concept of “conservatism” is not mentioned, suggest to mention that it remains a critical point that analyses are not set up to provide undue advantages for the new drug being tested. It would be very beneficial for the reader if the examples could be a bit more tangible. Please consider to elaborate a bit more on how the five different strategies are to be implemented e.g. regarding missing data imputation. Furthermore, it would ease the introduction of the topic ‘estimands’ if some of the current practices in clinical trials could be translated to estimands. 1. The guideline appears to reflect two different paradigms for conceptualizing an intercurrent event: a. Intercurrent events represent qualitative treatment outcomes of interest. An estimand attempts to summarize what actually happened, both classical and intercurrent events, as the complete treatment outcome including both quantitative and qualitative elements. Improved better methods better incorporate the qualitative intercurrent event information. b. Intercurrent events do not represent treatment outcomes of interest. An estimand ideally attempts to summarize a counterfactual scenario, what would have happened if the intercurrent event had not occurred. Improved methods better adjust for the qualitative intercurrent event information. While both approaches recognize intercurrent events as statistically informative rather than representing statistical noise, they lead to different directions both for study design and for methodological research. Suggest providing more guidance as to which conceptualization might apply in which circumstances. The guideline generally presents intercurrent events as potentially providing positive information. But the available methods to address intercurrent events generally appear to take a more counterfactual perspective. Suggest clearly articulating the goal separately from whether available methods reach that goal. 2. Available counterfactual approaches depend on strong assumptions, so when they are used, post-hoc checks must be made whether these assumptions remain plausible. Accordingly, the guideline very understandably focuses on sensitivity analyses. The goal for counterfactual approaches, however, is where possible more robust counterfactual methods which are less dependent on assumptions and

Stakeholder no.	General comment (if any)
	require fewer sensitivity analyses. Usable guidance must of course help industry use currently available methods today. Nonetheless, part of the purpose of the guidance should be to help inform methodological research of what is needed. To this end, recommend clearly identifying the goal of robust methods, indicating that the goal is not generally supportable with current methods, and indicate that the purpose of sensitivity analysis is not to make them a goal in themselves, but as a necessary adjunct to methods that make strong assumptions, designed to assess the appropriateness of the assumptions made. Recommend the guideline explicitly call on the research community to help develop more robust methods where possible and appropriate. 3. Suggest additional terminology to help clarify and distinguish the goals. As one possible set of terminology in the epistemological tradition, when a counterfactual approach is used, what is observed could be called a phenomenon, while the unobservable, counterfactual estimand of interest might be called a noumenon. Different terminology could be used when intercurrent events are conceived as introducing additional qualitative information to, rather than being counterfactual to, the estimand of interest. 4. Trial design and methods also inform and in some cases can conceal the estimand of interest. Suggest more discussion of trial design and observation methods to reduce intercurrent events and other sources of bias. This includes explicitly evaluating compliance, patient burden, and drop-out related characteristics of methods as part of decisions about what methods to use. a. Suggest discussing general preference for simpler and less intrusive methods that may have better compliance in the context; and considering compliance and patient burden as well as ability to reliably measure an endpoint in setting visit and assessment schedules. b. Suggest discussing discontinuities in observation (e.g. clinic visits or other discrete assessment required to observe endpoint whose analysis assumes continuous observation). Issues involved can include left censoring issues (e.g. event occurs before first scheduled assessment); overestimation bias (longer observation intervals increase overestimation); etc. Dependence of visit schedule on treatment schedule can result in additional confounding (treatment resulting in longer treatment delays may appear more efficacious). Because key elements of the estimand concept, including specifying the method as part of the variable, specifying how bias will be addressed, and appropriate sensitivity analyses, are appropriate to address associated bias, integrating observation methods which introduce confounding into estimand framework and specifying and addressing observation method issues as part of the required specification process would be helpful in introducing greater rigor, reliability, and attention to sources of confounding into clinical research.
18	EUCROF welcomes the opportunity to provide comments on the Addendum to the ICH Guideline E9. It is well appreciated to receive further guidance on statistical matters for clinical trials. This Addendum is a mix of guidance and tutorial. As a consequence, the document is

Stakeholder no.	General comment (if any)
	somewhat lengthy. Compared to the original E9 Guideline covering a good part of existing statistical principles for clinical trials in 35 pages, it seems disproportionate to add another 23 pages on estimands and sensitivity analysis. This may be acceptable but should be taken into consideration. Also, our members felt that it is not easy to understand as – at times – very long sentences are used. It is appreciated that the Addendum provides a different perspective for the planning stage, i.e. not only coming from the trial objective but also accounting for intercurrent events. However, it does not become totally clear whether or not it is expected that future protocols should always be based on primary estimands instead of primary endpoints. This Addendum is mainly related to clinical trials with large sample sizes. EUCROF thinks that more consideration should be put on trials with a smaller sample size. We appreciate the two examples described in the Addendum, however we think it would be also useful to describe a situation in which a small sample is analysed. We are seeing an increasing number of such trials, in particular trials in rare diseases. Protocol deviations: EUCROF would appreciate a clear statement whether a subject with an intercurrent event that represents a major protocol deviation, however had been accounted for during the planning phase, has to be excluded from the Per Protocol Analysis Set or not. Confounding factors and covariates: Adding information related to the analysis of covariates and confounding factors would be very helpful. Missing data and imputation: It would be very much appreciated to receive some sample strategies of imputation. Protocol content: Clarification as to what level of detail should be provided in the protocol and in the Statistical Analysis Plan (SAP), respectively, would be very helpful. For example, while it is understood that the pure fact of intercurrent events must be stated in the protocol, could the more detailed description of all (foreseeable) possible intercurrent events be stipulated in the SAP? Definition of analysis sets: In the context of the comment above, it should be re-emphasized in the Addendum that the analysis sets should be defined already in the protocol and not only in the SAP (which is often the case). What is the difference between confounding variables, covariates and intercurrent events? When reading the text and looking at Figure 1, it seems that an estimand seems like that: Estimand = Main estimator + covariates + confounding factors + error (this error may be related to measurement error, estimate error, etc.). Said so, covariates and confounding factors may be some kind of definition of the intercurrent events. The focus of this Guidance and in particular the Addendum is on Phase III studies. However, it is also applicable to Phases II and IV. While this is appropriate, it would be helpful to give some thought to Phase I trials and to other studies where subjects

Stakeholder no.	General comment (if any)
	(often healthy volunteers) are used as a biological model to address a scientific question. As an example (and there are many more), what is the primary estimand in a TQT study described in ICH E14? One or two sentences mentioning that there may be other situations would be very helpful. Composite endpoint: Composite endpoint, combination of various clinical events that might happen, such as heart attack or death or stroke, is commonly used in cardiovascular trials. The common practice to such data analysis is to assess the time from the randomisation to the occurrence of the first event. This approach is problematic since the endpoint could be dominated by the less important clinical event. No recommendation on such data analysis strategy is made in the draft Addendum. Non-normal data analysis: Non-normal data is ubiquitous in clinical trials. Mann-Whitney test can provide a p-value but does not provide any information on the treatment effect in terms of direction, magnitude and precision. Some discussions and guidelines on this topic would be of help to the practical statisticians. Multiplicity: Co-primary outcomes, multiple dosages, and multiple comparisons occur often in design and analysis of clinical trials. No guidelines are discussed on this topic in terms of trial design, analysis and reporting. Non-inferiority: Non-inferiority design is increasingly used in Phase III trials. Some guidelines on choosing the non-inferiority margins would be helpful.
19	In the generic example of section A.7, in the treatment-policy strategy a worded example in lines 645-647 is offered and lines 648-649 refer to how similar sentences could be constructed for the other examples. It would be helpful to demonstrate what those example sentences actually are for each of the examples too to help ensure that the wording employed by sponsors meets expectations.
20	We welcome the recognition and acknowledgement by ICH that designs and analytic approaches used in current clinical trial practice often have some level of mismatch between quantities of most importance and accurate interpretation. Traditionally, primary analyses had to be seen to satisfy an interpretation of the intention-to-treat principle and were often implemented in a manner which, in the presence of what this addendum calls “intercurrent events”, could lead to a disconnect between analysis results and clinical interpretation. This addendum provides a helpful framework to achieve a better alignment of trial design and results with clinically meaningful quantities. It would be helpful if the intended scope of the document was explicitly stated. For example, should this be viewed as applying to confirmatory studies (because of the emphasis on “regulatory decision-making”), or does it extend across development stages to earlier-phase trials? Does it apply only to parallel-group studies, or also to crossover studies, single-arm trials, non-inferiority or equivalence trials, etc.? Should the estimand framework be restricted to primary and key secondary endpoints, or to other outcomes (safety, PK, biomarkers, etc.)? Clarification would be very helpful. It would be helpful if more explicit linkage were made between the terminology in this document, used in manners which may not be familiar to many readers, relative to

Stakeholder no.	General comment (if any)
	current practices, in terms of concepts such as primary objective, primary analysis, key secondary endpoint, other secondary endpoints, supplemental analyses, exploratory analyses, etc. It's not completely transparent when the document is referring to an estimand related to its primary objective versus when it is referring more broadly to any estimand that is defined within a particular study (the usage of terms such as "main estimator" can also cause confusion in this regard). The inter-relationships of many of the quantities might be clarified graphically, for example, illustrating the relative positions of primary estimands, secondary estimands, main analyses, sensitivity analysis, supplementary analyses, etc. To achieve a precise understanding of the treatment effect, the addendum focuses strongly on the concept of an "effect", but far less on the "treatment" which is the object of the research. It seems that a relevant and explicit space is missing from the framework where the therapeutic intervention being studied can be precisely defined. Clear specification of the treatment is fundamental to the definition of the scientific question. It likely impacts the definition of the relevant population targeted by the therapeutic intervention. It is also key to support a principled approach to the identification of the relevant intercurrent events, and the specification of how to account for them. It may drive the identification of the variable of interest and impact the assessments required to support its measurements. Indirect allusions to the importance of this concept can be inferred within the addendum (e.g., lines 358-360 and 365-376). However, those implications of the treatment definition occur by default, in an implicit rather than explicit manner. If treatment were fully specified, perhaps as a fifth attribute of the estimand definition, placed first in the attribute list (or possibly second, after the population), the scientific question would be clearer and all other attributes could be defined meaningfully in reference to it. The motivation for this is particularly acute when complex therapeutic interventions are being studied. Some examples can be found in hematological oncology indications: • A treatment may consist of an overall strategy involving a complex sequence of interventions, such as induction, consolidation and maintenance treatments, each possibly with different drug combinations, doses and durations, and governed by a specific decision algorithm based on different outcomes. The study of each individual component separately from the rest may not be meaningful or relevant. The population amenable to the entire treatment strategy is different from the population targeted by a single step in the sequence. The treatment strategy may also incorporate other interventions, such as hematopoietic stem cell transplant, based on outcomes at different stages in the sequence. Such interventions could still legitimately be part of the overall treatment strategy (e.g., resulting from success of the previous steps and offering a greater clinical benefit to patients under certain conditions) and therefore not be considered as intercurrent events per se, even if their impact on outcome is possibly of a very different nature. • The definition of the comparator treatment may similarly be critical. For example, a simpler novel therapeutic approach may be proposed to replace a

Stakeholder no.	General comment (if any)
	complex standard of care strategy. The single administration of a cellular immunotherapy treatment may be considered to replace a sequence of several lines of standard drug-based treatments, the sequence of which follows a defined algorithm based on outcomes. The implications for the definition of the scientific question, e.g., on the actual comparison of interest, are very important. They govern the definition of the variable (e.g. relapse/progression after a given time point or after a sequence of events), the required assessments and the intercurrent events, which may have a different impact by treatment group. The tone of the document seems to indicate that there will always be one particular "effect" that can be identified to be "of interest" in a trial (for regulatory decision-making, for prescribers and patients, etc.), and that other versions are therefore of lesser interest. But for an effective treatment, different effects quantified using different ways of accounting for intercurrent events usually are just slightly different characterizations of the effectiveness. The document could make clearer, perhaps through specific examples, why and when one definition could be so much more important than others, particularly when its analysis and description requires unverifiable assumptions and thereby puts interpretation at some risk. Clinicians would play an important role in implementation of the principles and strategies described, so should be part of the target audience, but in its current form the document may not be very clear to clinicians. Clinicians would be able to more readily understand the differences between various estimand strategies if actual clinical research examples were used. In other words, the examples should be less generic and describe an actual clinical question referring to a specific type of patient, a specific disease or condition being studied, and a specific outcome measure being used for each estimand, to make these concepts more vivid to the broader audience. Although the guidance already broadly states in generic language why these different estimands might be used, it will be difficult for clinicians to grasp this unless actual real-world examples are provided for each type of estimand used to answer a specific research question. The principal stratum strategy is very challenging to comprehend, from a number of standpoints. For example, what are the clinical and regulatory considerations in using this strategy in actual practice as we often cannot know in advance which patients belong to the principal stratum? Further, is it acceptable to exclude patients who cannot tolerate or adhere to a *different* treatment (i.e., the comparator in a clinical trial)? Wouldn't patients who tolerate the investigational treatment, but may not tolerate a control treatment, almost always be part of any population of interest? Also, please note that the glossary defines this term in a manner that conflicts with the way it is described elsewhere in the document. It seems important that this strategy be better explained, motivated, and justified, probably through explicit examples. The document seems to consciously avoid use of familiar statistical terminology, such as hypothesis tests, parameters, covariates, analysis models, etc. We would expect that in implementing the strategies described, designs and analyses must follow a familiar hypothesis testing framework (e.g., sizing a study to have desired power to reject a null hypothesis under an assumed alternative, for a parameter value reflecting

Stakeholder no.	General comment (if any)
	an effect size of importance, and using a specified analysis model). Since much of the document is quite complex, it would clarify the meaning for readers if the linkage to standard hypothesis testing terminology was made explicit. It is recommended that the document more explicitly make a connection to analysis set terminology that is currently in common use (e.g., ITT, Full Analysis Set, PPS, etc.), and point out any distinctions. It's not clear whether this document extends ICH-E9 in this regard, or conflicts with it; if there are conflicts, then it seems important for this addendum to clarify any aspects of E9 which are no longer applicable. Since a given estimand strategy might require unverifiable assumptions, the document sensibly discusses the need for sensitivity analyses to investigate robustness. But there can never be certainty that these will indeed demonstrate robustness. Since a broad set of sensitivity analyses may be needed for a primary analysis due to the unverifiable assumptions, could the framework of this addendum put sponsors at added risk regarding regulatory interpretation, in case robustness is not shown? Might the document explicitly acknowledge this possibility (perhaps even as a factor that could at times influence the choice of the primary estimand)?
21	EFSPI/PSI agrees on the importance of clarifying objectives, assumptions and considering how to deal with intercurrent events during the planning and analysis stages of a trial. However, there are several areas where the Addendum can be improved and does not go into enough detail. The impact of introducing the Addendum without clear guidance on how to implement the framework in practice has not been considered and is a serious concern. The main issues to be addressed for a successfully implementation are outlined below 1. We acknowledge and agree that guidelines need to outline principles. However, there is a need for more case studies to support this addendum. It is proposed that the generic examples in section A.7.A are supplemented by a companion of case studies. The case studies should be a detailed description of a real situation where different strategies for handling intercurrent events are described and the pros and cons of each approach to address them as it relates to the question of interest are articulated. However, it is understood that the drafting group plans to address the lack of examples in the main text by producing case studies in slides that will accompany the Addendum. There are several issues with this approach. Firstly, it is understood that these case studies will be made available shortly, but the final version of the Addendum is unlikely to be completed until the end of 2019. Also, there is no consultation process planned for commenting on the case studies and hence if they could be improved upon there is no opportunity for this to happen, even though there is probably over a year available to do this. It is therefore strongly encouraged that this process is changed to allow comments on the case studies if the Drafting group does not include extensive examples in the main text of the Addendum. 2. The addendum clearly impacts existing disease specific guidance. Regulatory authorities should produce a prioritised list of disease specific guidance documents to revise in light of the Addendum. Given the Addendum will not be

Stakeholder no.	General comment (if any)
	finalised until 2019 work should start on revising disease specific guidance documents as a matter of priority to provide better guidance on estimand strategies that may be acceptable in specific disease areas/indications. What plans are there for review of these guidelines once the addendum is issued? 3. It is proposed the addendum explicitly highlights that there may be a difference between estimands used for hypothesis testing e.g. ones based on the randomisation with minimum assumptions and estimands used to quantify clinical benefit. 4. It is very important that how intercurrent events have been handled is transparently described in the drug label/SmPC . Please clarify in the Addendum how estimands will link to drug label/SmPC, in particular when a regulatory authority bases its decision on a different Estimand strategy to that pre defined by the Sponsor. 5. The addendum revisits the meaning and role of the analysis populations as outlined in ICH E9; in particular and role of the per-protocol analysis. It is a proposed that a section is added to the addendum that summarises the key changes to ICH E9 and those sections in ICH E9 that no longer apply. That is, clarity can be improved by summarising in one subsection the differences to ICH E9 in the use of ITT and PP analyses as outlined in the addendum. 6. Please clarify whether study discontinuation is an intercurrent event or missing data problem. The Addendum is ambiguous on this issue, and feedback from scientific meetings has not provided clarity. Similarly please can you clarify how death should be handled. 7. The Addendum suggests a Treatment Policy strategy cannot be applied for intercurrent events of death. For a study when there are a small number of deaths “unrelated” to disease/treatment then, please clarify what the appropriate strategy should be? 8. For non-inferiority/equivalence analyses, neither the appropriate strategy for handling intercurrent events nor the appropriate analysis set is specified, and in the case of analysis sets the document seems to contradict ICH E9. Because the inclusion of off-treatment data is likely to bias in favour of no difference between treatments, it seems that the treatment policy strategy and full analysis set are not appropriate in this setting. It is proposed that the addendum explicitly discusses non inferiority trials. A worked example for a non-inferiority trial in the proposed companion of case studies discussed in point 1 above would be helpful. 9. A worked example of an appropriate sensitivity analysis in the proposed companion of case studies discussed in point 1 above would be helpful . 10. Safety estimands: ICH E9 provides guidance specific to the evaluation of safety and tolerability (Section VI). It is proposed this topic is also addressed in the addendum. For example -are there strategies to handle intercurrent events specifically suited for safety

Stakeholder no.	General comment (if any)
	analyses (e.g. 'while on treatment')? -impact of handling of intercurrent events as treatment withdrawals or deaths to determination of number of subjects at risk and estimation of incidence, -should we analyse safety parameters 'as treated' or 'as assigned' Moreover, consideration should be given to the possibility that efficacy and safety estimands may use different strategies to handle intercurrent events and this has implications for how benefit risk evaluation is performed. 11. The addendum seems to be based on parallel-group phase III designs. It should be made clear in the introduction that this is the main focus of the Addendum. As well as points 5 and 7 above, some statements should be added to address other settings such as cross over studies, early phase setting, where treatment is given only for a short time like e.g. the conditioning therapy prior to stem cell transplantation in oncology, Bayesian frame work etc . Worked examples for a variety of scenarios in the proposed companion of case studies discussed in point 1 above would be helpful. Existing analytical methods for sample size calculations may not be applicable for some estimation analyses. Therefore, simulation needs to be performed to understand the operating characteristics of various strategies to handle intercurrent events. It is recommended that the Addendum includes a short section highlighting this issue.

2. Specific comments on text

Line no.	Stakeholder no.	Comment and rationale; proposed changes
40	3	**Comments:** Consider re-wording of the first part of the sentence. 'To properly inform the choices that are made by patients…' **Proposed change:** To allow for properly informed choices of patients and their prescribing physicians….
40	8	**Comments:** In Purpose and Scope, the opening sentence that says in order to inform choices for patients and physicians, clear descriptions of the effects of a medicine should be available. However, in the rest of the document it is apparent that clarity of the treatment effect and choice of estimand is to enhance the regulatory/sponsor dialogue. It is not apparent that estimands are primarily chosen to meet the specific need of patients and physicians, and in many cases the estimands of interest from physicians will not be the focus of regulators and are not guaranteed to be included in a label. e.g. line 335: 'Estimands based on the treatment policy strategy might also be more generally acceptable to support regulatory decision making, specifically in settings where estimands based on alternative strategies might be considered of greater clinical interest….' **Proposed change:** I would suggest that the first sentence indicates that the primary reason for ICH to propose a framework for treatment effects to be more precisely specified, is to facilitate discussion between sponsor and regulator. The focus on physicians/patients does appear to have a secondary priority (in this framework), as it is mentioned that these alternative estimands can be supplied as additional estimands, but there is no guarantee that these will be included in a label.
49-50	21	**Comments:** This is the only place in this document where the stage of clinical trial is mentioned. Does this imply that the addendum applies to confirmatory clinical trials only? **Proposed change:** Clarify whether the scope of the addendum is limited to confirmatory trials.
51	8	**Comments:** Clarify that sample needs to be large to have randomized trials free from baseline confounding

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: *Randomised trials are expected to be free from baseline confounding (for large sample)*
51	17	**Comments:** The randomisation could also be said to introduce causality by design, by keeping everything, but the treatment, similar in the two groups. If the trial is blinded this will be the case, also during the trial. Consider to include a sentence consider this causality by randomisation
51	21	**Comments:** The reference to clinical practice may need more explanation, suggest omit. Also suggest make clear that intercurrent events occur post randomisation. **Proposed change:** Randomised trials are expected to be free from baseline confounders, but in practice certain events will occur post randomisation that complicate … **Comments:** The randomisation could also be said to introduce causality by design, by keeping everything, but the treatment, similar in the two groups. If the trial is blinded this will be the case, also during the trial. **Proposed change:** Consider to include a sentence consider this causality by randomisation “Randomised trials are expected to be free from baseline confounding but…” **Comments:** The definition is incorrect. **Proposed change:** Change to “Randomised trials are designed to minimize the effect of confounding factors”
53 and thereafter including glossary	8	**Comments:** Multiple dictionary definitions of “intercurrent” typically restrict the usage to reference concurrent medical complications arising from a separate disease process. This is also the sense in which most clinicians understand the term. Using the term “post-randomization” rather than “intercurrent” is arguably more precise. **Proposed change:** *Substitute “post-randomization” for “intercurrent”.*
53, + rest of	15	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
text		newly introduced terms under apostrophes? **Proposed change:** " … these are denoted as 'intercurrent events' (see Glossary) …"
53	21	**Comments:** Here it refers the reader to the Glossary for the definition of 'intercurrent events'. Actually the Glossary just repeats what is in the previous sentence; albeit using slightly different wording. **Proposed change:** Omit reference to Glossary and use glossary wording in the main text.
54-55	15	**Comments:** examples for intercurrent event mentioned include: 'use of an alternative treatment' and 'treatment switching'; it might not be needed to mention both, as similar/the same event is meant?! **Proposed change:** one of those could be omitted, in particular in an introduction
56	15	**Comments:** "terminal events" is not immediately clear **Proposed change:** suggest rewording or clarifying what "terminal" entails
56	21	**Comments:** "…terminal events such as, in some circumstances, death." Death is always a terminal event, not only in certain circumstances. Also, death seems to be an intercurrent event in all situations as no data can be collected after its occurrence. **Proposed change:** Consider to delete "(…) in some circumstances (…)" Change to "…terminal events such as death."
57	8	**Comments:** Consider providing guidelines on how to define safety variables in the presence of intermittent events.
57	21	**Comments:** This is the only instance in the addendum where "safety" is mentioned in the context of variables, or in a broader sense "estimands". It might be worthwhile to specify whether the considerations in the addendum are meant to apply to efficacy questions only (the addendum is mainly talking about

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		"treatment effect", which is typically used in connection with efficacy) or whether the considerations equally well apply to safety questions.
58	12	**Comments:** ACRO recommends the following addition, for consistency and clarity. **Proposed change:** Add "(see Glossary)" after … intercurrent events, as done elsewhere in the document (e.g., line 70).
58 60-61 144 280-281	16	**Comments:** The wording "accounting explicitly for intercurrent events" (or similar wording in other places) may be misunderstood or misleading considering that the treatment policy estimand ignores intercurrent events (assuming that data after the intercurrent event has been collected). For example, in line 203-204 it is written: "Treatment policy strategy: The occurrence of the intercurrent event is irrelevant", i.e. here we would not 'account' for intercurrent events, or at least the expression 'account for' is confusing. **Proposed change:** Explain that "accounting explicitly for intercurrent events" or similar wordings include the use of a treatment policy estimands which is estimated using data collected after the intercurrent event.
59-60	3	**Comments:** Consider to add 'to' in the sentence to make it easier to read **Proposed change:** '…to ambiguity about the treatment effect to be estimated and <u>to</u> potential misalignment with trial 59 objectives.'
60	12	**Comments:** ACRO recommends replacing "The correct order is the reverse" (which fits strangely in the sentence) with the text proposed below. **Proposed change:** Replace the current test with "Therefore, the treatment effect to be estimated and the impact of intercurrent events should be considered prior to defining the efficacy and safety variables".
60	15	**Comments:** The sentence "The correct order is the reverse." reads quite prescriptive. **Proposed change:** "A more logical order is the reverse."

Line no.	Stakeholder no.	Comment and rationale; proposed changes
65	8	**Comments:** Add 'i.e.' before "translating the trial objective into a precise definition of the treatment effect that is to be estimated", as this is what the term 'estimand' describes. **Proposed change:** i.e. translating the trial objective into a precise definition of the treatment effect that is to be estimated.
65	21	**Comments:** Add 'i.e.' before "translating the trial objective into a precise definition of the treatment effect that is to be estimated", as this is what the term 'estimand' describes.
67	8	**Comments:** Add 'design' into the list of activities that the various disciplines are involved in, as trial design could encompass the clinical, scientific and statistical design of the trial, whereas trial planning could be considered to be a separate activity, which might be only be considered to be the operational planning of the trial. 'Design' is also in the title of section A2, so add here for consistency. **Proposed change:** trial planning, design, conduct, analysis and interpretation
70	16	**Comments:** Rather "sensitivity analyses"? In general, the use of the grammatical number of sensitivity analysis in the whole document appears confusing.
70-71	21	**Comments:** "This addendum clarifies the definition and the role of sensitivity analysis." The addendum should clarify the need for clear specification of the estimand and the criticality and role of sensitivity analyses. **Proposed change:** It should be clarified what is the change introduced by the estimand concept with regards to sensitivity analyses.
71-72	9	**Comments:** There are a few references to ICH E9 and they occur much later (the existence of this sentence could be forgotten by then, making e.g. what "Section 3.3.2" referred to somewhat bewildering). It would be better if ICH E9 is explicitly mentioned when it is being referenced (see 3rd paragraph on page 11, for example)

Line no.	Stakeholder no.	Comment and rationale; proposed changes
74-77	5	**Comments:** It is not clear if the meaning is exclusion of measurements due to intercurrent events or regardless. If the former the sentence should be modified to reflect this. If not then the meaning is not clear.
74-79 88-89 208-209	16	**Comments:** The intention to treat (ITT) principle is mentioned three times in the addendum and always in relation to the treatment policy strategy. However, the ITT principle is a more general concept not restricted to or defined by assessment of treatment policy. It simply defines a gold standard for the analysis population that best preserved benefits of randomization by including all randomized patients according to randomized allocation. Given appropriate trial designs any of the discussed strategies can and should be evaluated based on this analysis population. For a general discussion on our view on ITT, see also the general comment above. **Proposed change:** ITT should be defined as “including all randomized patients according to randomized allocation” and it should not be restricted to the treatment policy strategy only. This would allow ITT to remain the gold standard of analysing RCTs, would not restrict ITT conform trial objectives and would not require a new name for the fundamental principle ‘all randomized as randomized’.
74-109	15	**Comments:** This section is too long **Proposed change:** Proposal to shorten
78	20	**Comments:** Suggest changing “aim at exploiting” to “make use of” **Proposed change:** It remains undisputed that randomisation is a cornerstone of controlled clinical trials and that analysis should ~~aim at exploiting~~ ***make use of*** the advantages of randomisation to the greatest extent possible.
79	8	**Comments:** Add ‘primarily’ in front of ‘understanding the effect of treatment policy always targets the treatment effect of greatest relevance’, as there can be different treatment effects estimated, but the intent here is that treatment policy may not the most appropriate primary effect to estimate. **Proposed change:** the question remains whether <u>primarily</u> understanding the effect of a treatment policy always targets the treatment effect of greatest relevance

Line no.	Stakeholder no.	Comment and rationale; proposed changes
79-81	2	**Comments:** The following sentence is quite unclear: 'However, the question remains whether understanding the effect of a treatment policy always targets the treatment effect of greatest relevance to regulatory and clinical decision making.' Although a native English Speaker I cannot resolve what it is trying to communicate. What is the question here? **Proposed change:** reword sentence
79-81	21	**Comments:** "However, the question remains whether understanding the effect of a treatment policy always targets the treatment effect of greatest relevance to regulatory and clinical decision making." Comments: The sentence does not sound clear. Proposed change: Suggest to remove the word "understanding". Change to "However, the question remains whether the effect of a treatment policy always targets the treatment effect of greatest relevance to regulatory and clinical decision making."
82	12	**Comments:** ACRO recommends the following addition, for consistency and clarity. **Proposed change:** Add "(see Glossary)" after … intercurrent events, as done elsewhere in the document (e.g., line 70).
85-89	21	**Comments:** Lines 85-95 from "On one hand…" repeats material covered in e.g., lines 236ff.
87	16	**Comments:** The intercurrent events are irrelevant or difficult to interpret only if the intercurrent events do not reflect clinical practice (i.e. are artificially introduced by the trial). A distinction should be made between "observed" and "designed" intercurrent events. **Proposed change:** Add a paragraph on the external validity of a clinical trial setting with respect to the occurrence of intercurrent event.
87	21	**Proposed change:** "It" should be "they."
87-88	9	**Comments:** Suggest deleting the sentence "In the case of death, measurements after a

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		subject dies do not exist."
88-89	8	**Comments:** Asserts the benefit for analyses to continue collecting data post discontinuation rather than default to lost-to-follow-up by design. I know that the ethics topic is in a different guidance (E6, 4.8), where subjects' "willingness to continue participation in the trial" may change, and I believe should be able to change without judgement or persuasion (E6, 4.8.2-3). I wonder if anyone has already addressed the ethical concern, that upon discontinuation subjects' have every right to refuse further participation (I believe that every informed consent that I've read makes that statement)? The pros/cons of further participation should be considered (continued medical care vs continuing phlebotomy), because the baseline assumption that subject's best interest is to continue post discontinuation needs to be questioned, especially in the context of the topic: Estimands.
88-89	16	**Comments:** Lines 88/89 state the difficulty to fulfil the ITT principle in case data are missing. However, this is not a problem of fulfilling ITT but of reliably estimating treatment policy. **Proposed change:** The issue of missing data should be clearly separated from the ITT principle and from the handling of intercurrent events. Handling of missing data is a problem of reliably estimating a specific estimand, but not a requirement of the ITT principle. Avoiding a missing data problem by having complete follow up irrespective of intercurrent events might have let to the understanding of ITT as addressing treatment policy, but ITT is a much more general principle.
91	16	**Comments:** The "Purpose and Scope" section contains a central message of the document stating "This addendum invites consideration of the important distinction between non-adherence with, or withdrawal from, randomised treatment and discontinuation from the trial". However, in the following the document does not make use of the term discontinuation in the sense of "discontinuation from the trial" any more. In fact, a proper definition of a term for a patient withdrawing from the trial is missing ("subject withdrawals" is only used once in the context of sample size estimation in line 434). The ICHE9 Glossary term "dropout" is not used throughout the addendum, although a term is definitely needed due to the patient's right to withdraw from the trial at any time. The document could benefit from a clear definition for terms used to describe discontinuations/withdrawals/drop-outs from the trial in contrast to discontinuation of randomized treatment. Discussions with sponsors frequently showed that the terms "discontinuation" and "withdrawal" are used

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		for treatment discontinuation as well as discontinuation from the trial and are often used in an ambiguous way. Clear language is even more important as discussions on estimands using a hypothetical strategy showed that an intercurrent event such as a treatment discontinuation is often considered as simultaneously leading to a discontinuation from the trial. Intercurrent events are often described from the perspective of the missing data problem using ambiguous language. **Proposed change:** Introduce clear terms for study discontinuations/withdrawals/drop-outs in the sense of the ICH E9 Glossary term “dropout” and treatment discontinuations and emphasize the importance to distinguish between these.
91-92	15	**Comments:** it is suggested to phrase a complete sentence **Proposed change:** “ … from the trial. Another relevant distinction is the one between measurements that exist but have not been collected, and measurements that do not, or cannot, exist.”
91-92 163	16	**Comments:** Is “measurements that exist but have not been collected” the optimal wording or shouldn’t it rather be “existing values that have not been measured and recorded” ? Apparently “measurement” may be used for non-observed values (blood pressure whether recorded or not), but, e.g., blood pressure *measurements* would be related to the sampling (i.e. the trial) but not necessarily to the variable of interest itself. At least non-native speakers may struggle. **Proposed change:** Explain measurement in parentheses (existing values, whether recorded or not)
91-92	21	**Comments:** “also between measurements that exist but have not been collected, and measurements that do not, or cannot, exist”. **Comments:** The text should more clearly delineate between "cannot" and "do not". "Do not" relates to unmeasured but measurable. "Cannot" is unmeasurable (e.g. post death). **Proposed change:** Change to “also between measurements that exist but have not been collected, and measurements that do not exist (ie unmeasured), or cannot

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		exist (ie unmeasurable)"
94	12	**Comments:** ACRO recommends amending the statement "…..collected, present a missing data problem to be addressed. In turn methods to address…..." to read as follows. **Proposed change:** Replace the text with "…..collected, leads to a missing data problem that needs to be addressed. In turn, methods to address…..."
96	9	**Comments:** Suggest adding "of ICH E9" **Proposed change:** "Section 5.2" -> "Section 5.2 of ICH E9"
96	21	**Comments:** Despite the statement that "Thirdly, the concept of analysis sets is considered in the proposed framework," the addendum discusses the role of the per protocol analysis set only. Would it be useful to add a small section discussing analysis sets generally? How should an analysis set be defined where, for example, data after initiation of rescue medication are ignored and, in some cases, imputed? Section 5.2 of E9 refers to the analysis set as "the set of subjects whose data are to be included.." but the draft addendum makes clear that, in some cases, not all data on a subject are relevant for every estimand strategy. Should terminology be introduced for this situation – e.g. "modified analysis set" (see Line 840).
110	8	**Comments:** Clarify the robustness context in the document **Proposed change:** *Finally, the concept <u>of robustness of inferences</u>, ….*
110	21	**Comments:** Clarify the robustness context in the document **Proposed change:** Finally, the concept of robustnesss of inferences, ….
111-113	15	**Comments:** The meaning of this sentence is difficult to grasp. **Proposed change:** reword or delete?

Line no.	Stakeholder no.	Comment and rationale; proposed changes
111-113	21	**Comments:** Sensitivity analysis is used in two contexts; is there a better way to differentiate by introducing the terminology of supplementary analysis used later? **Proposed change:** To show how this guidance aligns with supplementary analysis, amend “In particular, a distinction is made between the sensitivity of inference to the particular assumptions of a particular analysis and the sensitivity to the choice of analytic approach more broadly.” To “In particular, a distinction is made between the sensitivity of inference to the particular assumptions of a particular estimator and supplementary analyses which investigate sensitivity to the choice of analytic approach more broadly.”
113-116	12	**Comments:** For clarity, ACRO recommends replacing the last sentence of the paragraph as follows. **Proposed change:** Replace the sentence with “With a precise specification of an estimand and with a pre-specified statistical analysis defined to a level that it can be replicated precisely, and that is aligned to the estimand, then regulatory interest can focus on sensitivity to deviations from assumptions and limitations in the data in respect of a particular analysis.”
113-116	15	**Comments:** In the overall context of decision-making, it is questioned if the main focus of “regulatory interest” should indeed rest on the robustness of analyses when applying varying assumptions. **Proposed change:** suggest rephrasing to highlight it as *one* and not *the* focus of regulatory interest
116	9	**Comments:** Change ‘of’ to ‘to’ **Proposed change:** ‘in respect of’ -> ‘in respect to’
119 - 120	15	**Comments:** To our understanding, formulating the “key scientific question” would be the first step in development, from which the trial objectives will be derived (not the other way around).

Line no.	Stakeholder no.	Comment and rationale; proposed changes
119-123	7	**Comments:** It is trivial that a clear scientific question is required before parameters are estimated. The well-known PICOS approach (participants, interventions, comparators, outcomes, and study design) should be taken into account. The given series of items on the one hand goes beyond the PICOS approach (handling of intercurrent events and specification of the effect measure), but on the other hand is incomplete (intervention and comparator are missing). **Proposed change:** The given series of items should build on the well-known PICOS approach with appropriate additions.
119-123	17	**Comments:** It is trivial that a clear scientific question is required before parameters are estimated. The well-known PICOS approach (participants, interventions, comparators, outcomes, and study design) should be taken into account. The given series of items goes on one hand beyond the PICOS approach (handling of intercurrent events and specification of the effect measure), but is incomplete on the other hand (intervention and comparator is missing). **Proposed change:** The given series of items should build on the well-known PICOS approach with appropriate additions.
119–123 151-157	13	**Comments:** Why are intervention and comparator not mentioned in the list of attributes defining an estimand as it is usually required when to fully describe a clinical study for answering a scientific question (e.g. well known PICOS approach)? **Proposed change:** Include intervention and comparator in the definition of an estimand.
120	17	**Comments:** What is meant by clear trial objectives? Should there be a one-to-one correspondence between the objective and the estimand? Or could there be several estimands addressing the same objective? **Proposed change:** Consider to include more guidance concerning this and/or update figure 1 with more estimands addressing the trial objective if relevant. This could also be included in the example at page 16.
120	21	**Comments:** What is meant by clear trial objectives? Should there be a one-to-one correspondence between the objective and the estimand? Or could there be several estimands addressing the same objective?

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Consider to include more guidance concerning this and/or update figure 1 with more estimands addressing the trial objective if relevant. This could also be included in the example at page 16.
122	21	**Comments:** Please substitute “handling” with “accounting for” for consistency across the document, since “handling” could be understood in the context of the estimator rather than the estimand. To make the difference between estimand and estimators/ estimates very clear, throughout the whole document, in the context of intercurrent events, the same wording should be used throughout the document. **Proposed change:** …the population, the variable, the strategy for intercurrent events…
123	18	**Comments:** The term “population-level summary” should be explained by adding it in the glossary.
124	8	**Comments:** It would be helpful to have a simple example of what an estimator could be (either here or within the examples in section A7)
124	20	**Comments:** This is the first reference to the term “main estimator”. While it seems that this is just a quantity that can be tied to any estimand, it might be good to explicitly mention that the word “main” does not imply linkage to a study’s primary objective, as this could be a source of confusion later in the document.
124-125	21	**Comments:** “The main estimator will be underpinned by certain assumptions.” Suggest adding an example for easy reading **Proposed change:** Change to “The main estimator will be underpinned by certain assumptions, e.g., no treatment crossover.”
126	12	**Comments:** There is a missing “the” at the end of the line. **Proposed change:** “… should be conducted in **the** …”
126-	9	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
127		Insert 'the' **Proposed change:** 'in form' -> 'in the form'
126-127	17	**Comments:** Please clarify how tipping point analyses that target estimates that deviates from a model by varying measures can still be used as sensitivity analyses under the set-up in Figure 1
128	20	**Comments:** It's unclear whether the intention is that Figure 1 should be viewed as applying to the primary estimand in a study, or additionally to estimands associated with secondary objectives, or possibly all relevant estimands included in a study analysis plan. Please clarify these scope-related issues where relevant within the document.
128/Figure 1	21	**Comments:** should the arrows to intercurrent events originate from "estimand" rather than "estimator"?
128-129	9	**Comments:** Although trial objective defines estimand, and estimand defines estimator as shown in Figure 1, there may be no appropriate estimator for the objective and estimand, and the estimand needs to be re-defined. Suggest adding an 'arrow' from estimator to estimand to illustrate this thought process. **Proposed change:** Adding an arrow from estimator to estimand in Figure 1
129	6	**Comments:** A punctuation mark is missing. **Proposed change:** "(…) given trial objective."
130-131	8	**Comments:** … distinguishes between the target of estimation (trial objective, estimand)…: Move trial objectives out of the parenthesis. **Proposed change:** *distinguishes between <u>the trial objective</u>, the target of estimation (~~trial objective,~~ estimand),*
130-131	21	**Proposed change:** … distinguishes between the target of estimation (trial objective, estimand)…:

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Move trial objectives out of the parenthesis. **Proposed change:** distinguishes between the trial objective, the target of estimation (trial objective, estimand),
130-133	21	**Comments:** "This will assist sponsors in planning trials, regulators in their reviews, and will enhance the interactions between these parties when discussing the suitability of clinical trial designs, and the interpretation of clinical trial results, to support drug licensing." This is already mentioned in Lines 66-68 and 82-83. As stated in general comment please remove the redundancies. **Proposed change:** Suggest deleting the sentence.
132-134	8	**Comments:** Add that a clear definition of estimand will allow an upfront understanding on how the data will be analysed and interpret *……when discussing the suitability of 133 clinical trial designs, and the interpretation of clinical trial results, to support drug licensing, prior to the completion of the study.*
132-134	21	**Proposed change:** Add that a clear definition of estimand will allow an upfront understanding on how the data will be analysed and interpreted ……when discussing the suitability of clinical trial designs, and the interpretation of clinical trial results, to support drug licensing, prior to the completion of the study.
135	21	**Comments:** Word missing **Proposed change:** In general, it is important to proceed sequentially and not allow for the choice of an estimator to determine the estimand,… **Comments:** It is not clear why there is an importance on proceeding sequentially. In practice, the real issue is handing of inter-current events and missing data. Addressing those in the addendum will then make for better estimation. **Proposed change:** Consider adding a mechanism to loop back through the framework if

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		estimation issues lead to a re-considering the estimand.
135-136	8	**Comments:** Last part of sentence sounds a bit unclear. **Proposed change:** Reorder and minor rewrite perhaps. "In order to address the scientific question, it is important to proceed sequentially, and not have the estimator determine the estimand." **Comments:** Adding to the above comment **Proposed change:** *Reorder and minor rewrite perhaps.* *"In order to address the scientific question, it is important to proceed sequentially, as depicted in Figure 1, , and not have the estimator determine the estimand."*
135-136	12	**Comments:** For clarity, ACRO recommends deleting the phrase "hence the" from the sentence. **Proposed change:** Delete "hence the" from the sentence.
135-136	18	**Comments:** Meaning not really clear. Please rephrase.
135 - 136	20	**Comments:** Suggestion to delete, or perhaps broadening, this sentence. In practice, it is often difficult to frame a scientific question of interest without some concurrent consideration of quantitative and analytical aspects. This also conflicts a bit with other parts of the document (e.g., lines 335-342).
135-136	21	**Comments:** While this may be true due to the words "in general," it could be misleading by not acknowledging that there often will be times where consideration of whether there is an estimator that could lead to a reliable estimate may well feed back on the choice of estimands. This topic was discussed well in Lines 297-308 and 335-338 and the guidance in Line 297 that "an iterative process may be required" is almost in conflict with the guidance in Line 135 that "in general, it is important to proceed sequentially". It is recommended that the sentence be followed by a clarification statement. **Proposed change:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		"In general, it is important to proceed sequentially, and not for the choice of an estimator to determine the estimand, and hence the scientific question that is being addressed. However, there are times when the estimand best reflecting the desired trial objectives cannot be reliably or robustly (i.e. without questionable assumptions) estimated by any feasible design and estimator, and in such cases, alternative estimands that may also address critical, related regulatory questions should be considered."
141	17	**Comments:** The treatment effect described here, as the counterfactual effect of a treatment given compared to when the treatment is denied, to a subject – how does this link to the five strategies described later? For example the treatment policy estimand seems to target an effect of being randomised to treatment -rather than the above described. **Proposed change:** Consider to describe how the five strategies can be said to help estimating the described treatment effect or why it is not the aim of the estimand
141	21	**Comments:** The treatment effect described here, as the counterfactual effect of a treatment given compared to when the treatment is denied, to a subject – how does this link to the five strategies described later? For example the treatment policy estimand seems to target an effect of being randomised to treatment -rather than the above described. **Proposed change:** Consider to describe how the five strategies can be said to help estimating the described treatment effect or why it is not the aim of the estimand
141-144	21	**Comments:** Restructure the second sentence (after the colon) for better readability. **Proposed change:** A central question for drug development and licensing is to quantify treatment effects. In a specific trial this may come down to the question of how the outcome of treatment compares to what would have happened to the same subjects under different treatment conditions (e.g. had they not received the treatment or had they received a different treatment).
144	21	**Comments:** Please use "definition" instead of "description" for clarification in the context of estimands. **Proposed change:** Intercurrent events need to be considered in the description definition of a

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		treatment effect…"
144-145	21	**Comments:** Intercurrent events do impact a "variable of interest" (which resides in the sample); however, it is more relevant to the discussion to note that these events—or, more precisely, the underlying processes generating the events—relate to the estimand (which resides in the population). **Proposed change:** Change "variable" to "estimand" or "population quantity to be estimated."
147	15	**Comments:** Could this consideration on relevance in some cases also extend to values of the variable *before* the intercurrent event (e.g. when applying a principal stratum strategy)?
151	17	**Comments:** The figure with four bubbles used by several presenters from the addendum group, could be used to illustrate the list of the four attributes to the estimand.
151, 159-160	21	**Comments:** Clarification is required on how to describe the population based on the inclusion/exclusion criteria i.e. whether this should be the full list of inclusion/exclusion criteria, a cross-reference to the relevant section of the protocol containing inclusion/exclusion criteria or some other appropriate summary.
151-153	15	**Comments:** The listing of estimand components implies four (independent) dimensions along which an estimand may be constructed. Is this true or are there dependencies between domains?
151-153, 159 …	15	**Comments:** When giving the general definitions, there needs to be reconsideration if we are talking about 'patients' in all those trials for which we define an estimand, and for which this addendum is generally applicable; think about e.g. vaccines trials were healthy subjects will be recruited, and the question what needs to be estimated is very relevant; **Proposed change:** use broader terms in definitions, throughout the whole document
151-157	7	**Comments:** : In the given series of items A to D the important items "intervention" and "comparator" are missing.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Please add the items “intervention” and “comparator” to the described items A to D.
151-157	11	**Comments:** The four components of the estimand that are provided miss arguably the most important one; the treatment regimen that is of interest / being compared. Without it, you cannot relate the estimand directly to a clinical question. Once this is defined, handling of intercurrent events becomes much clearer since they will all either be ‘part’ of the treatment regimen of interest, and hence ignorable (i.e. ‘treatment policy’), or a break from it, in which case they are either outcomes (i.e. ‘composite’) or confounding (i.e. ‘hypothetical’/’while on treatment’/’principle stratum’). This component would therefore help provide much needed clarity in defining more complex estimands, as well as providing a guiding principle to the handling of unexpected intercurrent event types that occur. More generally, the lack of clarity in the estimand definition regarding what treatment regimens are actually being compared may be the cause of much of the confusion and misunderstanding around how to define estimands in practice.
151-157	17	**Comments:** In the given series of items A to D the important items intervention and comparator are missing. **Proposed change:** Add the items intervention and comparator to the described items A to D.
151-157	21	**Comments:** “D. the population-level summary for the variable which provides, as required, a basis for a comparison between treatment conditions” Although trial level population is treatment specific free, the treatment level population is specific to, say, a randomized treatment, which is especially relevant for D. **Proposed change:** Replace “population“ by “patient set” Change to “D. the Patient-set-level summary for the variable which provides, as required, a basis for a comparison between treatment conditions”
151-159	15	**Comments:** It is unclear which population is meant here. The wording under point A implies the target population. The paragraph starting with l.159 writes that the population is reflected by in-and exclusion criteria (i.e. the study population). Whereas the second sentence of the paragraph refers to what

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		could be understood as the analysis population. To our understanding, an integral part of the estimand concept is to seek alignment between the population about which scientific claims are to be made and the population to which results from statistical analyses can be extrapolated to. It needs to be stated precisely what the population refers to. Especially, whether Part A of an estimand specification requires the definition of several populations (target-, study-, and/or analysis-population).
154-155	13	**Comments:** "how to account for intercurrent events" suggests reference to how the intercurrent events should be handled during analysis. However, analysis cannot be part of the definition of an estimand. The estimand needs to be defined on the population level, while analysis refers to estimation of the estimand from study data. **Proposed change:** Wording should be changed accordingly.
154-155	17	**Comments:** The specification C is hard to use in practice, without going into methods. For example, will MMRM "automatically" make use of all available data to influence the last observation, via the correlation, without any imputation going on. But of methods are to be kept out of the estimand specification (A-D), then C gets to be very vague **Proposed change:** If A-D are to be void of methods, could there be some "possible methods" part where stuff like this could be described?
156	8	**Comments:** Could the wording 'as estimated by the relevant estimator' be added here? The estimate and estimator are terms that have been introduced, but it would be helpful to indicate how they relate to the four components of an estimand. **Proposed change:** The population-level summary for the variable <u>(as estimated by the estimator)</u> which provides, as required, a basis for a comparison between treatment conditions.
156	21	**Comments:** Could the wording 'as estimated by the relevant estimator' be added here? The estimate and estimator are terms that have been introduced, but it would be helpful to indicate how they relate to the four components of an estimand. **Proposed change:** The population-level summary for the variable (as estimated by the estimator) which provides, as required, a basis for a comparison between

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		treatment conditions.
159 - 160	16	**Comments:** The target population is usually the population intended for the treatment indication. The inclusion and exclusion criteria should reflect the indication. They may, however, deviate from the intended indication for several reasons, which may constitute an issue with respect to external validity depending on the underlying assumptions. Shouldn't the desired estimand first be defined on the final population of interest? A subsequent question may then be whether the clinical trial is still capable to reliably estimate this estimand if the inclusion and exclusion criteria deviate from the indication. **Proposed change:** Describe that the population (as part of the estimand definition) reflects the intended treatment indication which is reflected by the specific inclusion and exclusion criteria, but that in some instances inclusion/exclusion criteria might deviate from the intended indication (i.e. also the 'estimand population') which then becomes an issue of external validity.
159-160	21	**Comments:** If the target population is the patients that are eligible to be included based on the inclusion/exclusion criteria in the protocol, the guideline could clarify if it means that we should exclude from analysis patients randomized but who turned out to have exclusion criteria discovered thereafter.
159-162/417	21	**Comments:** The phrases "target population" & "study population" might seem to avoid the identifiability problems noted above. However, "target population" is used in another fashion in the EMA draft "Reflection paper on the use of extrapolation in the development of medicines for paediatrics" undergoing public consultation as of Nov 2018. Therefore, continuing to use "target population" & "trial population" (Line 417) in the addendum could lead to confusion eventually, if discussion of any analysis or study requires referring to both the addendum & the reflection paper. Do we need to introduce new terminology for different "sets" of participants in the study so that it is clear who is included in summaries of demography etc and so that "population" is not used in reporting for this purpose?
159, 163, 170, 183	12	**Comments:** In ACRO's view, it would be more helpful if these paragraphs explicitly identified the particular component of the estimand that is referenced. **Proposed change:** Add text to identify the the particular component of the estimand that is referenced.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
160-162	21	**Comments:** "In some cases, a stratum of those patients may be of interest, defined in terms of a potential intercurrent event; for example, the stratum of subjects who would adhere to treatment." Restricting the analysis to subjects who adhere to treatment could be in contradiction with the earlier claim that ITT principle should be followed. How to describe the impact on the label? **Proposed change:** Clarification should be provided regarding how this scenario would be reflected in the label.
160-162	21	**Comments:** The use of the word 'stratum' is strongly perceived as related to stratified random sampling and this is not the case as described in these lines. Adding the word "subset" might help. **Proposed change:** "In some cases, a stratum (subset) of those patients may be of interest ..."
163	21	**Proposed change:** Consider adding that these are measurements on an individual (patient) level, as opposed to the "summary measure" attribute which is summarising over patients.
164	15	**Comments:** use "quantifications" instead of "quantities"
167	21	**Comments:** Average HbA1c, rather than area under the curve of HbA1c, is a more appropriate measure of clinical benefit. **Proposed change:** "for example when using measurements taken prior to discontinuation (e.g., area under the curve of HbA1c until discontinuation average HbA1c over time until discontinuation;"
172-173	16	**Comments:** If a patient dies it is confusing to say that planned measurements will be "not observed" but rather that they simply do not exist. **Proposed change:** change "will not be observed" in line 173 to "will not exist"
172-173	21	**Comments:** "For example, if a subject dies before a planned measurement of blood

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		pressure, the blood pressure will not be observed". This is a scenario of 'cannot' and should be unambiguous. **Proposed change:** Change to ""For example, if a subject dies before a planned measurement of blood pressure, the blood pressure cannot be observed".
173	15	**Comments:** "blood pressure will not be observed" sounds awkward; use "value" or "variable" instead of "blood pressure"
175-176	21	**Comments:** The last part is not true in general. Please substitute "will" with "might". **Proposed change:** If a subject discontinues treatment because of toxicity, the blood pressure may be observed but might reflect the lack of effect of the treatment when it is not taken.
178-180	12	**Comments:** For clarity, ACRO recommends rewording this sentence as proposed below. **Proposed change:** Reword the sentence as follows: "Taking use of rescue medication as an example, two different specifications could include 1) The combined effect of treatment and rescue medication (the intercurrent event) and 2) the effect of the treatment in the potentially hypothetical absence of taking rescue medication (the intercurrent event).
178-180	21	**Comments:** Please substitute the first "and" with "with" for better readability (due to the second "and" in the sentence). **Proposed change:** Taking use of rescue medication as an example, two different specifications include the combined effect of treatment and with any intercurrent event…
182	17	**Comments:** The intercurrent events such as discontinuation of treatment due to lack of efficacy or AE; or introduction of rescue medication, may reflect the trial design rather than clinical practice. If the estimated treatment effect is dependend heavily on the strategy for dealing with intercurrent effects – will it then be relevant for a future patient, who will have a different risk to experience similar intercurrent events as observed in the trial? **Proposed change:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Consider to include a discussion of the interdependence between trial design and occurrence of intercurrent events
182	21	**Comments:** The intercurrent events such as discontinuation of treatment due to lack of efficacy or AE; or introduction of rescue medication, may reflect the trial design rather than clinical practice. If the estimated treatment effect is depends heavily on the strategy for dealing with intercurrent effects – will it then be relevant for a future patient, who will have a different risk to experience similar intercurrent events as observed in the trial? **Proposed change:** Consider to include a discussion of the interdependence between trial design and occurrence of intercurrent events
183-187	21	**Comments:** This paragraph discusses the summary measure for the variable, the fourth element of the estimand full description. Neither this section nor other sections clearly describe the distinction between the summary measure component of the estimand, and the estimator depicted in Figure 1 (Line 128). It is recommended that the distinction between these is clarified using a generic example. **Proposed change:** "…under two different treatment conditions. The estimator (Figure 1) is distinct from the population level summary measure; the estimator is a specific statistical method for calculating the estimand. For example, the summary measure of an estimand may be stated as the mean change from baseline (of variable X at time T), and the estimator might be '…. Calculated with ANCOVA using covariates of A, B, C' ."]
187	8/21	**Comments:** …..under two different treatment conditions. **Proposed change:** ….under two different treatment regimens.
188	8	**Comments:** In the section on "Strategies for addressing intercurrent events", it will be helpful if the document provides guidelines on how to handle missing data that was resulted from the intercurrent event (for example, the subject took rescue medication but it was not able to collect data afterwards).
188	12	**Comments:** The four components of the estimand have been given in the previous paragraphs but then the document immediately talks about intercurrent events (which is component C, third on the list). ACRO recommends that A

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		and B should be mentioned at least beforehand, even if no extra detail is given (even just to say these are equivalent to population/endpoint definitions in E9), or point out why they are slightly different here. **Proposed change:** Add appropriate text on A and B.
188ff	15	**Comments:** The strategies outlined in this section appear to be more than simply strategies to deal with intercurrent events, but rather represent complete estimand classes. According to Section 3.1. only Part C of an estimand definition deals with intercurrent events, the proposed concepts however already imply specifications for all four parts.
188	16	**Comments:** Why do you use the wording “constructing strategies” instead of “defining estimands”? Finally, we wish to have a proper definition of what is to be estimated (taking intercurrent events into account), “constructing strategies” rather sounds like strategies to search or estimate something. **Proposed change:** Avoid terms like ‘constructing strategies’ and use ‘defining estimands’ or ‘defining strategies to handle intercurrent events’ instead.
188	21	**Comments:** In the section on “Strategies for addressing intercurrent events”, it will be helpful if the document provides guidelines on how to handle missing data that was resulted from the intercurrent event (for example, the subject took rescue medication but it was not able to collect data afterwards).
189	8	**Comments:** Depending on whether this document should follow US or UK English, replace ‘through’ with ‘to’. **Proposed change:** The estimand attributes A to D…
189	21	**Comments:** Depending on whether this document should follow US or UK English, replace ‘through’ with ‘to’. **Comments:** The term inter-related is not clear: it should be removed and completed by the chart with the four circles defining the estimands attributes for more clarity. “The estimand attributes A through D introduced in Section A.3.1 should not

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		be considered independently".
189-190	21	"The estimand attributes A through D introduced in Section A.3.1 are inter-related and should not be considered independently." **Comments:** A sort of introduction to intercurrent events is provided in paragraph starting line 40. However, no clear definition is provided. **Proposed change:** Add clear definition of intercurrent events here.
190-191	20	**Comments:** The meaning of the sentence is unclear. Suggest replacing the words "without reflecting how potential intercurrent events are reflected in" with "without addressing how potential intercurrent events affect". **Proposed change:** The description of an estimand will not be complete without ~~reflecting how potential intercurrent events are reflected in~~ ***addressing how potential intercurrent events affect*** the scientific question of interest.
190-191	21	**Comments:** Please substitute "reflecting" with "accounting for" for consistency Throughout the document, in the context of intercurrent events, for the same reason outlined in the comment for line 122 **Proposed change:** The description of an estimand will not be complete without accounting for how…
191	8	**Comments:** This states that 'at least five strategies may be considered'. This suggests that all studies should consider five or more strategies, so: (1) Are there more to consider? (2) All five may not be relevant/appropriate for every trial /endpoint. (3) By introducing the word 'may' this suggests that not all five need to be considered, but this contradicts the 'at least five' earlier in the sentence. Can this wording be toned down to suggest 'Several strategies may be considered'? **Proposed change:** <u>Several</u> strategies may be considered.
191-198	16	**Comments:** The five strategies how intercurrent events are reflected in the primary

Line no.	Stakeholder no.	Comment and rationale; proposed changes
272-276		question that are discussed in detail in the addendum are probably the most important ones. Nevertheless, these are not exhaustive and other strategies are thinkable. In particular, the five strategies were mainly developed starting from considerations for longitudinal data. In other contexts, different strategies may be considered. For example, for time to event endpoints where the event of interest does not include death (or only death due to a specific reason), a competing risk strategy could be appropriate to handle the intercurrent event death (or non-specific death). **Proposed change:** Emphasize stronger that there are other possibilities how intercurrent events are reflected in the primary research question than those described in the addendum.
192	8	**Comments:** It is stated that 'the strategies can be used alone or in combination'. Is it worth adding 'within a trial' to clarify that several strategies could be used within a study or for a particular endpoint depending on the number and type of intercurrent events? **Proposed change:** The strategies can be used alone or in combination, <u>within a trial or for a particular endpoint,</u> to address multiple different intercurrent events.
193-198	5	**Comments:** It is out of context before it is understood what is meant by strategies. **Proposed change:** Move after the 5 strategies are presented
194	15	**Comments:** "targeted" is not clear **Proposed change:** e.g., suggest rewording to "of (primary) interest"?
197	15	**Comments:** It is agreed that the relevance of each strategy will depend on the therapeutic and experimental context. It is assumed, however, that a more generic, higher-level concept of relevance could be expressed (e.g. a preferred 'regulatory strategy') that in turn would be expected to impact e.g., the abovementioned experimental context. It is felt that therapeutic and experimental context will guide the choice of a most suitable strategy and this is understandable when assuming therapeutic and experimental context as 'fixed'. They are *not*, however, and with reference to our general comment made above on the regulator's perspective,

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		must not there be a trade-off between the ease of decision making and the relevance of the question that is being decided upon?
197	21	**Comments:** Minor alteration of the sentence to include another important aspect to consider. **Proposed change:** The relevance of each strategy will depend on the objective as well as on the therapeutic and experimental context.
198	8	**Comments:** What is an 'experimental situation' in this context - is it referring to the type of clinical trial?
198	13	**Comments:** One would expect "estimator" instead of "estimate" since the planning stage view is assumed here. **Proposed change:** Change estimate to estimator.
198	21	**Comments:** What is an 'experimental situation' in this context - is it referring to the type of clinical trial?
200	6/8/18/20/21	**Comments:** The referenced section A3.4 does not exist. **Proposed change:** "(…) considerations are addressed in Sections A.3.3, A.4 and A.5." **Proposed change:** It could be deleted as most probably A3.3.2 is meant. However, with reference to A3.3, both subsections are covered
200-295	5	**Comments:** The 5 strategies can be categorized into 3 groups: 1) de jure - **Hypothetical strategy**; 2) de facto - **Treatment policy strategy**; 3) on-treatment observed data only - **Composite strategy**, **While on treatment strategy**, **Principal stratum strategy**. The 1st group needs modelling/imputation … etc. and requires stronger assumptions for missing data, while 2nd and 3rd groups need no modelling/imputation and requires weaker assumptions for missing data. No

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		sure **Principal stratum strategy** is useful since none can predict which patients will not have intercurrent events prior to randomization. **Proposed change:** Re-organize the section to reflect the different natures of the strategies.
201	8/21	**Comments:** Wouldn't an adequate description of the chosen strategy to be used be the estimand itself? Are we therefore allowed to deviate from the 5 descriptions given in the document? Could this get confusing when trying to compare different studies?
204-205	8	**Comments:** 'the value of interest is used regardless of whether or not the intercurrent event occurs' **Proposed change:** 'the value of interest is used regardless of whether or not the intercurrent event occurs (assuming that all values post an intercurrent event are collected).'
204-205	20	**Comments:** Suggest replacing the word "irrelevant" with "ignored", and recommend adding a sentence immediately following this one such as "Randomization is preserved by this strategy."
204 - 212	8	**Comments:** The argument strongly equating the intent-to-treat (ITT) principle with a treatment policy estimand appears to be predicated on an ITT definition that embodies three requirements: 1) all patients included in the analysis, 2) all patients analysed according to their assigned treatment, and 3) all patients contribute data to support the desired analysis. Following the arguments laid out in Leuchs et. al. (Pharmaceutical Statistics 2017; 16: 12–19), the first two elements are widely recognized as the root of an ITT conforming analysis while the third condition is arguably stricter than originally intended when the ITT principle was first laid out. The section as currently written asserts that a treatment policy estimand would be impossible in the case of death of some subjects. This assertion is true under the implicit definition of an ITT population meeting all three criteria but is not necessarily true if the third condition is excluded (e.g. use of retrieved dropout data from an appropriate treatment arm and covariates to support imputation of a continuous variable otherwise missing due to death). Additionally, in some regulatory circles, there is a tendency to conclude that if an estimand definition isn't explicitly treatment policy, then it also isn't an ITT conforming analysis. This assertion would only be necessarily true if the ITT definition includes the third criterion.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		The operational ITT definition is still unclear in the current addendum (and also remains unclear in the original E9 guidance). To facilitate clear communication and concordance between regulatory agencies as well as between regulators and sponsors, a clear and precise definition of ITT should be provided.
204-212	21	**Comments:** Use of the words “irrelevant” and “ignored” are appropriate if using an ITT approach to estimating a treatment policy. If a marginal structural model is employed, these intercurrent events are accounted for as part of the estimand. **Proposed change:** "The value of the variable of interest is used regardless of whether the intercurrent event occurs. For example, when specifying how to account for rescue medication as an intercurrent event, observations on the variable of interest are used regardless of whether rescue therapy was taken.
206	12	**Comments:** For clarity, ACRO recommends rewording this sentence as proposed below. **Proposed change:** Reword the sentence as follows: “For example, when specifying how to account for rescue medication as an intercurrent event, occurrence of the intercurrent event is ignored and the observations on the variable of interest are used regardless of rescue medication intake”.
208-209	21	**Comments:** This document, appropriately, does not name any specific strategy as leading to what should be named the “intention to treat analysis” (ITT), and only notes that the treatment policy strategy ‘reflects’ the description in ICH E9 for ‘intention to treat principle’. There remains in many venues unclear usage of the term “ITT”, and often regarded that the treatment policy strategy for intercurrent events is the only approach that can lead to a “ITT analysis”. This sentence in the guideline may seem to imply this as well. This guideline will add valuable clarity to study design and analysis protocols and discussions by improving clear understanding of the term. **Proposed change:** “If applied across all types of intercurrent events, this reflects the comparison described in the ICH E9 Glossary (under Intention to Treat Principle) as the effect of a treatment policy. The term ‘ITT principle’ should be used to mean the principle of including all randomised study subjects in the analysis set, using data and methods as defined by the estimand. Furthermore, there is no single strategy for intercurrent events that should always be applied for an analysis to be consistent with the ITT principle, nor should any specific

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		analysis be named as 'the ITT analysis'. "
210	8	**Comments:** It is stated that "this [treatment policy] strategy cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects", however this cannot be known up front, therefore should this still be defined a priori even though there is a risk that not all subjects will have data after the intercurrent event. Should some guidance be provided on how to handle/impute missing data for such subjects? (As this is likely to be relatively common).
210-211	21	**Comments:** Clarify whether "do not exist" means that it's not possible for the event to exist (i.e. following death) or missing data for other reasons (e.g. lost to follow up as subject left the country and didn't want to continue). Is death the only reason for measurements not to exist? As written this section implies the treatment policy strategy cannot be applied to a study where even a single subject dies. If deaths are unrelated to treatment/disease this seems unreasonable. Please clarify if this is the intent. **Proposed change:** In general this strategy cannot be implemented when values for the intercurrent event do not exist for all subjects (unless the number of subjects is sufficiently small that this can be disregarded).
210-211	16	**Comments:** Please emphasise "not exist" to avoid misunderstandings and add something along the line "in contrast to just missing".
210-211	19	**Comments:** The text seems to imply that the treatment policy strategy **cannot** be implemented when values for the variable after the intercurrent event do not exist for **all** subjects. Lines 485-492 however describe using missing data methodology to account for missing data scenarios under the treatment policy strategy so this text seems quite restrictive and at odds with further descriptions accounting for missing data.
210-211	20	**Comments:** Please clarify: Is the intention really to say that the treatment policy strategy can only be used if a measurement is available for *all* patients (though potentially not under the conditions originally planned, e.g. after stopping treatment)? This seems restrictive, and suggests that implementation of the pre-specified strategy may depend on things that occur during the trial. Is it not allowable within a treatment policy strategy to utilize an analysis that imputes missing values, for example?

Line no.	Stakeholder no.	Comment and rationale; proposed changes
210-212	7	**Comments:** It is incorrect that the treatment policy strategy "cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects". For example, imputation techniques can be used to also include subjects with missing data after the intercurrent event. **Proposed change:** Please change the statement that the treatment policy strategy cannot be implemented to the statement that the treatment policy leads to problems when values for the variable after the intercurrent event do not exist for all subjects.
210-212	8	**Comments:** …this strategy cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects. **Proposed change:** Could we also state that 'In the case of values that were planned to be collected but were not, suitable imputation methods can be used (see section A7.1)'
210-212	9	**Comments:** Is there any general recommendation on how to handle missing values due to death?
210-212	13	**Comments:** It might be possible to implement the treatment policy strategy "when values for the variable after the intercurrent event do not exist for all subjects". For example, imputation techniques can be used to include also subjects with missing data after the intercurrent event. **Proposed change:** Change the statement that the treatment policy strategy cannot be implemented to the statement that the treatment policy leads to problems when values for the variable after the intercurrent event do not exist for all subjects.
210-212	15	**Comments:** does this imply: whenever we have a risk of patient dying during the trial before the primary endpoint (the variable) is measured, one cannot plan for treatment policy estimand? As the risk for death is always there, what is left?
210-212	17	**Comments:** It is not true that the treatment policy strategy "cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects". For example, imputation techniques can be used to include also

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		subjects with missing data after the intercurrent event. **Proposed change:** Change the statement that the treatment policy strategy cannot be implemented to the statement that the treatment policy leads to problems when values for the variable after the intercurrent event do not exist for all subjects.
210-212	21	**Comments:** It is stated that an estimand based on the treatment policy cannot be constructed with respect to a variable that cannot be measured due to death. However, even in a trial of a non-life-threatening disease, it cannot be excluded that death may occur, in particular in trials with large sample size and/or long duration. Does that mean that this strategy can never be fully achieved, and that other strategies, such as composite strategy or hypothetical strategy (eg. had no patient died) should be considered for such trials? **Comments:** …this strategy cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects. **Proposed change:** Could we also state that 'In the case of values that were planned to be collected but were not, suitable imputation methods can be used (see section A7.1)'
210-213	5	**Comments:** It is not realistic to assume that there will be 0% early termination, regardless of efforts to maintain patients in the study. Hence, when reading this the impression is that this strategy cannot be used in practice. This is discussed later in the documents when methods to address missing data are mentioned. **Proposed change:** Remove this sentence or soften it by saying that if data is not available, appropriate analytic approaches should be considered.
217	21	**Comments:** both "multiple" and "different" are used to describe approaches **Proposed change:** "There are multiple approaches that can be considered under this label."
217-224	21	**Comments:** An example of a composite strategy would help the reader to better understand the concept. We propose inclusion of an example such as the one

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		below in this section. **Proposed change:** A composite strategy may be implemented in HCV trials where data on response are collected after treatment discontinuation? If a subject prematurely discontinues treatment but still responds at PT Week 12 he is considered a responder. On the other hand, response data after rescue medication initiation are considered non-responders.
221-223	21	**Comments:** The terminology “numerical variable” should be clarified. Does it mean ordinal? Continuous or both? i.e. is it allowed to use a composite strategy with a continuous endpoint (provided you can define what a “extreme unfavourable value” is for the continuous endpoint)?
222	12	**Comments:** For clarity, ACRO recommends replacing “ascribed” with “assigned”. **Proposed change:** Replace “ascribed” with “assigned”.
223	13	**Comments:** It is unclear what is meant by “area-under-the-curve” (which curve?). Please specify.
223-224	9	**Comments:** Does the area-under-the curve approach require that large values of the continuous variable be favoured (that is, large values indicate good outcome, and small values indicate bad outcome)? Because otherwise, the composite strategy may interpret the occurrence of the intercurrent event as a good thing.
223-224	20	**Comments:** Please clarify how area-under-the-curve could be an important quantity for parties such as patients and prescribers (for example, how is this readily interpretable if the distribution of intercurrent event times if potentially
223-224	21	**Comments:** The use of an AUC based on values prior to the intercurrent event does not seem to reflect the composite strategy. It is for example mentioned as an example of the while on treatment strategy in line 743). This strategy seems also biased in case the variable would spontaneously deteriorate over time, which would be the case in progressive diseases (such as Alzheimer, Parkinson…)
223f and	15	**Comments:** It is unclear where the composite strategy using an Area under the curve

Line no.	Stakeholder no.	Comment and rationale; proposed changes
264ff		analysis differs from the on-treatment strategy.
225	12	**Comments:** For clarity, ACRO recommends rewording this sentence as proposed below. **Proposed change:** Reword the sentence as follows: “Sometimes an event being considered as intercurrent is itself the most meaningful variable that can be measured for quantifying the treatment effect of interest. This can be the case with death: the fact that a subject has died may be much more meaningful than observations before death, and observations after death will not exist.”
225-226	5	**Comments:** Baseline Observation Carried Forward (BOCF) can be seen, in some instances, as an extreme value approach yet regulatory agencies and the 2010 NRC report on missing data had reservations on its usage. **Proposed change:** Add clarification when such an approach can be used and what kind of justification should be provided.
225-228	16	**Proposed change:** Using the intercurrent event as the primary variable should be discussed in a separate paragraph not related to composite strategies. Could this even be a different strategy to be mentioned or is it rather just selecting another variable?
232	17	**Comments:** The naming “Hypothetical” seems unnecessary negative. In a causal inference manner of speaking all comparisons are “hypothetical”, and what makes the “hypothetical” strategy more so than the “principal strata”? **Proposed change:** Use another term to describe the strategy, for example “Scenario” – and require that the assumed scenario is described precisely.
232	21	**Comments:** Under what regulatory setting will the hypothetical strategy be appropriate? **Proposed change:** Give specific situation(s) under which a regulator might be interested in the hypothetical strategy.
232-247	7	**Comments:** We question the validity and utility of the hypothetical strategy. Even if a valid parameter estimation could be performed in the hypothetical scenario that an observed intercurrent event had not happened, what is the value of this

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		estimation in practice where intercurrent events *are* occurring? Moreover, no methods are available to estimate estimands in hypothetical scenarios with a low risk of bias. Maybe there are situations where estimands for hypothetical scenarios make sense as additional information for hypothesis generation or sensitivity analysis. Therefore, the hypothetical strategy should not be described as an option for the main analysis. **Proposed change:** Please delete the hypothetical strategy from the available options for the main data analysis. Define the hypothetical strategy as a possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
232-247	9	**Comments:** This section was not written clearly. Detailed examples are needed beyond what was written in subsequent sections.
232-247	12	**Comments:** This section talks about the hypothetical event being 'if rescue medication had not been available'. Conceptually, it is possible to see why it would be useful to answer this question but, practically, it is very difficult to understand how the analysis would look. ACRO recommends referencing a different intercurrent event which can be addressed via the hypothetical strategy.
232-247	17	**Comments:** I question the validity and utility of the hypothetical strategy. Even if valid parameter estimation could be performed in the hypothetical scenario that an observed intercurrent event had not happened, what is the value of this estimation in practice where intercurrent events *are* happening? Moreover, no methods are available to estimate estimands in hypothetical scenarios with low risk of bias. Maybe there are situations where estimands for hypothetical scenarios make sense as additional information for hypothesis generation or sensitivity analysis. Therefore, the hypothetical strategy should not be described as an option for the main analysis. **Proposed change:** Delete the hypothetical strategy from the available options for the main data analysis. Define the hypothetical strategy as possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
233	21	**Proposed change:** "A scenario is envisaged in which the intercurrent event would not occur" "A hypothetical scenario is envisaged with regard to the intercurrent event"
233-235	21	**Comments:** Suggested change in text for clarity.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: "… the value to reflect that scientific question of interest is that the value which the variable would have taken in the hypothetical scenario defined."
233-243	16	**Comments:** In the past an "attributable estimand" was often discussed that addresses the treatment effect if the effect is lost after an intercurrent event (usually addressed with placebo imputation). This 'estimand' does neither fit into the hypothetical category as defined nor any of the other four categories in the addendum: According to the addendum, the hypothetical estimand aims at a scenario where the intercurrent event would not occur, for example, the effect if no rescue medication had been available. However, if no rescue medication is available, there are two options for a patient who would have been given rescue medication in the trial: either continuing treatment without rescue medication, or discontinuing treatment at all (for example if he cannot tolerate the treatment). Only the first option seems to be covered by the definition of the "hypothetical" scenario in the addendum but the second option may be more relevant in many situations. **Proposed change:** Either a broader definition of the hypothetical strategy is needed or, alternatively, a 6th strategy "if another intercurrent event than the one actually observed had happened".
236-237	21	**Comments:** Please add a clarification what "not been available" means as we think this could be any of the items below. Intercurrent event never occurred or Rescue medication never approved Rescue medication not been made available to the subjects
237	14	**Comments:** Consider adding the term 'counterfactual' somewhere within this example (e.g. before the point on line 237, add, 'i.e., counterfactual event') since this is a well-known analysis strategy and enables the reader to tie the description in the paragraph with known methods
244-245	21	**Comments:** As above, an example will help the reader understand the concept here. We propose inclusion of an example such as the one below in this section. **Proposed change:** For example, the data after rescue medication was initiated would be excluded, and the model and/or imputation method used should assume or

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		impute missing data as if the results were continuing in the same trend as before the rescue medication was initiated.
245-247	21	**Comments:** Please correct important typo which alters the sentence meaning. **Proposed change:** For example, the hypothetical condition might usefully address both the use of a rescue medication and *non-adherence* to treatment as intercurrent events in order for an estimand to be precisely described.
247	8	**Comments:** Could we also add that suitable imputation methods may be used to reflect the hypothetical condition (see section A7.1).
247	21	**Comments:** Could we also add that suitable imputation methods may be used to reflect the hypothetical condition (see section A7.1).
248	17	**Comments:** The guidance for using the principal strata strategy in the current version is very limited. **Proposed change:** In the example section suggest to add suggestions for what steps would be involved in doing such an analysis **Comments:** A key assumption when selecting a principal stratum, is often that it should include patients that would complete the study on placebo. Such a selection may target a population with a high number of placebo responders, which could be counterproductive to showing effect in a study. **Proposed change:** Indicate that in a number of cases principal strata may not be relevant
248	21	**Comments:** The guidance for using the principal strata strategy in the current version is very limited. **Proposed change:** In the example section suggest to add suggestions for what steps would be involved in doing such an analysis
248-263	7	**Comments:** The "principle stratum strategy" is a purely hypothetical construct. Due to the given reason (confounding), principal strata could not be formed by subsets of

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		patients without intercurrent events. Therefore, no methods are available to deal adequately with purely hypothetical principal strata. **Proposed change:** Please delete the principle stratum strategy from the available options for the main data analysis. Define the principle stratum strategy as a possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
248-263	12	**Comments:** The sentence in lines 254-255 talks about the principal strata effect that should be distinguished from subgroups based on trial data, but the previous paragraph gives the example of adherence to determine the principal stratum. In ACRO's view, adherence is a component of trial data, therefore we are unclear what is meant by this sentence. Also, the meaning of line 260 "membership in a principal stratum must then be inferred, usually imperfectly from covariates" is unclear, as is the intended guidance to sponsors on this point. **Proposed change:** Clarify this section to ensure the meaning and guidance to sponsors is clear.
248-263	13	**Comments:** It is unclear whether in practice the principal stratum strategy would be considered adequate in situations where members of a principle stratum cannot be identified in advance, which will typically be the case. What claims could be derived for a treatment with superiority proven by analysis using a principle stratum strategy and adequate methods to address confounding, sensitivity analysis etc., given that in practice it cannot be told whether or not the patient falls into the principle stratum? **Proposed change:** Please provide considerations and examples on the applicability of the principal-stratum strategy.
248-263	17	**Comments:** The "principle stratum strategy" is a purely hypothetical construct. Due to the given reason (confounding) principal strata could not be formed by subsets of patients without intercurrent events. Therefore, no methods are available to deal adequately with purely hypothetical principal strata. **Proposed change:** Delete the principle stratum strategy from the available options for the main data analysis. Define the principle stratum strategy as possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
248-	21	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
263		It looks like that unless eligible patients are identified before randomization, for example based on a run-in period (i.e. a randomized withdrawal trial), this strategy will always rely on untestable assumptions and that no robust estimator can be proposed (see also later comments on lines 728-735). This makes the relevance of this estimand questionable. It is also unclear what would be the impact on the labelling on this estimand. **Comments:** It will aid clarity to note the distinction between the other strategies which address what observed data to include in the analysis set and how to address missing data to enable use of the analysis set data, versus the principle stratum strategy which addresses intercurrent events by defining which study subjects should be included in, or entirely excluded from, the analysis set. **Proposed change:** "… because different subjects will experience different intercurrent events on different treatments. The principle stratum strategy differs from several other strategies by defining which study subjects will be represented in the analysis set, rather than how to include the observed data or occurrence of missing data in the analysis set."
248-276	16	**Comments:** Both, principal stratum and while-on-treatment strategies do not fully cover the treatment effect and should usually be accompanied by the proportion of subjects in the different strata or the analysis of the time-to-intercurrent event, respectively, if used. **Proposed change:** Include sentences that highlight this issue.
249	21	**Comments:** Line 249 introduces the principal stratum approach, and states that the principal stratum can define the target population. This is inappropriate. As a fundamental principal, the target population should be one that a treating physician can identify; however, the principal stratum cannot be identified by a physician. There is a subsequent example on line 423, where the principal stratum consists of patients who tolerate the treatment. Since patients must be treated in order to determine who tolerates it, it is of questionable use to a physician that the group of patients who should receive the medication is the group that tolerates it. It would not be sufficient to claim that physicians should simply treat all potential patients until tolerability is determined, because the effect (positive or negative) of the treatment in the subset of patients who don't tolerate it would need to be accounted for.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
251-253	9	**Comments:** How would this group of patients be identified prior to enrolment?
251-253	21	**Comments:** As written, the sentence “In other words, a principal stratum is a subset of the broader population who would not experience the intercurrent event” is misleading since it implies that a principal stratum is always a subset of the population in which an intercurrent event would not occur. With reference to the previous sentence (line 250-1), the principal stratum of interest is the subset of the population who would not experience the intercurrent event on either treatment. **Proposed change:** “In other words, the relevant principal stratum in this case is the subset of the broader population who, would not experience the intercurrent event”
254	21	**Comments:** Please better clarify the difference between principal strata and subgroups.
255-261	21	**Comments:** from “Principal stratification….In contrast” repeats the definition in the Glossary.
255, 714, 857	21	**Comments:** "... patient's potential intercurrent events on both treatments ..." The Glossary defines four possible principal strata. However, an implementation of this would be the patients who would not have had an intercurrent event of treatment withdrawal on treatment B (the novel treatment) irrespective of whether they had an intercurrent event on treatment A (the standard of care). An example of which would be a treatment for a chronic condition for an estimand of interest to a Payer who will only be interested in those patients who are taking the novel treatment long-term in comparison to a treatment policy estimate of the efficacy on the standard of care in this population. This seems a more useful analysis than the no intercurrent event under either treatment which is primarily discussed in the amendment, which is a harder population to conceptualise. **Proposed change:** Consider adding the above as an example of using the principal strata strategy.
259	21	**Comments:** "... randomised controlled trial because each patient will be observed on one treatment only". This is true for a parallel group study. However, in a cross-over study these patients could be identified (under some assumptions at least, e.g. on wash-out).

Line no.	Stakeholder no.	Comment and rationale; proposed changes
260	21	**Comments:** Please clarify what is meant by “inferred from covariates”. How is this done? **Comments:** Example(s) of the statistical model that should be used in that case should be provided. In addition the use of wording "as imperfectly" seems to indicate that the approach is controversial, in that case is it pertinent to mention it?
263	14	**Comments:** Principal stratum strategy line 263 - add as another caveat: generalizability of the trial results should be considered, this may be a challenge if the principal stratum does not make up the vast majority of enrolled patients.
264	20	**Comments:** Suggestion that the “while on treatment” strategy be better motivated, perhaps with a more specific example. This description suggests that the times and duration of assessment of the variable in question are irrelevant, and perhaps stay constant during the period of interest. Can we expect that this can be meaningfully interpreted when the on-treatment durations differ across treatments? Despite the terminal illness example mentioned, this strategy seems to occupy an extremely narrow and rare niche, and clearer rationale and justification seem needed.
264	21	**Comments:** The name while on treatment strategy could suggest that "last value under treatment" could fall under this heading. Consider to make a more explicit statement whether or not this approach could be considered as a "while on treatment strategy".
264-271	7	**Comments:** The restriction of the data analysis to the period of treatment continuation leads to serious problems due to different follow-up times. Therefore, this strategy should be avoided in general. Maybe there are situations where the “while on treatment” estimand makes sense as additional information for hypothesis generation or sensitivity analysis. However, the “while on treatment” strategy should not be described as an option for the main analysis. **Proposed change:** Please delete the “while on treatment” strategy from the available options for the main data analysis. Define the “while on treatment” strategy as a possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
264-271	17	**Comments:** The restriction of the data analysis to the time period of treatment

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		continuation leads to serious problems due to different follow-up times. Therefore, this strategy should be avoided in general. Maybe there are situations where the on treatment estimand makes sense as additional information for hypothesis generation or sensitivity analysis. However, the while on treatment strategy should not be described as an option for the main analysis. **Proposed change:** Delete the while on treatment strategy from the available options for the main data analysis. Define the while on treatment strategy as possible supplementary analysis for hypothesis generation or sensitivity analysis in special situations.
264-271 743	13	**Comments:** For the while on treatment strategy, the use of the “average of the designated measurements while on randomised treatment” can lead to problems in case of different follow-up times. **Proposed change:** Please comment on how to deal with the case of unbalanced times on treatment when the while on treatment estimand is considered.
264-276	14	**Comments:** "While on treatment strategy" – One might consider that the "holy grail" is the modelling of a joint process of treatment discontinuation/modification and effect while on treatment. The "while on treatment" strategy by itself does not seem to further the goal of improving treatment choice unless discontinuation/modification is exogenous (which is not likely to be the case in any interesting circumstance)
265	21	**Comments:** For clarity, please consider minor alteration to the sentence. **Proposed change:** Only response to treatment prior to the occurrence of the intercurrent event is of interest.
265-267	21	**Comments:** “If a variable is measured repeatedly, its values up to the time of the intercurrent event may be considered to account for the intercurrent event, rather than the value at the same fixed timepoint for all subjects.” Please clarify, are the measurements up to the intercurrent event considered to account for the intercurrent event or rather to account for the response to treatment? If no imputation of measurements after the intercurrent event is needed then please explicitly clarify. **Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		The content of this sentence relating to how to deal with different durations of observations for patients or treatments requires further detail as the meaning was not clear to us as a reader.
266-267	21	**Comments:** The wording/sentence structure is ambiguous. **Proposed change:** "… its values up to the time of the intercurrent event may be considered to account for the intercurrent event the only relevant values in this strategy rather than the value at the same fixed timepoint for all subjects."
272	21	**Comments:** The text should not be restricted to 5 possible strategies, the possibility of adding new strategies that might be devised in future should be allowed for. **Proposed change:** "Altogether, five different strategies are considered in this section. Other strategies for an intercurrent event are not precluded, and should be considered if relevant and appropriate."
272-273	5	**Comments:** It is not clear if participation based in principal stratum based on covariates should be estimated at the end of study when these covariates are being estimated or based on an algorithm which is not estimated in the study and is pre-defined. In case of the former, what are the potential implications on type I error?
272-276	7	**Comments:** The five strategies are listed on the same level although only two strategies should be used as the main analysis in practice. **Proposed change:** Please divide the list of strategies into two parts. One part with options for the main analysis (treatment policy, composite) and a subordinate part with options for supplementary analyses in special situations (hypothetical, principal stratum, while on treatment).
272-276	15	**Comments:** The message of this additional paragraph is unclear. Does it represent a conclusion, or a further recommendation? **Proposed change:** Proposal to include it under the respective strategies or to consider re-wording.
272-	17	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
276		The five strategies are listed on the same level although only two strategies should be used as main analysis in practice. **Proposed change:** Divide the list of strategies into two parts. One part with options for the main analysis (treatment policy, composite) and a subordinate part with options for supplementary analyses in special situations (hypothetical, principal stratum, while on treatment).
272-276	17	**Comments:** The section concerns the previous five subsections. A separate subsection for these lines could help the reader to acknowledge this.
272-276	21	**Comments:** This paragraph is not a continued description of the topic 'While on treatment strategy', but a final overview of the all topics under Section A.3.2. It is recommended that this paragraph becomes a separate summary, preceded by a heading. The other comments provided here relating to this overview/summary will also be more clearly relevant if this is added. **Proposed change:** "Diversity and precision of strategies Altogether, five different strategies are considered in this section." **Comments:** Section A7 provides simple generic examples of applying strategies to create an estimand. Even the examples with two intercurrent events are simplified to enable concise presentation. Actual studies, however, will often have multiple types of intercurrent events, each of which needs to have a strategy selected to create a comprehensive estimand. This paragraph, or new paragraph within the newly named subsection, should state the expectation that real study planning will often require explicit identification of all types of expectable intercurrent events, and a diversity of strategies across the intercurrent events might be selected to address the events. Although this thought is briefly expressed in a later portion of the document (Line 288), this will be a valuable location to also express it to avoid potentially leaving the impression that the just described strategies are to be selected among for uniformly applying to all events for any individual estimand. **Proposed change:** "… (iii) the effect during adherence. Actual studies, particularly long or complex, or in complex clinical circumstances, may be expected to have multiple types of intercurrent events. A well-defined estimand will identify a strategy for each type of intercurrent event. Often a study objective will be best served by employing a variety of

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		strategies across the event types in defining an estimand."
279-285	21	**Comments:** Lines 279-285 to "…trial results" repeat material that is covered elsewhere, e.g., in lines 144-150.
279-289	16	**Comments:** Examples and discussions in the addendum as well as in estimand discussions beyond the addendum, mostly focus on intercurrent events that have a permanent influence, meaning the intercurrent event happens at a specific time point and changes the conditions under which a patient is observed permanently (e.g. patient discontinues treatment prematurely after 3 weeks). However, the estimand framework is more generally applicable also to intercurrent events that have only a 'temporary' influence for a single patient (e.g. treatment interruptions, or intake of rescue medication that has an effect only for a limited time). These 'temporary' intercurrent events can and should also be addressed. **Proposed change:** It should be clarified in the addendum that intercurrent events are not limited to those that have a permanent influence but include also those with a temporary influence. For events with a temporary influence, the 'duration' of the intercurrent event may need to be considered for the construction of an estimand (for example by imputation of values during rescue medication is effective for a hypothetical estimand)
281-283	11	**Comments:** While this makes sense, in practice fine-grain differentiation will be statistically extremely difficult or impossible and so multiple intercurrent event types will need to be grouped together: Most clinical trials will have insufficient patients to allow for differential handling of more than 2-3 types of intercurrent event in a statistically sound manner (e.g. by within-group imputation). It can also be extremely difficult to objectively classify many types of these events; stated reasons may not be the real underlying reason (for instance, patient withdrawing consent). A patient might be labelled as discontinued, but then may or may not have gone on to take other medication. Although these issues may be reduced by improved training and recording, they are difficult to remove entirely as classification of many types of intercurrent events is subjective and/or ambiguous. In general, even when handled sensibly, analysing different types of intercurrent event in different ways will result in increased statistical complexity, variance and likelihood of analysis failure **Proposed change:** We would welcome guidance on this topic, possibly including a framework/structure for recording and classifying intercurrent event types in

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		an objective manner, and potentially standard groupings.
281-283	21	**Comments:** "The construction of the estimand should address each intercurrent event that may occur in the clinical trial and that will affect the interpretation of the results of the trial" **Comments:** That sentence is in conflict with Lines 285-286 ("It may be impractical to foresee every relevant kind of intercurrent event"). **Proposed change:** Change to ""The construction of the estimand should address the practically foreseeable intercurrent events that may occur in the clinical trial and that will affect the interpretation of the results of the trial" **Comments:** While this makes sense, in practice fine-grain differentiation will be statistically extremely difficult or impossible and so multiple intercurrent event types will need to be grouped together: Most clinical trials will have insufficient patients to allow for differential handling of more than 2-3 types of intercurrent event in a statistically sound manner (e.g. by within-group imputation). It can also be extremely difficult to objectively classify many types of these events; stated reasons may not be the real underlying reason (for instance, patient withdrawing consent). A patient might be labelled as discontinued, but then may or may not have gone on to take other medication. Although these issues may be reduced by improved training and recording, they are difficult to remove entirely as classification of many types of intercurrent events is subjective and/or ambiguous. In general, even when handled sensibly, analysing different types of intercurrent event in different ways will result in increased statistical complexity, variance and likelihood of analysis failure **Proposed change:** We would welcome guidance on this topic, possibly including a framework/structure for recording and classifying intercurrent event types in an objective manner, and potentially standard groupings.
282	20	**Comments:** Suggest changing "may occur" to "is anticipated to possibly occur" to avoid a conflict with line 286 ("impractical to foresee every relevant kind . . .")
285	21	**Comments:** Please add further clarification why these specific criteria do not affect interpretation of trial results, since we cannot see why such criteria are not expected to affect the interpretation of trial results.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
286-288 also 585-588	20	**Comments:** These sentences seem unclear. Please clarify the distinction between the impact of unforeseen intercurrent events on the planned analysis versus additional post-hoc analyses that may be defined to address them, for example.
286-288	21	**Comments:** “the effect on what the chosen analysis estimates” is quite unclear. An example might be useful.
288	12	**Comments:** For clarity, ACRO recommends rewording this sentence as proposed below. **Proposed change:** Reword the sentence as follows: “Trial reporting should then discuss not only the way unforeseen intercurrent events were handled in the analysis but also the effect on what the chosen analysis estimates.”
289-290	14	**Comments:** we agree with this statement, and it is essential. The construction of the estimand(s) in any given clinical trial is a multi-disciplinary undertaking including clinicians, statisticians and other disciplines involved in clinical trial design and conduct. **Proposed change:** Is it worth also adding this at the top of the document, e.g., within section A.1. Purpose and Scope, so it is clear who this guidance is for (=not only clinical statisticians)?
291 - 293	15	**Comments:** This discussion should also be part of the CSR/dossier and consider prior knowledge and literature.
297	21	**Comments:** Suggested change in text for clarity. **Proposed change:** “An iterative process may be required to construct the estimands of interest.”
299-301	16	**Comments:** Whereas the estimand itself is not defined by the study setting, the set of realistic clinical trial options may impact the usefulness of certain estimands. With other words, a quantity to appreciate the drug’s effectiveness may only be useful if it can actually be measured or reliably estimated. In that sense, e.g. a composite estimand may be selected in some settings primarily because of its feasibility to be measured. The document could further

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		comment on the inter-relation between feasibility and desired measure. **Proposed change:** Inter-relation between feasibility and desired estimand could be discussed in an additional paragraph in section A.3.3.1 or A.3.3.2.
299-301	21	**Comments:** Minor alteration of the sentence for clarity. **Proposed change:** "…which is reliable for inference regarding the estimand can be derived."
300	21	**Comments:** Methods of data collection (including accurate recording of the occurrence of intercurrent events) must also be considered. **Proposed change:** Suggest add "and methods of data collection"
300-306 and 574 and 635-640	5	**Comments:** The text in lines 300-306 explains that intercurrent event may be unforeseen at the planning stage. If so, the estimator that takes into account intercurrent events cannot always be pre-defined. This notion is not consistent with the requirement/principle of pre-defined primary end-point. Also row 574 and 635-640 mention that estimator should be pre-defined. This contradiction between requirement for pre-definition and the fact that intercurrent is not always foreseen should be addressed, and the question of how estimators that used intercurrent events that were not foreseen at the planning stage (or post hoc estimators) will be considered by regulators should be addressed.
302-303	7	**Comments:** The formulation "Some estimands, in particular those that are estimated using the observed data, …" is unclear and makes no sense. If it means that an estimand is sometimes defined by the data observed, the statement is invalid because theoretical parameters should not be defined by the data observed. If it means that some estimands are estimated by the data observed and others not, the statement is of no use, because an estimand is only meaningful if it is estimable by means of the data observed. **Proposed change:** Please delete or revise the statement "Some estimands, in particular those that are estimated using the observed data, …".
302-	9	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
303		Regarding the sentence “Some estimands, in particular those that are estimated using the observed data,”, what other kinds of estimands are there?
302-303	17	**Comments:** The formulation "Some estimands, in particular those that are estimated using the observed data, …" is unclear and makes no sense. If it is meant that an estimand is sometimes defined by observed data the statement is invalid because theoretical parameters should not be defined by observed data. If it is meant that some estimands are estimated by observed data and others not, the statement is useless, because an estimand is only meaningful if is estimable by observed data. **Proposed change:** Delete or reformulate the statement "Some estimands, in particular those that are estimated using the observed data, …".
302-304	20	**Comments:** Suggest deleting the phrase “in particular those that are estimated using the observed data”; this is confusing and does not really seem needed. **Proposed change:** Some estimands~~, in particular those that are estimated using the observed data,~~ can be robustly estimated making few assumptions, whereas other estimands require more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions (see Section A.5.1).
302-305	21	**Comments:** “Some estimands, in particular those that are estimated using the observed data, can be robustly estimated making few assumptions, whereas other estimands require more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions” It seems the clause,“…in particular those that are estimated using the observed data…” is confusing since any estimand would use observed data. It is just that some estimands may also use external data (e.g. those incorporating the hypothetical or principal stratum approaches). **Proposed change:** Change to “Some estimands, in particular those that are estimated using only observed data from the study, can be robustly estimated making few assumptions, whereas …”
302-308	21	**Comments:** There is much emphasis on the robustness of estimating using the observed

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		data. This might not be true in some situations. For example, in rheumatoid arthritis, subjects who do not respond well are allowed to escape to the test treatment. If a treatment policy is requested for this "escape" intercurrent event, the comparison of interest might become between the drug and same drug with a delayed start. A hypothetical strategy might be more useful. **Proposed change:** "… trial design and analytic approach would need to be considered. In some circumstances the estimand most 'robustly estimated' may be not addressing a useful question and a less robust, but more relevant estimand may be preferred as the primary estimand."
303	13	**Comments:** "in particular those that are estimated using the observed data". All estimations will make some use of observed data, so maybe "using exclusively observed data" may be more appropriate. **Proposed change:** Clarify and revise accordingly, e.g. "using exclusively observed data".
303	15	**Comments:** "using the observed data" is slightly misleading as observed data would likely always be used. Is "using only observed data" meant? Please clarify/revise
303	16	**Comments:** All estimands are estimated using observed data, even if these observed data are used for missing data handling. **Proposed change:** Change the sentence to 'Some estimands, in particular those that are estimated using only the observed data without requiring further assumption (e.g. for missing data handling), can be robustly estimated making few assumptions, whereas other estimands require more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions'
304	21	**Comments:** Clarify which estimands are considered able to be robustly estimated, and which may be more sensitive to changes in assumptions. Are treatment policy and composite estimands the robust ones and hypothetical, principal stratum the less robust ones?
306	16	**Proposed change:** Replace "to derive a reliable estimate" by ".. to provide a reliable estimator".
306	21	**Comments:** Where significant issues exist to develop an appropriate trial design or to

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		derive a reliable estimate for a particular estimand, an alternative estimand, trial design and analytic approach would need to be considered. It would be appreciated to clarify in an example when different estimands might be possible to evaluate different strategies for addressing intercurrent events. E.g. for Oncology would it be possible to use overall survival time as estimand for treatment policy strategy (de facto estimand) and PFS time as estimand for hypothetical strategy (de jure estimand)? This would avoid the need of making several assumptions difficult to check when using methods like Rank Preserving Structural Failure Time
306-308	21	**Comments:** In the described situation, is a discussion in the study protocol expected that another estimand would be more appropriate, but cannot be reliably estimated?
309 ff (A.3.3.2)	15	**Comments:** As to the experimental context, one could also imagine that different estimands might be appropriate in different phases of drug development. Earlier phases being more interested in describing the “pure” pharmacological drug effect (without any intercurrent events) whereas later stages will likely try to be more reflective of the therapeutic setting/clinical use of the product (irrespective of non-adherence or use of rescue medication).
318-326	21	**Comments:** This discussion begs the question of a “robust estimate” ever exists in cases where we need to model what would have happened to a given subject under a different treatment. **Proposed change:** Provide one or more specific examples of robust estimators in these more challenging settings.
321	3	**Comments:** Suggestion to change text to ‘could adhere to treatment’ instead of ‘~~can~~ adhere to treatment’ as this would better reflect the hypothetical nature of the principal stratum.
325-326	20	**Comments:** It cannot be “agreed” that a robust estimate can be obtained; this will depend on the data. We suggest conveying something along the lines of “ideally if a sensible sensitivity plan can be described”.
326 460	16	**Comments:** Please here and in other places: It is the *estimator* (or estimation function) that is robust, but not the *estimate* or the *result*. The estimate may be reliable (since the estimator is robust).

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Robustness refers to the operating characteristics of an estimation procedure (related to the design or the analysis) if the underlying assumptions are changed. Robustness appears to be used in a rather vague manner: Estimation function (estimators) may be more or less robust depending on their bias with respect to a given estimand when underlying assumptions are changed. Robustness of an estimand would remain unclear or even not applicable. Maybe an estimand can be defined as robust if it captures (as single parameter) maximum information of the drug's effect. However, the concept of robust estimands would need further reflections. **Proposed change:** Throughout the whole document, robustness should be used in a precise manner in relation to estimator (estimation function) and not the estimate/result. Robustness of an estimand could be discussed but would need a proper definition.
329-331	21	**Comments:** We modified the sentence because we did not understand why "for use in treatment naïve subjects" is mentioned and thought it was possibly a mistake as the sentence has the same meaning for non-treatment naïve patients. **Proposed change:** If the treatment is proposed for use in treatment naïve subjects as part of a treatment policy…
331-334	21	**Comments:** In some clinical trials, the number of subjects with an intercurrent event could be small. Statistical inference suggested by the addendum in Lines 331-332 for an additional estimand and analysis pertaining to the intercurrent event could be misleading due to small sample size and lack of the power. **Proposed change:** … inference can be complemented by defining an additional estimand and summary analysis pertaining to the intercurrent event itself…
335-338	11	**Comments:** The recommendation to use treatment policy estimands even when they are known to be less clinically-relevant is strongly disagreed with, and would result in worse regulatory decisions. If the estimand of clinical interest were hypothetical, then use of treatment policy as a surrogate estimand would not improve decision making. This is because the treatment policy estimate will typically be used and reported without allowance made for it not reflecting the most clinically relevant estimand, and the significance testing upon which regulatory decision making is primarily based cannot account for this discrepancy in 'meaning'. Use of treatment policy in this case therefore does

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		not remove bias, it merely moves it, hides it, and amplifies it: The known statistical biases of hypothetical estimation (whose potential impact is quantifiable by sensitivity analysis) are exchanged for clinical bias in treatment policy, which is harder to identify or quantify as it occurs outside a statistical framework, but which logically must on average be greater in magnitude (else the treatment policy estimator would also be a less biased estimator of the hypothetical estimand than the hypothetical estimator itself). Where treatment policy is less clinically relevant, the use of it due to its 'unbiased estimation' in practice simply represents the introduction of a strong conservative bias to the estimation of the desired (e.g. hypothetical) estimand. Where the primary concern is robustness and there are no better methods available, the correct response is to perform sufficient sensitivity analysis to assess the impact of reasonable deviations of assumptions. As this is already mandatory good practice, there is no need to deliberately choose clinically less-relevant estimands. **Proposed change:** Remove this recommendation entirely.
335-338	15	**Comments:** This statement seems of great importance for the overall topic (please also see previous comments on general regulatory preference and on balancing clinical meaningfulness against reliable estimates/robust inference). **Proposed change:** This could be discussed more extensively and more prominently in the document.
335-338	20	**Comments:** This seems an important message, but would be clearer if the order of its two main ideas were reversed. For example, an estimand of greatest clinical importance is identified, but perhaps an adequate sensitivity plan cannot be determined; in such cases a readily-interpretable treatment policy estimand that still has clinical importance and relevance for regulatory decision making may be elevated in importance.
335-338	21	**Comments:** The recommendation to use treatment policy estimands even when they are known to be less clinically-relevant is strongly disagreed with, and would result in worse regulatory decisions. If the estimand of clinical interest were hypothetical, then use of treatment policy as a surrogate estimand would not improve decision making. This is because the treatment policy estimate will typically be used and reported without allowance made for it not reflecting the most clinically relevant estimand, and the significance testing upon which regulatory decision making is primarily based cannot account for this discrepancy in 'meaning'. Use of treatment policy in this case therefore does not remove bias, it merely moves it, hides it, and amplifies it: The known

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		statistical biases of hypothetical estimation (whose potential impact is quantifiable by sensitivity analysis) are exchanged for clinical bias in treatment policy, which is harder to identify or quantify as it occurs outside a statistical framework, but which logically must on average be greater in magnitude (else the treatment policy estimator would also be a less biased estimator of the hypothetical estimand than the hypothetical estimator itself). Where treatment policy is less clinically relevant, the use of it due to its 'unbiased estimation' in practice simply represents the introduction of a strong conservative bias to the estimation of the desired (e.g. hypothetical) estimand. Where the primary concern is robustness and there are no better methods available, the correct response is to perform sufficient sensitivity analysis to assess the impact of reasonable deviations of assumptions. As this is already mandatory good practice, there is no need to deliberately choose clinically less-relevant estimands. **Proposed change:** Remove this recommendation entirely.
335-342	21	**Comments:** This paragraph states that in some cases the estimand based on the treatment policy strategy may be the one best suited to support regulatory decision making (e.g., hypothesis testing) because it may best support robust inference when estimands of greater clinical interest cannot be formulated with similar robustness. The sentence encouraging use of a treatment policy estimand is then stated to be "still relevant". This phrasing may be misunderstood to mean that if the treatment policy estimand supports robust inference and regulatory decision making, then it is also the relevant estimate to consider for all purposes. The last sentence in the paragraph only partly ameliorates that. **Proposed change:** "… that are agreed to support a reliable estimate or for robust inference. An estimand based on the treatment policy strategy might offer the possibility to obtain a reliable estimate of a treatment effect that is still adequately relevant. In this situation, it is recommended to retain those estimands that are considered to be of greater clinical relevance and to present the resulting estimates all estimands and resulting estimates along with a discussion of the relevance and limitations of each, in terms of trial design or statistical analysis, for that specific approach.
336-338	21	**Comments:** It is unclear what is meant by "specifically in settings where estimands based on alternative strategies might be considered of greater clinical interest, but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference". Does it refer to "Principal stratum strategy". Does it mean that treatment policy strategy would be used instead of this strategy? Are there any case studies that underlie this

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		statement in the draft ICH E9 (R1)?
337-338	7	**Comments:** The following statement is unclear "… but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference." **Proposed change:** Please clarify what is meant by the statement "… but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference."
337-338	17	**Comments:** The following statement is unclear "… but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference." **Proposed change:** Please clarify what is meant by the statement "… but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference."
341-342	8/21	**Comments:** …present the resulting estimates along with a discussion of the limitations..' Would it be possible to indicate that this information is also useful to include in a label, particularly for physicians/
343-348 and 663-691	12	**Comments:** These sections recommend dichotomising the data but there is no mention of the loss of power that results from this approach. It is therefore not clear why this is recommended. Also, in the composite strategy example, if this strategy is used as a sensitivity analysis then it is likely to have less power than the primary analysis. ACRO recommends that it be made clear in the assessment of the strategy that, to conclude consistency between the continuous and binary approach, it is not required to reach significance for both but that the conclusions around the treatment effect estimate should be consistent. **Proposed change:** Clarify why the dichotomous approach is recommended and its effect on power, and that in the assessment of the composite strategy, to conclude consistency between the continuous and binary approach, it is not required to reach significance for both but that the conclusions around the treatment effect estimate should be consistent.
343-348	14	**Comments:** it's unclear here if this is the recommended or to be avoided.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
343-348	21	**Comments:** Please do not recommend dichotomizing a continuous variable. There is substantial literature discussing the fact that this approach has tremendous cost in statistical efficiency, while failing in its goal to address the question of clinical meaningfulness. **Proposed change:** Balance this section with the disadvantages of dichotomizing a continuous variable and include the possibility of retaining discrimination in the continuous variable
370-371	3	**Comments:** It seems that a part of the sentence has gone missing, as sentence in line 371 does not start with a capital. 'Use of a treatment other than the one assigned will commonly be considered as an intercurrent event. prohibited by the protocol or use of a subsequent line of therapy.'
370-371	6	**Comments:** The text isn't well written, it seems that there's a drafting error. **Proposed change:** "Use of a treatment other than the one assigned will commonly be considered as an intercurrent event, prohibited by the protocol or use of a subsequent line of therapy."
370-371	21	**Comments:** These two lines seem not to correctly connect. Likely, there is a full stop at the end of line 370 that should not be there. **Proposed change:** Remove full stop at the end of line 370 or re-word: Use of a treatment prohibited by the protocol or use of a subsequent line of therapy will commonly …
370-375	13	**Comments:** The meaning of this paragraph is unclear. Please reword for clarification.
371	9	**Comments:** Delete 'prohibited by the protocol or use of a subsequent line of therapy'
371	15	**Comments:** there is a fragment of a sentence … **Proposed change:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		delete?
377	14	**Comments:** the document doesn't really give any guidance on how such deviations should be handled, and what needs to be done in those cases. **Proposed change:** Could be complemented with some references to other sources if not adding a brief suggestion in the text itself.
377-378	21	**Comments:** “The choice of estimands for studies with objectives to demonstrate non-inferiority or equivalence requires careful reflection” This sentence adds little to the exposition. All elements of study design require careful reflection. **Proposed change:** Suggest deleting this sentence.
377-390	16	**Comments:** The distinctions between non-inferiority and superiority may require further conceptual considerations: Whereas the relevant estimand to be estimated may in principle be the same, the derivation of a relevant non-inferiority margin and the requirement of a sensitive trial may render certain estimands unfeasible in a non-inferiority setting. **Proposed change:** Discuss the conceptual considerations of estimands in superiority and non-inferiority trials in more detail (e.g. at the end of section A.3.3.2).
377-390	21	**Comments:** While the rest of the document stays away from specifying definitions of analysis sets, this concept is introduced in this paragraph with a discussion around the FAS and no mentioning of the Per Protocol Analysis Set. It is unclear what strategies could be employed for the intercurrent events of “protocol violations and deviations, non-adherence and withdrawals “ i.e. whether subjects with these events should be eliminated from the population definition in one type of estimand and, if so, how this type of estimand might be defined. **Comments:** This paragraph discusses considerations for non-inferiority trials. Another possible consideration concerns determination of the non-inferiority margin, which is often based on a historical trial comparing the active control to a placebo.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Clarify or discuss what estimand should be used to estimate the active control's effect. There may be some rationale for using the same estimand in the historical trial and the non-inferiority trial.
378	9	**Comments:** Suggest adding "of ICH E9" **Proposed change:** 'Section 3.3.2' -> 'Section 3.3.2 of ICH E9'
378-381	21	**Comments:** For clarity, it is recommended that the term "conservative" be defined.
379	2/3/8/9/12/15/21	**Comments:** Typo: he' = 'the' **Proposed change:** "he importance" -> "the importance"
379	18	**Comments:** In Section 3.3.2 it is stated that such trials are not conservative in nature and the importance of minimising the number of protocol violations and deviations, non-adherence and withdrawals is indicated.
380	9	**Comments:** Suggest adding "of ICH E9" **Proposed change:** 'Section 5.2.1' -> 'Section 5.2.1 of ICH E9'
381	15	**Comments:** "critically" instead of "seriously"?
382	15	**Comments:** The topic of non-inferiority and equivalence could receive more attention, and the comments here only scratch the surface. It could be argued whether in an equivalence (and to a lesser degree also NI) setting the hypothetical or principal stratum estimands would be more appropriate than e.g. a treatment policy estimand.
388	15	**Comments:** …"violations from the target population" – what does that mean? **Proposed change:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		rewording
388-390	21	**Comments:** It would be good to have an example of “a measure of treatment effect with high sensitivity to detect differences between treatments.” More generally, more specific advice on analysis of non-inferiority studies would be helpful.
391	3/8/9/12/13/15/18/19/21	**Proposed change:** The phrase “trial sign” should be “trial design”.
391	15	**Comments:** Concerning trial designs a discussion whether and which measures to minimize intercurrent events would be appropriate (this may depend on the estimand).
393	21	**Comments:** Would it only be the primary trial objectives? As stated later in the addendum it is important that the study is designed appropriately for the secondary (including safety) objectives too. Please provide clarity on what is intended by primary objectives – We interpret this to mean primary and key secondary endpoints for which a labelling claim will be made – is this correct? **Proposed change:** Please clarify that this applies to key secondary endpoints too.
402-407	16	**Comments:** This section discusses the importance of collecting all relevant information to reliably estimate an estimand. Focus is very much on the collection of outcome data despite potential intercurrent events. It is not at all focused on the reliable collection of data on the occurrence of relevant intercurrent events. For any strategy information on intercurrent events is needed and lack thereof (e.g. if patients are lost to follow up information on intercurrent events that may have occurred after dropout is missing) will be a missing data problem different to the extensively discussed problem of missing outcome data. This distinction should be elaborated more clearly in the addendum. **Proposed change:** Differentiate between missing data for the clinical outcome (which will differ depending on the strategy) and missing data for the intercurrent event itself. Hence, focus discussion not only on thorough collection of outcome data but also on thorough collection of all information on relevant intercurrent events. Furthermore, raise the issue of having missing data for the intercurrent event that is a problem different from having missing data on the outcome.
402-	16	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
419		This paragraph contains the important message to collect reliable information on intercurrent events ("distinguish intercurrent events of interest from residual missing data and thus potentially improve the primary analysis"). The section should be improved by stating more explicitly that robustness of combination strategies that treat intercurrent events differently relies on robust information on the intercurrent events themselves. Sponsors propose combination strategies (either pure hypothetical or hypothetical with treatment policy) treating distinct intercurrent events differently with not much focus on the classification of intercurrent events. It should be emphasized more that misclassification could be problematic for combination strategies. This may be viewed independently from the distinction between intercurrent events and missing data and could be regarded a problem of missing information on intercurrent events (see also separate comment on lines 402-407 above).
408	15/21	**Comments:** There is a typo (comma, not colon) **Proposed change:** change to "and, depending on the proportion of missing data, this"
409	21	**Comments**: It seems not easy to distinguish "intercurrent events" and "residual missing data". **Proposed change:** Consider adding more examples of what would be considered "residual missing data".
412-413	13	**Comments:** The example confuses cessation of study treatment with premature end of recording study data. There is no inherent reason to stop documentation when the treatment is discontinued, as it is correctly emphasized in lines 403-407 in the same section. (The document is unclear in lines 85-87 in that regard.) **Proposed change:** An example which directly relates to a plausible reason why data couldn't be collected would be desirable here.
413	12	**Comments:** ACRO recommends that the phrase "treatment discontinuation due to lack of efficacy" should simply be "lack of efficacy" since it is likely that treatment discontinuation would be in the actual case record form (CRF) question. **Proposed change:** Delete "treatment discontinuation due to". **Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Comments: The use of the term 'correctly' in this context is not appropriate, the term 'accurate' is clearer and more correct. **Proposed change:** …perhaps a generic 'loss to follow up' should be more accurately be recorded as 'treatment discontinuation due to lack of efficacy'.
413	21	**Comments:** If "treatment discontinuation due to lack of efficacy" is considered an intercurrent event of interest, it would generally be helpful if the criteria that constitute lack of efficacy were determined in advance. **Proposed change:** Change sentence to "For example, perhaps a generic "loss to follow up" should correctly be recorded as "treatment discontinuation due to lack of efficacy", with lack of efficacy criteria defined in the protocol."
415-419	21	**Comments:** In general, a clinical trial in itself is not reflecting clinical practice. There are in general more visits and more assessments/interventions in a clinical trial compared to normal clinical practice and this in itself is likely to impact retention of subjects. Furthermore, we conduct multi-regional trials and clinical practice differs across regions and countries. We need to adhere to the highest standards in the participating countries. Also, when comparing e.g. two insulins, we need titration targets/schemes in order not to favour one insulin to the other and this titration needs to be monitored closely and action taken if no good explanation exists for a deviation from the titration algorithm. This type of titration is typically not reflecting clinical practice, but something we need to implement to ensure sufficient titration and avoid biased comparison. **Proposed change:** Please consider softening the first sentence and delete "titration schemes" from the example.
420-429	21	**Comments:** Sequential multiple assignment randomized trial (SMART) designs in which patients requiring rescue therapy are re-randomized to a specific rescue therapy, allow for the estimation of treatment policy estimands for a specific frontline and a specific rescue therapy. **Proposed change:** Add SMART designs to the list of examples of non-standard trials in row 421.
421	21	**Comments:** Consider adding cross-over designs (completers)

Line no.	Stakeholder no.	Comment and rationale; proposed changes
422	16	**Proposed change:** Include underlined word: ... regarding their implementation, *interpretation* and …
424	21	**Comments:** It seems that a randomized withdrawal design is described here. Suggest to write "(..) subjects who can tolerate a treatment using a randomized withdrawal design with a run-in period (..)".
425-426	21	**Comments:** The dialogue should also agree on the treatment to be used in the run-in period. More in particular, whether it could be placebo.
430-431	8	**Comments:** Regarding sample size calculations, will an approach using simulations be the preferred approach to take into consideration the expected intercurrent events?
430-434	21	**Comments:** The impact on sample sizes could be discussed more deeply. In many situations it could be difficult to target an effect-size that accounts for the impact and handling of intercurrent events. The impact on sample size and possible inflation should be discussed in such situations. An example would be of interest.
431-434	8	**Comments:** It is stated that “Where all subjects contribute information to the analysis, and where the impact of intercurrent events and their handling is reflected in the effect size that is targeted and the expected variance, it is not usually necessary to inflate the calculated sample size by the expected proportion of subject withdrawals”. However, it is possible that several estimands will be estimated and these may vary with regards to whether they utilise data from all subjects or not, therefore the sample size may need to account for a range of different estimands.
431-434	9	**Comments:** The sentence ‘it’s not usually necessary to inflate the calculated sample size by the expected proportion of subject withdrawal’. This is not true for studies with survival endpoints for which the withdrawal proportion needs to be considered in the sample size calculation.
435 Section 7.2.	8/21	**Proposed change:** Change to Section A.7.2. But also check, is that reference accurate? Did you mean Section A.7.A instead?
435-	21	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
443		This paragraph mentions the need to have consistent definition for the variable of interest, but does not mention the potential impact of different intercurrent events across trials.
444	21	"More generally, a trial is likely to have multiple objectives translated into multiple estimands." **Comments:** Sample size should account for intercurrent events. Therefore, it is not clear how to include intercurrent events in sample size estimation. **Proposed change:** Consideration and guidance on how to address multiplicity issue and impact on sample size should be provided.
444-446	21	**Comments:** It would be appropriate to acknowledge that the treatment policy estimand is not always required. **Proposed change:** "A trial design that is suitable for one estimand might not be suitable for other estimands of potential importance. In addition, the treatment policy estimand is not always required. Trials with multiple objectives and endpoints …"
452-453	16	**Proposed change:** It should rather be ".. should reflect the outcomes in a group of subjects treated with a given treatment as related to the outcomes in the same subjects if they were not treated or treated with another treatment": The phrase used in the document relates to a clinical trial but not to the quantity to be estimated by the trial.
453-454	5	**Comments:** The rationale for the proposed change below is that generic reasons such as "loss to follow up" are often inappropriately selected by investigators due to insufficient emphasis on the importance and lack of specific guidance as to how to accurately record reasons for missing data. Ongoing and monitoring and querying to conform to protocol guidance should help improve the collection of informative reasons for missing data. Propose to add clarity. **Proposed change:** A prospective plan **(including protocol guidance, investigator training, and ongoing monitoring)** to collect informative reasons for why data intended for collection are missing may help
454	21	**Comments:** Clarification is suggested.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: "… so that the effect of treatment can be isolated from any treatment unrelated differences between the groups of subjects on which the comparison is based."
459-461	18	**Comments:** This sentence considers the sensitivity analysis, but power assessments are not mentioned at all. **Proposed change:** The robustness of the results to the underlying assumptions should be assessed through sensitivity analysis and power calculation aligned to the estimand.
462-471	11	**Comments:** The primary concerns around the robustness of estimation of hypothetical estimands (lines 468-471) arise from handling of missing data; without it, statistical assumptions are essentially the same as with treatment policy. However, in real data, even with the best plans possible for patient follow-up after discontinuation/rescue etc., there will always be missing data in later phase clinical trials. Lines 462-464 therefore present an unduly positive and misleading picture of treatment policy (in the presence of missing data, the complexity of analysis and assumptions are arguably greater for treatment policy). Lines 465-468 state that a composite endpoint may need no further statistical assumptions regarding missingness, even when it is present. However, this document encourages the differentiation between missingness and intercurrent events (e.g. lines 483-485), and it is likely that some of the missing data cannot be treated as 'failure' in a clinically reasonable way (e.g. where no intercurrent event had occurred). Therefore, treatment policy and composite strategies may reduce missing data, but in practice are unlikely to lead to its elimination. Wherever there is missing data, assumptions automatically become strong and untestable, and hence all five estimands strategies outlined require the same types of assumption that are so criticised in this document. That these assumptions have to be made is therefore inevitable and should not be criticised as a flaw of any strategy. Where the strategies differ is in the amount of (relevant) missingness that they generate, and therefore the sensitivity and robustness of their analysis to the assumptions. This is a classic missing data (MNAR) issue which should be handled (for all primary analyses) by sufficient sensitivity analysis. **Proposed change:** Cover the problem of missing data requiring strong, untestable, assumptions without reference to specific estimands (since it applies to them all). Focus on use of sensitivity analyses to assess robustness of results to these assumptions. It would be fair to state that some estimands are more robust to deviations in these assumptions than others.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
462-471	21	**Comments:** The primary concerns around the robustness of estimation of hypothetical estimands (lines 468-471) arise from handling of missing data; without it, statistical assumptions are essentially the same as with treatment policy. However, in real data, even with the best plans possible for patient follow-up after discontinuation/rescue etc., there will always be missing data in later phase clinical trials. Lines 462-464 therefore present an unduly positive and misleading picture of treatment policy (in the presence of missing data, the complexity of analysis and assumptions are arguably greater for treatment policy). Lines 465-468 state that a composite endpoint may need no further statistical assumptions regarding missingness, even when it is present. However, this document encourages the differentiation between missingness and intercurrent events (e.g. lines 483-485), and it is likely that some of the missing data cannot be treated as 'failure' in a clinically reasonable way (e.g. where no intercurrent event had occurred). Therefore, treatment policy and composite strategies may reduce missing data, but in practice are unlikely to lead to its elimination. Wherever there is missing data, assumptions automatically become strong and untestable, and hence all five estimands strategies outlined require the same types of assumption that are so criticised in this document. That these assumptions have to be made is therefore inevitable and should not be criticised as a flaw of any strategy. Where the strategies differ is in the amount of (relevant) missingness that they generate, and therefore the sensitivity and robustness of their analysis to the assumptions. This is a classic missing data (MNAR) issue which should be handled (for all primary analyses) by sufficient sensitivity analysis. **Proposed change:** Cover the problem of missing data requiring strong, untestable, assumptions without reference to specific estimands (since it applies to them all). Focus on use of sensitivity analyses to assess robustness of results to these assumptions. It would be fair to state that some estimands are more robust to deviations in these assumptions than others.
464-465	7	**Comments:** It is correct that "Estimation for an estimand … will require stronger and untestable assumptions if measurements are not collected following intercurrent events." Therefore, every effort should be made to collect all relevant data after the occurrence of an intercurrent event. **Proposed change:** Please add the statement that every effort should be made to collect all relevant data after the occurrence of an intercurrent event.
464-465	13	**Comments:** It is correct that "Estimation for an estimand … will require stronger and untestable assumptions if measurements are not collected following

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		intercurrent events." **Proposed change:** Add a statement about the importance of collecting relevant data after occurrence of intercurrent events in order to avoid this situation.
464-465	17	**Comments:** It is correct that "Estimation for an estimand … will require stronger and untestable assumptions if measurements are not collected following intercurrent events." Therefore, any effort should be made to collect all relevant data after occurrence of an intercurrent event. **Proposed change:** Add the statement that any effort should be made to collect all relevant data after occurrence of an intercurrent event.
467	15	**Comments:** difficult sentence **Proposed change:** reword or delete “and the associated assumptions”
467-468	21	**Comments:** Not a logical sentence, edited for clarity. **Proposed change:** Using a composite strategy it may be possible to perform an analysis without need for imputation or modelling of response after an intercurrent event and without the associated assumptions of such modelling or imputation.”
468-469	21	**Comments:** Did the author intentionally distinguish between "a strategy that requires a hypothetical scenario" and "the hypothetical strategy"?
468-471	7	**Comments:** It is correct that "… the estimation of estimands constructed using a strategy that requires a hypothetical scenario to address an intercurrent event entails careful specification of the hypothetical conditions and will necessarily rely on modelling assumptions that are untestable …". Therefore, the corresponding analysis should not be used as the main analysis for decision-making. **Proposed change:** Please add the statement that methods relying on strong untestable assumptions should not be used as the main analysis for decision-making.
468-471	17	**Comments:** It is correct that "… the estimation of estimands constructed using a strategy

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		that requires a hypothetical scenario to address an intercurrent event entails careful specification of the hypothetical conditions and will necessarily rely on modelling assumptions that are untestable …". Therefore, the corresponding analysis should not be used as main analysis for decision making. **Proposed change:** Add the statement that methods relying on strong untestable assumptions should not be used as main analysis for decision making.
471	21	**Comments:** The sentence starting “In a randomized….”. Unclear what this is telling me and not sure if it’s correct in all cases. **Proposed change:** Rephrase or remove this sentence.
472-473	7	**Comments:** It is correct that "… estimation of a treatment effect within a principal stratum of the population will be confounded unless the subjects within that stratum can be identified before randomisation." If the subjects can be identified before randomisation, the principal stratum strategy is nothing more than a conventional subgroup analysis. If this is not the case, the principal stratum strategy can only be used as a supplementary analysis but not as the main analysis for decision-making. **Proposed change:** Do not use the term "principal stratum strategy" for situations of a conventional subgroup analysis. In all other cases, do not describe the principal stratum strategy as an option for the main analysis.
472-473	17	**Comments:** It is correct that "… estimation of a treatment effect within a principal stratum of the population will be confounded unless the subjects within that stratum can be identified before randomisation." If the subjects can be identified before randomisation the principal stratum strategy is nothing else than a usual subgroup analysis. If this is not the case, the principal stratum strategy can only be used as supplementary analysis but not as main analysis for decision making. **Proposed change:** Do not use the term "principal stratum strategy" for situations of a usual subgroup analysis. In all other cases, do not describe the principal stratum strategy as an option for the main analysis.
472-473	20	**Comments:** “unless the subjects within that stratum can be identified before randomization”: Is it correct to interpret that this suggests that in cases

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		where they can be identified, only patients within the principal stratum should be randomized? And then the principal stratum and the full study population are identical? Please clarify.
472-473	21	**Comments:** "In a randomised trial, estimation of a treatment effect within a principal stratum of the population will be confounded unless the subjects within that stratum can be identified before randomisation." **Comments:** It is unclear what the sentence is supposed to illustrate.
474-476	21	**Comments:** A comparison of adherers on drug vs. control could be informative. Regardless of the reasons for non-adherence a notable difference in adherence to therapy is an important observation in considering the utility of the investigational drug especially in a study with an active comparator. The fact that they are different and what makes them different may also be informative for treatment decisions should the investigational drug eventually be authorised for use. **Proposed change:** Clarify that it is only inappropriate to compare outcomes in these two strata.
477-479	21	**Comments:** These lines seem to suggest that a "preconceived" set of covariates is undesirable, but pre-specification of model terms is generally regarded as good statistical practice. **Proposed change:** Clarify whether the covariate list should be pre-specified or not.
479	21	**Comments:** Clarification regarding the intended meaning of the word 'labelled' is requested.
479-481	16	**Comments:** Isn't it rather "interpretation of effect is difficult if intercurrent event rate and timing is different between arms" instead of "strong assumptions are needed"? **Proposed change:** Change sentence to "For the labelled while-on-treatment strategy interpretation of the treatment effect estimate is difficult when the occurrence and timing of the intercurrent event is related to treatment."
480	21	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Suggest spelling out what you mean by “stronger assumptions”.
482-497	16	**Proposed change:** Reference should be made to missing data imputation methods (in general)
483-485	2	**Comments:** while ‘data’ are always plural, ‘information’ is always singular so replace ‘are’ with ‘is’ after information.
484	15	**Comments:** delete “that are”
491-496	21	**Comments:** Prediction model based on other patients who discontinued treatment but for whom data collection continued is probably a good approach, but requires that a sufficient number of patients are available to build a stable model. If we assume for example a 10% discontinuation rate, and 50% of these patients have data available, this would lead to build a model from 5% of patients, which will likely be unstable and result in highly variable predictions. Would control-based imputations or imputations based on external data be considered an acceptable approach, providing the underlying assumptions are justifiable?
498-536	12	**Comments:** It is unclear in the document whether sensitivity analyses are intended to serve a different purpose for estimands compared to the usual purpose of assessing the robustness of primary conclusions. If so, it would be useful to highlight the differences. **Proposed change:** Clarify the purpose of sensitivity analyses with regard to estimands and highlight any differences from the usual use of a sensitivity analysis.
498-536	16	**Comments:** In addition to sensitivity analyses for a given estimand, the investigation of the operating characteristics, e.g. type-1 error control of the proposed estimation functions (“estimators”) whether primary or used as sensitivity analyses is important in many settings due to the uncertainty with respect to relevant assumptions on which the estimation function is based upon. **Proposed change:** Add a paragraph on the upfront evaluation of the operating characteristics (especially bias and type-1 error) of the used estimators whether primary or used as sensitivity analyses.
500-501	21	**Comments:** This statement may be misinterpreted to mean that estimates used for

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		inference (e.g., hypothesis testing) should be 'absolutely' robust. The need, often, to balance reliability and relevance may lead to using an estimate with high relevance but not the highest reliability as the best available option. Qualification of the term 'robust' is thus recommended. **Proposed change:** "Inferences based on a particular estimand should be adequately robust to limitations in the data and deviations from the assumptions used in the statistical model for the main estimator.
503-511	20	**Comments:** This paragraph contains important concepts, but seems too concise to be completely clear. The relationships among primary estimands, secondary estimands, main estimators, sensitivity analyses, supplementary analyses, etc., and their relative priority for interpretation, should be described more clearly to avoid any interpretation on the part of a reader.
506-508 870-873	16	**Comments:** Two different types of analyses are discussed: sensitivity analysis and supplementary analyses. To aid clarity and avoid misunderstandings both types of analyses should be distinct classes. Considering one estimand, this would be the case when sensitivity analyses are defined as analyses that still address this one estimand but use different sets of assumptions and when supplementary analysis are defined as analysis address different estimands (for the same outcome and intercurrent event). This distinction is not clear enough in the addendum and especially the description of supplementary analysis in the glossary is too imprecise by just stating that supplementary analysis are a broader set of analyses than sensitivity analyses. **Proposed change:** Clearly highlight that sensitivity and supplementary analysis are distinct classes: the former addressing the primary estimand and the latter addressing different estimand than the primary. Also the definition of supplementary analyses in the glossary should be adapted correspondingly: "**Supplementary Analysis:** Is a general description for analyses that are conducted in addition to the main and sensitivity analysis aligned to the target estimand to provide additional insights into the understanding of the treatment effect. The term describes a broader class of analyses that evaluate different estimands than the main and sensitivity analyses."
508-509	16	**Comments:** "Each supplementary analysis may refer to a different estimand, or a different

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		estimator to the same estimand." It is unclear whether this sentence means "supplementary analysis address different estimand than primary or address different estimator to the primary estimand" or "supplementary analysis address different estimand than primary or a different estimator for this different estimand". It should be the latter. **Proposed change:** Change to "Each supplementary analysis may refer to a different estimand than the main and sensitivity analysis, or refer to another estimator to this different estimand."
508-509	21	**Comments:** "Each supplementary analysis may refer to a different estimand or a different estimator of the same estimand." The last part is confusing, as estimators focusing on the same (i.e., main) estimand can better labelled as "sensitivity estimators".
509	15	**Comments:** According to text, a supplementary analysis could also be one using a different estimator for the same estimand. This is rather close (identical?) to the definition of a sensitivity analysis, so: every sensitivity analysis is also a supplementary analysis, but not vice versa?
509	21	**Comments:** It should be useful to add an example of different estimators to the same estimand for a supplementary analysis. In our understanding the use of different estimators for the same estimand should be considered only as sensitivity analyses so maybe the sentence should be reworded.
519-536	12	**Comments:** ACRO recommends that it would be helpful to re-explain here how sensitivity analyses are used to assess the robustness of primary conclusions, e.g. clarify that not all sensitivity analyses have to reach significance but there should be consistency in the treatment effect estimate. It is unclear how tipping point analyses help to assess consistency of the conclusion on the effect size and more explanation would be helpful. **Proposed change:** Add text to re-explain how sensitivity analyses are used to assess the robustness of primary conclusions, and provide more explanation of how tipping point analyses help to assess consistency of the conclusion on the effect size.
525	15	**Comments:** One could also argue to the contrary, that the characterization of "relevant"

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		deviations will always be challenging. **Proposed change:** delete the statement
526	20	**Comments:** What is the “original analysis” referred to here – an analysis of all the non-missing cases? Please clarify and re-phrase.
533	15	**Comments:** It is unclear how a degree of clinical relevance can be quantified (if at all) in relation to a treatment effect difference. A specific method for this purpose should be given as an example.
537-559	21	**Comments:** It is unclear what is the difference between supplementary analysis (eg. targeting a different estimand from the same variable or endpoint) and estimand for secondary trial objectives (eg. for a secondary variable or endpoint) Could it be clarified whether it is expected that results from supplementary analyses should confirm the conclusions from the primary analysis (to match what was done before with FAS and PP set), or that since it is addressing a different question, different results may be expected.
546	9	**Comments:** Suggest adding “of ICH E9” **Proposed change:** ‘Section 5.2.3’ -> ‘Section 5.2.3 of ICH E9’
546-559	15	**Comments:** This section can be expected to have immediate major impact on protocol development concerning the judgement of adequacy of (primary) PPS analysis, regardless of whether the estimands concept can be adopted in the particular trial or not. **Comments:** This paragraph appears to make a case against a per protocol analysis, stating that the value of the PPS analysis is limited to investigating whether the extent of protocol deviations compromises confidence in the trial results. To us this still appears a worthwhile objective. In general, one should add a discussion about the merits of analyses that are conservative for some clinically relevant estimand, but by themselves do not provide unbiased estimates of this estimand.
546-559	17	**Comments:** The sections seems to indicate that it will not be necessary to do statistical

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		analyses of the PP analysis set. Does this also hold for non-inferiority trials? **Proposed change:** Consider to have clear guidance concerning this
548-550	21	**Comments:** This section states that consistent results between the full analysis set and per protocol set increase confidence in the study results, but also states that the per-protocol results can have “severe bias”. It is not clear why consistency with a potentially severely biased estimate would be reassuring, or conversely, why lack of such consistency would make one less confident about a study’s results. **Proposed change:** Resolve this apparent contradiction.
549-559	20	**Comments:** The recommendation regarding PPS seems a bit vague. It seems that PPS analyses are discouraged; but is the suggestion that they should no longer be performed, or at best, should have very limited interpretation? And therefore that ICH-E9 is outdated in this regard? As alluded to in lines 553-5, PPS analyses make the unverifiable assumption of treatment comparability, but other approaches that seem preferred in this document may also be subject to biases if their assumptions are not satisfied. Also, as the use of PPS in non-inferiority studies is widespread, it would seem that recommendations in that setting should be explicitly mentioned.
556-559	21	**Comments:** The use of a per-protocol analysis, especially in the context of non-inferiority/equivalence trials seems to be revisited in this addendum. Is a per-protocol analysis still considered relevant? If not, should the Section 5.2.3 in the original ICH E9 be amended?
560	21	**Comments:** Suggest in this section, more realistic examples could be given, appropriate for each strategy, laid out in such a way which could be plausibly used in a protocol template; that would promote good practice. E.g. stating attributes of the estimand, assumptions and how they will be investigated.
564	21	**Comments:** “protocol and the analysis plan” Does this refer to two separate documents still (i.e. protocol and SAP); does this mean that the detailed descriptions would be contained within the SAP? Or does this mean that the protocol now needs to have a more robust analysis plan section (beyond describing attributes of the estimand).

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Clarification should be provided regarding the details to be provided in the protocol versus the SAP.
564-569	21	**Comments:** Is it really a requisite to have the sensitivity analysis fully specified in the protocol? It may be reasonable to provide high level elements in the protocol and then provide technical details in the SAP? Same comment also applies to other sections of the addendum.
566-569	21	**Comments:** Sensitivity analysis for secondary analysis may not always be required. **Proposed change:** each with a corresponding main estimator and a suitable sensitivity analysis", suggest to change to "each with a corresponding main estimator and if appropriate a suitable sensitivity analysis.
567-569	12	**Comments:** The sentence should be expanded to explain whether sensitivity analyses are required for all secondary endpoints. **Proposed change:** Explain whether sensitivity analyses are required for all secondary endpoints.
567-575	21	**Comments:** What is meant by properly documented for estimands other than the primary? Please clarify which should be specified in the protocol and which could be left for the analysis plan.
569-570	17	**Comments:** The text suggest that even explorative analyses should be described by estimands, that seems like a lot of documentation to go into for example a protocol **Proposed change:** Suggest to clarify that only analyses to support claims (primary, key secondary) should be documented to the level of estimands
569-570	21	**Comments:** The text suggest that even explorative analyses should be described by estimands, that seems like a lot of documentation to go into for example a protocol **Proposed change:** Suggest to clarify that only analyses to support claims (primary, key

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		secondary) should be documented to the level of estimands
571-575	18	**Comments:** Our main question here is why it is not a requirement to specify in detail an estimand for each pre-planned exploratory question, since for statistical purposes each variable must be clearly defined so no doubts are raised, especially when there are minor variations on primary or secondary estimands.
574-575	9	**Comments:** Please provide an example.
578-581	5	**Comments:** Suggest omitting the definitive recommendation of changing one assumption at a time. In sake of simplicity it may good to start with changing one assumption at a time however then to change more than one assumption and see if there are interactions that significantly change outcomes. Statistical analyses tools should be selected in order to properly handle such simultaneous changes and yet enable to realize the effects of the individual factors and their interactions. **Proposed change:** “When planning and conducting a sensitivity analysis, it is useful to investigate the impact of changing only one assumption at a time. It is recommended however to consider changes of several assumptions simultaneously in case some interactions among those assumptions significantly affect estimands.”
579-580	21	**Comments:** A statement such “Beyond these aspects, the conventional considerations for trial design, conduct and analysis remain the same. ” could appear earlier in the addendum. **Proposed change:** A similar statement may be introduced in Section A.1. (for example after line 70-71. It could set the stage for the whole addendum.
580-582	21	**Comments:** It is not clear when one needs to account for multiple testing in the context of additional estimands. Do you consider sensitivity analyses for missing data a source of multiplicity? **Proposed change:** Clarify when multiple estimands lead to a need for type I error control.
584	21	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		It is not fully clear what “pre-specified” means. Do you mean “before first patient first visit” (implemented in trial protocol or SAP). Often, “pre-specified” is interpreted as being before breaking the blind.
584-585	12	**Comments:** It is not clear why “analyses introduced while the trial was still blinded” are referenced. Such changes should be pre-specified and taken into account in the statistical plan. **Proposed change:** Delete “analyses introduced while the trial was still blinded”.
585-586	21	**Comments:** The added text adds more information to the sentence to help the reader better understand the intercurrent event concept. **Proposed change:** Intercurrent events that were not foreseen at the design stage but were identified during the conduct of the trial should be discussed to specify both the way the intercurrent events were handled in the analysis and the effect they had on the chosen analysis estimates and the interpretation of the trial results.
585 - 588	15	**Comments:** Difficult to understand. Please reword.
589 Section 7	15	**Comments:** There is redundancy at some text passages relating to missing data, sensitivity and additional supplementary analyses. Suggestion to screen for redundancies, and to shorten; A.7.1 gives examples of the 5 strategies using a very similar structure and wording, which is repeated for each of them. While repetition is helpful on the one hand, the specifics/ differences could be highlighted (e.g. by using Italics or in bold) to be more obvious. While it is understandable that the 5 strategies are presented only in a very abstract way, real examples would be illustrative as well. Moreover, the potential strengths and weaknesses of each strategy could be better presented. Graphs for each of the scenarios might be helpful as well. It would be most helpful if examples are already presented in the way that is outlined in the guidance. For example: • Main estimator and sensitivity analyses should be specified (Section A.2). This should include all necessary details, for example whether subjects enter the analysis as randomized or as treated. • Necessary assumptions to ensure a conservative and efficient analysis

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		should be documented (Section A.5.2.1). In this regard it appears strange why ANOVA and not ANCOVA is suggested for the analysis of change to baseline variables. Moreover, for the while on treatment strategy it is unclear whether the usual assumptions of ANOVA (variance homogeneity) are plausible. Similar, in the composite strategy example the population summary is the difference in response proportions. However, logistic regression which estimates the odds ratio is suggested as the appropriate analysis.
589	21	**Comments:** Section A.7. appears to include only conceptual statements indicating the interpretation of the estimand without specifying how the intercurrent event is handled in the estimators. Additional detail on this aspect would be helpful.
589-627	12	**Comments:** ACRO recommends that, for consistency, it would be helpful to use the flow diagram on page 5 of the document to present this and the other examples. **Proposed change:** Use the flow diagram on page 5 of the document to present this and the other examples.
593	21	**Comments:** Delete this paragraph unless it can be specified how this is related to estimands.
597	15	**Comments:** It appears as if the assertion that a randomised clinical trial best addresses the scientific question is independent of the estimand but that is contrary to the previous suggestion that the definition of a meaningful estimand should precede the choice of appropriate study design.
604-624?	21	**Comments:** This section reads as a non sequitur immediately following the introduction to the vignette. **Proposed change:** Break this example out into its own section, analogously to Sections A.7.1 and A.7.2 for one and two intercurrent events (i.e. this one would be an example with zero intercurrent events).
606-607	20	**Comments:** Please clarify how it is recommended to handle patients who are erroneously enrolled despite not satisfying the inclusion/exclusion criteria.
608, 636,	21	**Proposed change:** For clarity, consider changing the phrase "change from baseline to month

Line no.	Stakeholder no.	Comment and rationale; proposed changes
695, 718		six..." to "change from baseline at month six…"
610	12	**Comments:** It is not clear why the term “treatment conditions” is used rather than “treatment groups”. If “treatment conditions” is retained, it should, for clarity, be added to and defined in the glossary. **Proposed change:** Replace “treatment conditions” with “treatment groups” or add and define “treatment conditions” in the glossary.
610	20	**Comments:** Please clarify whether “difference in variable means” refers to the raw means, or to estimates of model parameters.
614-624	13	**Comments:** This part is repeated by L652-L661 (ending “…the treatment groups.”). The paragraph L614ff seems to be the one that is in the wrong place since the section considers the case where no intercurrent events occur, however, intercurrent events and rescue switchers are mentioned in L621 and L623. **Proposed change:** Consider inserting a headline “A.7.0 No intercurrent event” between lines 603 and 604, and suitable adaptation of L614ff.
615	7	**Comments:** The method for statistical analysis is described as "… analysis of variance model with treatment group as a factor …". In the situation considered, the corresponding ANOVA model is reduced to the conventional *t*-test. **Proposed change:** Please replace "analysis of variance model" by "*t*-test".
615	17	**Comments:** The method for statistical analysis is described as "… analysis of variance model with treatment group as a factor …". In the considered situation the corresponding ANOVA model reduces to the usual *t*-test. **Proposed change:** Replace "analysis of variance model" by "*t*-test".
615, 652, 682	13	**Comments:** The given examples for recommended statistical methods (analysis of variance and logistic regression) are trivial for the respective situations. Instead of giving recommendations for trivial situations, recommendations for statistical methods should be given for situations where the choice of

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		appropriate methods is really unclear (hypothetical strategy, principal-stratum strategy). **Proposed change:** Recommendations for the choice of statistical methods for complex situations as the hypothetical strategy and the principal-stratum strategy should be given.
616-617	20	**Comments:** It is stated that missing data need to be predicted; does this allow the possibility that the missing data could ever be ignored, and not predicted, if there is a justifiable fully random mechanism for the missingness, e.g., analytical failure, accidental death, etc.? This may at times be an assumption more realistic than those required for other approaches recommended in this document.
616-617 also 756	20	**Comments:** It is explicitly stated that prediction / imputation / interpolation is required. But not all relevant methods require explicit prediction of all missing values (e.g., modeling approaches). Shouldn't this statement be modified to be a bit more general?
617-618	2	**Comments:** The missing at random assumption is rarely defensible in randomised trials. The comment here appears to be recommending this approach. Prof Ian White's work on missingness would appear to challenge the idea of undertaking MI based upon MAR assumptions in randomised trials.
617-618	18	**Comments:** We think this sentence should be rewritten so an example of imputation is considered. **Proposed change:** For instance, missing data may be imputed based on similar subjects who remained in the trial, for example, by Random Forest procedures.
616-620	21	**Comments:** When it comes to missing data, the addendum focuses on imputation and interpolation methods. Other methods such as MMRM are not mentioned. **Proposed change:** It will be helpful to include examples on what missing data methods could be applied as analyses for each intercurrent event strategy.
617/654/789/	21	**Comments:** Is the first sentence referring to how the estimate reflects uncertainty in

Line no.	Stakeholder no.	Comment and rationale; proposed changes
823		imputations? ”In the case of missing measurements, data need to be predicted based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial.” **Proposed change:** replace “data need to be predicted” with “the estimate needs to be predicted” or “data need to be imputed” and add these terms to the glossary.
617-618	21	**Comments:** Minor alteration of the sentence because we think that only using “similar patients …” without further accounting in the analysis would not be a correct approach, however one which might be often incorrectly used. Therefore, we added “also”. **Proposed change:** …missing data may be imputed based on also using similar subjects who remained in the trial.
618-620	21	**Comments:** “…and information on the intercurrent event”: Discrepancy to line 609 (“…no intercurrent events to be taken into account”) Please delete. **Proposed change:** Similarity may be established based on the same baseline covariates, the same randomised treatment arm, and the same measurement history. and information on the intercurrent event.
620-622	21	**Comments:** This sentence suggests that “Sensitivity analyses should be pre-specified in the trial protocol”. However, in the past, sensitivity analyses were more commonly defined in the statistical analysis plan rather than the protocol. **Proposed change:** Change to “Sensitivity analyses should be pre-specified in the trial protocol or the statistical analysis plan”. **Comments:** In an example of trial without intercurrent events, the guideline says: “For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity, may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event.”. The last reference to an intercurrent event appears a typo in this example without intercurrent events.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
626, 627	12	**Comments:** This sentence appears unnecessary given line 609 and details included in glossary. **Proposed change:** Delete the sentence.
626-627	20	**Comments:** It's not clear what this sentence means and why is it needed. We suggest removing it, or re-phrasing.
626-627	21	**Comments:** Move this sentence to line 611 where it makes the meaning clearer.
629 and beyond	8	**Comments:** For each of the example cases it would be good to include the text description of the objective related to the estimand. This would be very helpful as a guide to include in protocols. **Proposed change:** Include text descriptions as indicated: e.g. *Estimand:Treatment policy strategy* *"Compare experimental drug X and placebo in terms of improving endpoint Y at* *6 months for all randomized patients without regarding adherence to randomized treatment"* *Estimand:Hypothetical* *"Compare experimental drug X and placebo in terms of improving endpoint* *Y at 6 months for all randomized patients as if all patients had remained in the trial* *and received treatment as planned without rescue medication until 6 months"* *Estimand:Composite* *"Compare experimental drug X and placebo in terms of a clinical responder* *at 6 months (Responder defined as achieving a pre-specified threshold of endpoint Y* *and not requiring rescue)"*
629 and beyond	21	**Comments:** For each of the example cases it would be good to include the text description of the objective related to the estimand. This would be very helpful as a guide

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		to include in protocols. **Proposed change:** Include text descriptions as indicated: e.g. Estimand:Treatment policy strategy “Compare experimental drug X and placebo in terms of improving endpoint Y at 6 months for all randomized patients without regarding adherence to randomized treatment” Estimand:Hypothetical “Compare experimental drug X and placebo in terms of improving endpoint Y at 6 months for all randomized patients as if all patients had remained in the trial and received treatment as planned without rescue medication until 6 months” Estimand:Composite “Compare experimental drug X and placebo in terms of a clinical responder at 6 months (Responder defined as achieving a pre-specified threshold of endpoint Y and not requiring rescue)”
633	5	**Comments:** One challenge in designing non-inferiority trials is the choice of primary analysis set between Full Analysis Set (FAS) and Per-Protocol Set (PPS). Guidance E9 states: "In an equivalence or noninferiority trial, use of the full analysis set is generally not conservative and its role should be considered very carefully." This addendum creates a great opportunity to reduce the differences between the FAS and PPS according to the definition of the estimand. This opportunity is suggested in lines 629-632: “Where a majority of intercurrent events are handled through the construction of the estimands, the number of remaining protocol violations and deviations will be low and analysis of the PPS might not add additional insights.” It may be helpful to elaborate on this point with examples, such as those suggested below. **Comments:** Additional wording starting at line 633 could include: "For example, for an estimand based on treatment policy strategy or composite endpoint strategy, certain intercurrent events not resulting in missing data such as use of rescue medication might not be considered protocol violations in deriving the PPS. Similarly, for an estimand based on while-on-treatment strategy, certain pre-defined intercurrent events resulting in treatment discontinuation would not be considered protocol violations in deriving the PPS."

Line no.	Stakeholder no.	Comment and rationale; proposed changes
634-662	21	**Comments:** This is a good example that makes mention of a trial design (2-arm parallel) and model (ANOVA) when all rescue therapies are considered equal and no adjustment for rescue therapy is made. This addendum/supplement would benefit greatly by also including an example where not all rescue therapies are considered equal, and the distribution of the types of rescue therapy taken differs by initial treatment group. **Proposed change:** Discuss adjustment for rescue therapy and SMART designs.
635	5	**Comments:** This sentence indicates that estimands need to be completely specified in the protocol, however elsewhere in the document it specified that there are potentially unexpected intercurrent events that could occur. It would appear practically in this situation that the protocol would need to be amended each time, which is an extra burden and may be impractical. **Proposed change:** As a matter of practice, foreseen intercurrent events and estimands should be prespecified in the protocol, however the final estimand should be completely specified in the statistical analysis plan, which may differ slightly from the protocol.
640	21	**Comments:** In this case, treatment policy is the wrong phrase as nobody would ever be prescribed placebo. So in this case, the true treatment effect will be attenuated due to an increased level of switching from placebo; this is the wrong message to give patients. **Proposed change:** Clarify the effect being measured is relating to the clinical trial but would not be reflective of clinical practice.
648	21	**Comments:** Minor alteration of the sentence for greater consistency with the content in the paragraph. **Proposed change:** A similar sentence can be constructed for each of the examples strategies below,...
652	8/21	**Comments:** In the Treatment Policy estimand example, a possible method of analysis for obtaining the population-level summary is an Analysis of Variance. Would this be considered to be the 'estimator' – if so, it might be helpful to include that

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		clarification here.
652-661	20	**Comments:** This paragraph exactly duplicates lines 615-624, and largely repeats lines 682-689 as well. It is confusing when contents are repeated even when applied to different strategies. We recommended that some repetition be reduced to improve of the ease of reading.
652 - 662	15	**Comments:** Repetition of a whole paragraph from previous chapter (614 – 624) **Proposed change:** Shorten and cross-refer
653-656	8/21	**Comments:** It will be helpful if Treatment Policy Strategy example is elaborated to cover the case where the data are missing after the intercurrent event
655-656	21	**Comments:** “For example, missing data may be imputed based on similar subjects who remained in the trial” This line is redundant to Line 617-618 (“For example, missing data may be imputed based on similar subjects who remained in the trial”) **Proposed change:** Suggest deleting one or the other.
655-657 and 685-687	12	**Comments:** The sentence “For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event” is repeated again on the following page. **Proposed change:** Cross-reference to reduce repetition within the document.
675	21	**Comments:** Is this guidance to now collect reasons for missing data? Lack of efficacy cannot be assumed. **Proposed change:** Clarify the importance of collecting more specific details relating to why data may be missing and/or reasons for switching.
682	7	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		In the situation considered, the use of logistic regression is not required. A simple 2x2 table with an adequate statistical test would be sufficient. **Proposed change:** Please replace "logistic regression" by "2x2 table with an adequate statistical test".
682	17	**Comments:** In the considered situation the use of logistic regression is not required. A simple 2x2 table with adequate statistical test would be sufficient. **Proposed change:** Replace "logistic regression" by "2x2 table with adequate statistical test".
682	21	**Comments:** Section is highly repetitive of other sections. **Proposed change:** Consider aligning paragraph to reduce text
683	21	**Comments:** Can MAR assumption be used for intermittent missing values, i.e., missing data before the event?
686-688	21	**Comments:** Alteration of the sentence to be consistent with other examples **Proposed change:** Similarity may be established based on the same baseline covariates, the same randomised treatment and the same measurement history and information on the intercurrent event.
688	13	**Comments:** The prospective planning of sensitivity analyses should be standard and is contributing to the validity of the interpretation of results. However, it is nearly impossible to pre-empt all possible situations of missingness. Rather than planning all sensitivity analyses in the protocol upfront it might be considered to explain the analyses strategy. **Proposed change:** Please change accordingly.
692-713	7	**Comments:** We question the usefulness of a hypothetical setting in which it is assumed that rescue medication was not available. No regulatory decisions should be based upon such an analysis.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Please clearly describe that an analysis in hypothetical settings may be used as a supplementary analysis in special situations.
692-713	9	**Comments:** This section needs a specific example of how hypothetical strategy would be done.
692-713	17	**Comments:** I question the usefulness of the hypothetical setting to assume that rescue medication was not available. No regulatory decisions should be based upon such an analysis. **Proposed change:** Describe clearly that an analysis in hypothetical settings may be used as supplementary analysis in special situations.
692-713	21	**Comments:** For the hypothetical strategy, it would be helpful to give more details on how the underlying assumptions would be spelled out (in this hypothetical setting) and which type of sensitivity analyses could be conducted.
698-700	21	**Comments:** It should be clarified that the clinical relevance of this estimand relies on an untestable assumption: specifically, that the treatment has no impact on the effectiveness of the rescue medication. Suppose hypothetically that the treatment affects the patient in such a way that the rescue medication, which provides benefit in most situations, actually causes harm in this situation. This would be critical clinical information (since presumably the rescue medication will be used in clinical practice), but something ignored by the proposed estimand. The effect described here could lead to an average ITT effect of 0 and an average non-ITT effect > 0. If such a hypothetical situation is possible, and if the ITT effect is the more clinically relevant in that situation, then ITT should be chosen estimand. This same issue could be described for most or all of the non-ITT estimands. **Proposed change:** Clarify the clinical relevance of estimand using the hypothetical strategy.
701-703	21	**Comments:** In this scenario, data through month six on subjects who switch to rescue medication may still be useful to predict the measurements under the hypothetical estimand in a sensitivity analysis, depending on the assumptions. For example, one assumption may be that a subject's response at six months under the hypothetical of not being offered rescue medication is less than or equal to the response under the scenario of being offered rescue medication.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Under this assumption, the observed response at six months on rescue medication is useful in predicting the response under the hypothetical of not being offered rescue medication. **Proposed change:** Reword to make it clear that data collected after the intercurrent event may be useful in some cases even when a treatment policy estimand itself is not of interest.
701-704	21	**Comments:** “There would be no need to collect measurements after switching to rescue medication, unless there is interest in alternative trial objectives that would require such data (e.g. to collect safety information even after the intercurrent event)”. Some of the structural model based approaches would require collection of these data. **Proposed change:** Change to “There would be no need to collect measurements after switching to rescue medication, unless there is specific interest related to alternative trial objectives (e.g. to collect safety information even after the intercurrent event) or statistical methodology requiring the use of structural model based approaches”.
704-705	21	**Comments:** Upon a first reading, this might look like there is a typo here and it should read “is regarded as missing”. Does this imply that immaterial data that are removed (or not collected) should not be treated as missing in the resulting analysis/imputations? If data are not considered relevant then should we introduce terminology for this and add to glossary (see next comment) e.g. immaterial data (or could use alternative words such as extraneous, inapplicable, redundant, irrelevant).
705-708	21	**Comments:** We suggest to add: “A statistical analysis [..] to subjects. If values are collected after the event they must not be used in the analysis.”
706-708	9	**Comments:** How would this be done in practice? Please provide an example.
706-709	21	**Comments:** Regarding the sentences "A statistical analysis for this estimand will rest on assumptions about the measurements that would have been observed under the hypothetical setting where rescue medication was not available to subjects. Generally, the assumptions needed for such predictions cannot be

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		verified based on the observed data so that a sensitivity analysis will be necessary to assess the robustness of conclusions.", it is unclear what kind of primary analysis (i.e. estimator) is expected. **Proposed change:** Because this part is a general example, it would be better to include some examples of the estimator in hypothetical strategy (e.g. MMRM, MI, WGEE, etc) as in treatment policy strategy.
707-708	8/21	**Comments:** It will be helpful to provide some examples of prediction for the hypothetical strategy
708	21	**Comments:** We propose addition of an example to help the reader better understand the concept. **Proposed change:** For example, using a Random Coefficients Model extrapolates data in a line based on observed data available for each patient.
708-711 735-737	13	**Comments:** For the hypothetical and the principal-stratum strategy, the choice of the appropriate statistical methods for the main analysis and the extent and type of expected sensitivity analyses are unclear. **Proposed change:** Advice on appropriate methods for analysis and examples for situations where these estimands are regarded as suitable should be given.
714	21	**Comments:** Would it be relevant to provide an example that matches with the strategy proposed?
714-739	21	**Comments:** We would appreciate clarity on how to handle subjects who have an intercurrent event, despite the population of interest being chosen to avoid this e.g., subjects are only included if no rescue medication was required during a run-in period, but a subject still takes rescue medication during the trial
715-716	20	**Comments:** It seems questionable to exclude patients from a population of interest because they would have required rescue on an alternate treatment. Patients who do not need rescue on an investigational treatment, but would have needed rescue on a comparator, are usually of particular interest, and may

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		often be the patients who would benefit most.
718, 746-747	21	**Comments:** Lines 746-747 indicate that the intercurrent event is the use of rescue medication. However, this is not clear in the definition of estimand (see Line 718). **Proposed change:** “B. Variable: change from baseline to month six in the designated measurement, and no switching to rescue medication occurred.”
724-725	7	**Comments:** It is not difficult to identify members of this hypothetical population in advance; it is, in general, impossible. **Proposed change:** Please describe that it is, in general, impossible to identify members of this hypothetical population in advance and that such an analysis should only be used as a supplementary analysis in special situations.
724-725	17	**Comments:** It is not difficult to identify members of this hypothetical population in advance; it is, in general, impossible. **Proposed change:** Describe that it is, in general, impossible to identify members of this hypothetical population in advance and that such an analysis should only be used as supplementary analysis in special situations.
725-727	21	**Comments:** It is not clear how a randomised withdrawal design helps in targeting “patients that would not require rescue medication”. More explanation will be helpful.
728-39	21	**Comments:** please clarify more precisely what an “appropriate” analysis is for this estimand.
728-735	21	**Comments:** In this section, it is well explained what is not correct “A suitable analysis cannot be achieved by restricting the analysis to those subjects who did not switch to rescue medication”, but the way to obtain “an appropriate analysis to account for this confounding” is not described. An example where such analysis is possible, could help for clarity.
730-731	21	**Comments:** These individuals cannot, in general, be identified even after data collection

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		(as mentioned in lines 257-259). **Proposed change:** Remove the words “in advance.”
731-737	20	**Comments:** The document clearly discourages analyses of the non-rescue subset due to the potential non-comparability, but suggests that other approaches would be appropriate; however those might make other non-verifiable assumptions and also might be prone to bias. Is this somewhat of a double standard?
735	7	**Comments:** It is correct that "An appropriate analysis needs to account for this confounding." However, no possible methods are described, not even in an exemplary way. Indeed, no method is available that guarantees to account for all known and unknown confounders. **Comments:** Please add that there is no robust method available in practice to deal with all known and unknown confounders and that the corresponding analysis should only be used as a supplementary analysis in special situations.
735	17	**Comments:** It is correct that "An appropriate analysis needs to account for this confounding." However, no possible methods are described, not even in an exemplary way. Indeed, no method is available which guaranties to account for all known and unknown confounders. **Proposed change:** Add that there is no robust method available in practice to deal with all known and unknown confounders and that the corresponding analysis should only be used as supplementary analysis in special situations.
735	21	**Comments:** The guidance should give clear steer for the reader as to what is needed. The requirement to conduct ‘an appropriate analysis’ in this section does not add sufficient detail to guide the reader in this context. **Proposed change:** Please add more details about the appropriate analysis.
740-747	21	**Comments:** The terminology ‘average’ can have multiple meanings and is inaccurate in this context. **Proposed change:** Correctly describe which type of ‘average’ in the three instances where the

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		word is mentioned in this section.
740-761	21	**Comments:** Please add clarification or details on whether and how treatment duration should be accounted for.
743	7	**Comments:** The defined variable "average of the designated measurements while on randomised treatment" frequently leads to serious problems because the corresponding comparison is unfair due to different follow-up times. **Proposed change:** Please describe the problems of unfair comparisons due to different follow-up times and add that the corresponding analysis should only be used as a supplementary analysis in special situations.
743	17	**Comments:** The defined variable "average of the designated measurements while on randomised treatment" frequently leads to serious problems because the corresponding comparison is unfair due to different follow-up times. **Proposed change:** Describe the problems of unfair comparisons due to different follow-up times and add that the corresponding analysis should only be used as supplementary analysis in special situations.
743, 760-761	12	**Comments:** It would be helpful to include an explanation as it is unclear why the variable changes to be “average of the designated measurements while on randomised treatment” rather than simply using the change from baseline to last measurement on treatment. Also, lines 760-761 state that “considering alternative choices for the variable definition by focussing on the last measurement while being on treatment, leading to different estimands.” ACRO recommends that further explanation be given as to what the different estimand would be of u vs “last measurement on treatment>”sing “average” **Proposed change:** Include explanations on these points.
748-750	7	**Comments:** There is almost always an interest in trial objectives that would require the collection of data after switching to rescue medication. **Proposed change:** Please revise the statement and state that, in general, the collection of data after switching to rescue medication is required.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
748-750	17	**Comments:** There is almost always interest in objectives requiring to collect data after switching to rescue medication. **Proposed change:** Reformulate the statement that in general the collection of data after switching to rescue medication is required.
748-750 813	13	**Comments:** There is usually interest in objectives requiring collecting data after switching to rescue medication. **Proposed change:** Reformulate the statement that in general the collection of data after switching to rescue medication is required.
754	20	**Comments:** There are many aspects of this situation that could be envisioned, both in regard to interpretability and statistical assumptions, that could suggest that the analysis mentioned here as “appropriate” would be highly inappropriate (for example, interpretability could be severely limited if the on-treatment duration differs across treatments). It might be better instead to emphasize that there are aspects of statistical model building which are not addressed within the framework presented and are beyond the scope of this document, but can be quite important.
754-755	18	**Comments:** This analysis does not account for heteroscedasticity introduced by different number of measurements available to calculate the mean. There may be reasons to ignore this heteroscedasticity but these should be mentioned.
755-756	12	**Comments:** It is not clear why interpolation is recommended if the analysis is based on average results on treatment, as any intermittent missing measurement would be populated with the average at the visits that are present. **Proposed change:** Explain why interpolation is recommended.
755-757	21	**Comments:** The way to “interpolate” intermittent missing measurement is not clear. It looks like intermittent missing data may be imputed assuming they are missing at random (providing uncertainty is taken into account), but it is not clear.
757-	21	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
758		"…the assumptions of the interpolation method": add same wording as in other examples ("the assumptions of the imputation method") to be consistent through the whole document. If there is a reason why it is not added here then please clarify. **Proposed change:** the assumptions of the interpolation imputation method
758-761	21	**Comments:** It looks like a "LOCF" approach. Can it be clarified.
762	21	**Comments:** The document does not make it clear what to do if a subject has more than one intercurrent event. For example if a patient switches treatment and subsequently receives rescue medication how is this handled in the defined Estimand.
763-764	21	**Comments:** It is unclear why only discontinuations due to AE are taken into account. How would discontinuation for other reasons be considered?
764	21	**Comments:** Provide an example where all types of treatment discontinuations are accounted for or eliminate "due to an adverse event" from the second intercurrent event definition. If too complex, please acknowledge that a clinical trial would have to deal with all types of treatment discontinuations as intercurrent events (and potentially other intercurrent events).
771-772	21	**Comments:** To be consistent throughout the document, please describe all estimand strategies in this section as well.
773	12	**Comments:** The word "both" should be replaced with "two" for consistency with lines 762 and 797. **Proposed change:** Replace the word "both" with "two".
787	21	**Comments:** Section is highly repetitive of other sections. **Proposed change:** Consider aligning paragraph to reduce text
790	2	**Comments:**

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Again the implicit assumption of missing at random. This will almost always be wrong or at least an un testable assumption, and risks giving the sponsor an easy ride in terms of avoidable missing data. It also provides encouragement that missing data may be surmountable, and thus reduce efforts to minimize those; it seems a retrograde step.
794-796	21	**Comments:** It is suggested that this be clarified to indicate that these supplemental estimands could be used for estimation instead of hypothesis testing. **Proposed change:** "… such as contrasting the proportion and timing of rescue switchers and treatment discontinuations due to adverse events between the treatment groups. These supplemental estimands could be used for estimation instead of hypothesis testing."
802	7	**Comments:** Again, the consideration of the hypothetical setting in which rescue medication would not be available is of no use in practice (see above). **Proposed change:** Please add a clear statement that the corresponding analysis should only be used as a supplementary analysis in special situations.
802	17	**Comments:** Again, the consideration of the hypothetical setting that rescue medication would not be available is useless in practice (see above). **Proposed change:** Add the clear statement that the corresponding analysis should only be used as supplementary analysis in special situations.
807 - 810	15	**Comments:** It´s not logical that "withholding rescue medication" would be considered "unethical", but at the same time subjects could remain "untreated after discontinuation due to an AE". This is not a very plausible/convincing scenario. **Proposed change:** Could a better example be found?
813	7	**Comments:** There is almost always an interest in trial objectives that would require the collection of data after switching to rescue medication. **Proposed change:** Please revise the statement and state that, in general, the collection of data

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		after switching to rescue medication is required.
813	17	**Comments:** There is almost always interest in objectives requiring to collect data after switching to rescue medication. **Proposed change:** Reformulate the statement that in general the collection of data after switching to rescue medication is required.
817-20	21	**Comments:** we would suggest to add that missing data methods should be used to handle missing values, or values observed after the event and hence excluded from the analysis.
823	2	**Comments:** It is unclear why data will be missing in the case of those who stop a drug due to adverse events; we have good success in govt funded trials of separating discontinuation of study drug and follow up, achieving good rates of follow up in those who discontinue treatment. A strength of the randomised experiment is that it avoids the need for assumptions; this is not a good circumstance to let them in (the occasion when we are interested in a measurement of say fatigue in palliative care however is a real time when we want to think about these things (that is a real example of a Govt funded trial).
823-828 / 831-835	21	**Comments:** It seems unclear why missing measurement wording is repeated in the text after the bullets. **Proposed change:** Delete “In case of missing measurements, data need to be predicted based on plausible assumptions while accounting for the added uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event” in one of the two places.
826-828	21	**Comments:** Please add “e.g. timing” everywhere where the following wording is used or at least when the sentence is used for the first time (line 657): “Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event.” **Proposed change:** to line (657) (first occurrence): “Similarity may be established based on the

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event, e.g. timing." **Proposed change:** to line (827): "Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event, e.g. timing."
831-835	21	**Comments:** "In case of missing….intercurrent events" seems to be redundant with lines 823-827
833	2	**Comments:** The document keeps on referring to imputation based upon MAR assumptions; you really should for balance add MNAR and some 'worst case' scenarios eg assumption that the missing subjects in the experimental group have a very poor outcome and the missing subjects in the control condition have a good outcome.
840	21	**Comments:** Objective and variable (or endpoint) are used within R1. To have a comprehensive Glossary of terms used within this Guidance we would suggest adding Objective and Variable (or endpoint) to the Glossary. As there is often a mix between Objectives, Endpoints, and associated statistics (e.g. EMA Guidance on the evaluation of anticancer medicinal products in man regards ORR as an endpoint; standard statement in presenting results of a trial is ….endpoint met…) adding both terms might be beneficial for clarification **Proposed change:** Objective: Determine the scientific research questions the clinical trial should answer and will lead to defining the estimands. Endpoint: An endpoint is an individual subject based quantitative measurement intended to reflect the effect of a drug – maybe extend by – as required by the objectives
840/Glossary	21	**Comments:** For completeness, it would be helpful to add the following definitions, noting that the distinctions between variable and endpoint, and population and analysis set are clarified (as not always clear in the original E9): "population" - the set of patients who might be exposed to the investigational treatment (which is the focus of statistical inference). "analysis set" - the set of observed data to be included in the analysis "Endpoint" – a quantity which is derived from one or more variables (e.g. change from baseline, AUC, Cmax, time to event or censoring) observable on a single subject that directly addresses study objectives.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		"missing data" – value(s) that, though intended to be observed per-protocol, have not been observed "immaterial data"-data that are not considered to be relevant to the estimand which may not be measured or if measured, may be removed prior to analysis (e.g. data after initiation of rescue medication in the hypothetical estimand). "modified analysis set" - the set of observed data to be included in the analysis where immaterial data are intentionally excluded. "subjects" is used for the participants in a clinical study to distinguish from "patients or individuals in the wider population" "variable" - the measure which could be measured directly on any patient in the population (and which we intend to measure on the subjects in our study). "impute" - assign (a value) to a missing data point by inference under certain assumptions (which should be specified). "predict" –to estimate the value of an unknown quantity (e.g. the estimand).
841-845	14	**Comments:** The definition of estimand not fully clear in this glossary (though the idea comes through in the text earlier). **Proposed change:** Modify the glossary definition as follows and link it to the other definitions: Estimand: Is an estimate that addresses the scientific question of interest posed by the trial objective, the question pertains to a specific population. Attributes of an estimand include the population of interest, the variable (or endpoint) of interest, the specification of intercurrent events reflected in the scientific question of interest, and the estimation method by which the estimate will be derived from the data collected during the trial
841-849	20	**Comments:** The definitions given here seem unclear. It is unconventional to define "estimator" as an "analytical approach". Conventionally, this would be called "estimation (method)", for example "maximum likelihood estimation". A conventional definition of estimator is typically in relation to what is estimated, e.g. "any statistic whose values are used to estimate f(theta) where f(.) is some function of the parameter theta."
842-845	7	**Comments:** The important items "intervention" and "comparator" are missing. **Proposed change:** Please add the items "intervention" and "comparator".
842-845	17	**Comments:** The important items intervention and comparator are missing.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		Proposed change: Add the items intervention and comparator.
843-845	21	**Comments:** Update text here in line with any change to the attribute wording. Proposed Change: for example, change to “Attributes of an estimand include the population, the variable(s) and endpoint of interest, the specification of how to account for intercurrent events, and the population-level characteristic defined on the endpoint of interest which is the focus of our estimation.
847	8	**Comments:** Add that the estimate is the numerical value ‘of the population-level summary for the variable (or endpoint) of interest’, otherwise this doesn’t link with the general description of an estimand, and also it doesn’t state what the ‘numerical value’ is supposed to represent in this context. **Proposed change:** Estimate: Is the numerical value of the population-level summary for the variable (or endpoint) of interest, computed by an estimator based on the observed clinical trial data.
847	12	**Comments:** ACRO recommends that it would be informative and helpful to add an example here. **Proposed change:** Add an example.
847	21	**Comments:** Add that the estimate is the numerical value ‘of the population-level summary for the variable (or endpoint) of interest’, otherwise this doesn’t link with the general description of an estimand, and also it doesn’t state what the ‘numerical value’ is supposed to represent in this context. Proposed Change: Estimate: Is the numerical value of the population-level summary for the variable (or endpoint) of interest, computed by an estimator based on the observed clinical trial data.
849	21	**Comments:** It is suggested that the term “intention to treat principle” be defined as this can be misunderstood to also mean that all observations on a study subject must be used in study analysis. **Proposed change:** “Is the analytic approach to compute an estimate from observed clinical trial

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		data. Intention to treat principle: All study subjects randomised to the study are included, in some manner, in the analysis set, and allocated to the group according to randomisation. This does not relate to how each subject's post-baseline information is used, or what data observations are used in calculating the study result."
850-852	8	**Comments:** Clarify whether Study Withdrawal would be considered to be an intercurrent event or not, This event would 'preclude observation of the variable', and therefore it would meet this definition as currently stated in the glossary. However, if Study Withdrawal is not considered to be an intercurrent event then could this definition be amended in order to clarify this? For example "Events that occur after treatment initiation and before study withdrawal, and either preclude observation…" (If that is the correct interpretation of this). **Proposed change:** Events that occur after treatment initiation and before study withdrawal, and either preclude observation of the variable or affect its interpretation.
850-852	12	**Comments:** ACRO recommends that it would be informative and helpful to add an example here. **Proposed change:** Add an example.
850-852	16	**Comments:** In line with the comment that one should distinguish between intercurrent events that have a permanent influence on outcome and those that only have a temporary influence the glossary definition of Intercurrent events could be changed to reflect this. **Proposed change:** change glossary definition to: **"Intercurrent Events:** Events that occur after treatment initiation and either preclude observation of the variable or affect its interpretation. Intercurrent events can either have a permanent influence on the interpretation of the variable or only affect it temporarily."
850-852	21	**Comments:** Clarify whether Study Withdrawal would be considered to be an intercurrent event or not, This event would 'preclude observation of the variable', and therefore it would meet this definition as currently stated in the glossary.

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		However, if Study Withdrawal is not considered to be an intercurrent event then could this definition be amended in order to clarify this? For example "Events that occur after treatment initiation and before study withdrawal, and either preclude observation…" (If that is the correct interpretation of this). **Proposed change:** Events occurring after treatment initiation and before study withdrawal, that affect interpretation of a variable.
850-852	21	**Comments:** Does the definition of intercurrent events need to be more precise, as any event resulting in missing data would fit the current definition? For example, the case of a batch of blood samples lost due to laboratory problems which result in a marker for disease not being measurable would meet this definition. However, the preclusion of these observations would affect all treatments arms equally so it should not meet the criteria to be classed as an intercurrent event. Suggest delete the phrase "preclude observation of the variable"? **Proposed change:** Change "Events that occur after treatment initiation and either preclude observation of the variable or affect its interpretation." to "Events, occurring after treatment initiation and before study withdrawal, that affect interpretation of a variable."
851-852	11	**Comments:** The part of the definition of intercurrent events stating that they may preclude observation is misleading since, although true, the estimand is unchanged whether or not there is any missing data. Missingness may also occur for reasons other than intercurrent events, such as missed visits. Since the document contains recommendations to ensure follow-up after intercurrent events occur, including missingness in the intercurrent event definition is unnecessary. In the definition provided there is also no linkage between intercurrent events and changes of treatment; this is the key to defining intercurrent events since it is these treatment changes (including death) that weaken the clarity of the causal relationship between treatment and outcome that is supposed to be ensured by randomisation. **Proposed change:** Events occurring after treatment initiation (including death) that change the treatment being administered.
851-852	21	**Comments:** Are events that occur after randomisation and before treatment initiation and either preclude observation of the variable or affect its interpretation considered as intercurrent events?

Line no.	Stakeholder no.	Comment and rationale; proposed changes
853-856	12	**Comments:** ACRO recommends that it would be informative and helpful to add an example here. **Proposed change:** Add an example.
853-856	16	**Comments:** The missing data description in the glossary should also include a short remark on the difference between missing data on the outcome and missing data on the intercurrent event itself (i.e. lack of reliably collecting information on intercurrent events) **Proposed change:** change glossary definition to: **"Missing Data:** Data that would be meaningful for the analysis of a given estimand but were not collected. This may concern missing information on the clinical measurements and/or missing information on the occurrence or type of the intercurrent event. Missing data on the clinical outcome should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event."
854	12	**Comments:** The term "not collected" might imply there is no field on the CRF, which should have been considered when designing the CRF. **Proposed change:** Replace "not collected" with "not available".
857 - 863	2	**Comments:** The glossary definitions of principal stratification and principal strata are inadequate for a reader. I have tried these out on colleagues who are statistically literate but have not seen these concepts before and they were not able to understand them.
859-862	20	**Comments:** The definition used here involving four strata seems to strongly conflict with all previous discussions in the addendum, which explicitly mention only two (patients who would not experience the event on either treatment, and the complement). This seems to be a very important distinction for the document to describe and address with clear terminology and consistency.
841	2	**Comments:** Estimand is an arcane term which is not included in standard dictionaries, and

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		not in general use in the statistical community. It is an unhelpful construct in this document (not least as it is rather poorly explained) and non English readers (eg those using English other than as a first language) really struggle with the concept. This is the wrong document to introduce such nomenclature, and it is not central to the arguments.
863-865	12	**Comments:** It is not clear why "principal stratification" (the quantity being defined) is included within the definition. **Proposed change:** Include further explanation.
867	12	**Comments:** In ACRO's view, different assumptions would result in a different estimand. We therefore recommend the change proposed below. **Proposed change:** Replace "differing assumptions" with "differing model assumptions".
867	21	**Comments:** we would suggest to write "[…]same estimand but different estimators, […]".
872	21	**Comments:** Why are supplementary analyses a "broader class" than sensitivity analyses? Please clarify the relation between supplementary & sensitivity analyses. **Proposed change:** "The term describes a broader class of analyses than sensitivity analyses. A supplementary analysis may target a different but related estimand using the same variable(s) (e.g. repeating a composite endpoint with a different threshold) or investigates the attributes of the intercurrent event (e.g. investigating time to start rescue medication or proportion on rescue medication)."
960-966 and 970-975	5	**Comments:** There is a redundancy in these two sentences which discuss how to handle missing data in case there will be for the same estimand. **Proposed change:** Keep only one of them, but seems that a better flow will be to keep 972-975 and delete 962-966.
After 979	8	**Comments:** It would be beneficial to add a generic example, placed in an added section (A.8), that has death as the intercurrent event. This new example can be similar to the generic example already provided (section A.7) but where the

Line no.	Stakeholder no.	Comment and rationale; proposed changes
		disease is life-threatening, as occurs in many oncology trials.
979	21	**Comments:** It would be beneficial to add a generic example, placed in an added section (A.8), that has death as the intercurrent event. This new example can be similar to the generic example already provided (section A.7) but where the disease is life-threatening, as occurs in many oncology trials.
Processes	21	Recommendations outside the scope of a Guideline The following recommend processes, rather that stating requirements with regard to estimands. Although good processes would help to lead to good estimands, it is questionable that the process of arriving at an estimand should be laid down in Guideline. Lines 57-60 up to "…is the reverse" [requires that intercurrent events be considered before defining safety/efficacy variables; this may lead to efficiency but does not seem appropriate as instruction in a Guideline]. Lines 66-68 "It aims…address" addresses processes rather that what is required of an estimand. Line 119 to "…(Figure1)" and Figure 1 itself addresses processes rather that what is required of an estimand. Lines 130-138 and 297-308 address processes rather than attributes required of an estimand. Lines 392-409 up to "…(Section A.5)" addresses processes rather than attributes required of an estimand.
???	21	**Text that does not support the Guideline or add substance to the Guideline** Lines 123-4 "A suitable…selected" is an unnecessary introduction to the next sentence. Lines 183-187 from ", e.g. the mean change…" examples of summary measures are not needed for the Guideline. Lines 217-231 Such extended examples of estimands are suited to a tutorial paper but not to a Guideline. Lines 335-342 describe suppositional positions of regulators, but do not provide actual guidance. Lines 343-348 belongs in a survey/tutorial paper on estimands, rather than a Guideline. Lines 444-449 constitute a vague warning on including multiple objectives. The issue of multiplicity is covered by other guidelines. Suggest either omit or simply refer to other Guideline(s).