Article:
Morga, A., Latimer, N.R., Scott, M. et al. (2023) Is intention to treat still the gold standard or should health technology assessment agencies embrace a broader estimands framework? : Insights and perspectives from the National Institute for Health and Care Excellence and Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen on the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use E9 (R1) Addendum. Value in Health, 26 (2). pp. 234-242. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2022.08.008

Abstract

Objectives

The ICH E9 (R1) addendum will have an important impact, especially with respect to the Intention-to-treat (ITT) principle, on the design and analysis of randomised controlled clinical trials which represent crucial sources of evidence in HTAs. This paper brings together a task force of health economists and statisticians in academic institutes and the pharmaceutical industry, to examine the implications of the addendum from the perspective of NICE and IQWiG, and to address the question of whether the ITT principle should be considered the gold standard for estimating treatment effects.

Methods

We review the ITT principle, as introduced in the ICH E9 guideline. We then present an overview of the ICH E9 (R1) addendum and its estimand framework, highlighting its premise and the proposed strategies for handling intercurrent events (ICEs), and examine some cases among submissions to IQWiG and NICE.

Results

IQWiG and NICE appear to have diverging perspectives around the relevance of the ITT principle and, in particular, the acceptance of hypothetical strategies for estimating treatment effects, as suggested by examples where the Sponsor proposed an alternative approach to the ITT principle when accounting for treatment switching for interventional, oncology trials.

Conclusions

The ICH E9 (R1) addendum supports the use of methods that depart from the ITT principle. The relevance of estimands using these methods depends on the perspectives and objectives of payers. It is challenging to design a study that meets all stakeholders' research questions. Different estimands may serve to answer different relevant questions or decision problems.

Highlights (75 words limit for each highlight statement)

1. What is already known about the topic?

Randomised controlled trials play a critical role in the Regulatory approval of new medicinal treatments, as well as in HTA and pricing and reimbursement assessments. Historically, the central role of the intention-to-treat analyses in the design, conduct and interpretation of clinical trial data has long been recognised. The ICH E9 (R1) addendum, however, opens the door to the possibility of estimating treatment effects on the basis of principles that depart from ITT.

2. What does the paper add to existing knowledge?

To the best of our knowledge, there have been limited attempts to assess the challenges and opportunities, from an HTA perspective, of the addendum to date. To help with this shortfall, we present the differing perspectives of two HTA Agencies, IQWiG and NICE, particularly on the implications of departing from the ITT principle, based upon our understanding of published methods guidance, publicly available agency documents, and our experience of working with these agencies.

3. What insights does the paper provide for informing healthcare-related decision making

Our paper helps address the following questions: If different stakeholders are interested in different research questions, how can the estimand framework reconcile these differing needs, when designing a confirmatory clinical trial? What approaches, alternative to treatment-policy and ITT principle, would be expected to inform the decision-making processes of different stakeholders? As the implementation and awareness of the addendum continues, the understanding of its implication for HTA processes will become even more critical.

Introduction

The principle of Intention-to-treat (ITT) has been accepted as the standard for the design and analysis of randomised controlled clinical trials (RCTs) since its conception in the 1960s.[1,2] Following its formal inclusion into the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 guideline on the "Statistical Principles for Clinical Trials",[3] and together with concealment of treatment allocation to subjects via randomisation and blinded outcome assessment, the ITT principle is seen as the cornerstone of sound statistical analysis and regulatory decision making, for confirmatory trials and for superiority studies in particular.

The ICH E9 (R1) "Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials"[4] (adopted in November 2019, and in the course of implementation by Health Authorities, as per Table 1), however, elaborates that, in the presence of post-randomisation (or "intercurrent") events, decision makers may be interested in estimating treatment effects on the basis of principles that depart from ITT. Confirmatory RCTs represent a crucial source of evidence in the Regulatory review of novel medicinal technologies, as well as in Health Technology Assessments (HTAs). This paper looks at the implications of the addendum from the perspective of two HTA Agencies: the Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen- IQWiG), an important actor in the pricing and reimbursement processes in Germany, and the National Institute for Health and Care Excellence (NICE), which oversees HTA processes in England and Wales. This paper brings together a task force of health economists and statisticians in academic institutes and the pharmaceutical industry, with the aim of addressing the question of whether, in

light of the addendum, the ITT principle can still be seen as the gold standard for estimating treatment effects for confirmatory RCTs by these two HTA Agencies.

Methods

We review the ITT principle, as introduced in the ICH E9 guideline.[3] We then present an overview of the ICH E9 (R1) addendum, highlighting its premise and the proposed strategies for handling intercurrent events (ICEs). We discuss where the perspectives of IQWiG and NICE on the ITT principle may diverge, considering current methodological guidance, views provided on the addendum itself and the authors' direct experience with the two HTA agencies, as well as using examples from historical appraisals. Further, we look at the perspectives that different stakeholders have in their respective decision-making processes, and how these contribute to shaping the question addressed in clinical trials. We then elaborate on the assumptions and main limitations associated with the strategies presented in the addendum, highlighting some of the potential trade-offs that HTA Agencies are confronted with.

The ITT principle: what does it actually mean and why do we use it?

Established in 1990, the ICH brings together Regulators and the Pharmaceutical Industry with the purpose of achieving greater international harmonisation in the development, registration and maintenance of safe, effective and high-quality medicines.[5] This harmonisation mission is accomplished through several guidelines, covering quality, efficacy, safety and multidisciplinary issues. The ICH E9 Harmonised Guideline on "Statistical Principles for Clinical Trials", finalised in 1998, sets out the principles of statistical methodology applied to clinical trials for marketing applications submitted in the ICH regions, and asserted that the best way of

determining the effect of a "treatment policy" is to perform the evaluation on the basis of the intention to treat a subject; that is, on the basis of the *planned* treatment regimens to which study participants were randomised to.[3] It follows that, from a regulatory perspective in which one is planning and analysing a confirmatory trial under the ITT principle, we are estimating the average effect of a subject *being randomised* to receive an experimental treatment versus a control, irrespective of potential events that may have taken place post-randomisation[6] (e.g., use of rescue therapy), with subjects followed up to the specified data collection points for the variable of interest.

This (first) part of the ITT principle is rarely contested for two reasons: first, the subjects normally receive the correct medication to which they were randomised, thereby negating any concerns of analysing a subject as having received the wrong treatment; and, second, researchers understand and readily accept the need to uphold the randomisation by analysing a subject based on what they were randomised to, rather than what they actually received. The second part of the ITT principle, as stated by ICH E9, requires that subjects should be followed up, assessed and analysed as members of the group they were allocated to, irrespective of any possible post-randomisation event that could have affected the treatment or outcome (e.g., lack of compliance to treatment regimen, or use of rescue therapy): once randomised, always analysed.[7, 8]

ICH E9 goes on to say this is difficult to achieve. Indeed, the ability to retain subjects in a study for complete follow up was the key focus of the National Research Council report on missing data in 2010.[9] For some Sponsors, this led to a significant change in clinical trial mindset where discontinuation from treatment was separated from withdrawing from study. Unfortunately, this is still challenging to achieve today, and this remains a significant issue to overcome.

As shown in figure 1, the ITT principle calls for data collection on the relevant endpoint for both intervention and comparator until the planned point of collection, even in presence of post-randomisation events.

Despite the well-known and well-accepted constraints imposed via the trial design, the ITT approach, focusing on the effect of the treatment policy, thereby goes a long way to estimating the effects of treatment in real practice, including events that comprise treatment compliance (e.g., comorbidities, drug-drug interactions, switching to alternative treatments). By embracing the ITT principle, Regulators sought confirmation of the treatment effect in a setting that exists once the drug is marketed – i.e., *generalisation.* Consequently, the designs of pivotal trials have focussed on addressing the question of what happens in clinical practice and have been much less interested in answering hypothetical or "what if?" scenarios.

The ICH E9 further clarifies the ITT principle by introducing the Full Analysis Set (FAS), defined as close of possible to the ITT ideal of including *all* randomised subjects. Most importantly, it warns of the impact on both the subject data and trial conclusions should there be violations of the protocol that occur after randomisation, particularly if their occurrence is related to the treatment assignment. It also notes, however, that in most cases such data are appropriate to include to be consistent with the ITT principle.

In essence, the ICH E9 extensively promoted the ITT principle and carefully qualified the expectations around its implementation. What concerns, therefore, prompted the need to revisit the recommendation around the ITT principle, and what alternative approaches have been recommended when determining the treatment effects of novel medicinal technologies?

The estimand framework and alternative approaches to estimating treatment effects

One of the main premises of the ICH E9 (R1) addendum is that, in presence of ICEs, the estimation of effects on the basis of the ITT principle may not always address the clinical question(s) of relevance to Regulators and other decision makers. The addendum advocates the adoption of a clear framework in the design, conduct, analysis and interpretation of a clinical trial. The description of the objective of a trial should be reflected in the clinical question of interest, through the construction of estimands, and the specification of their five attributes. In constructing each estimand, the relevant strategy for handling ICEs should now be clearly specified. Importantly, alongside the treatment policy strategy reflecting the ITT principle, other strategies for addressing ICEs are also proposed (figure 2).

Figure 3 illustrates the application of the estimands framework in a hypothetical oncology trial, showing how the clinical question of interest affects the strategy for handling some expected ICEs and, thus, potential departures from the ITT principle.

If, ultimately, it is the research question that shapes the estimands, it becomes critical to understand which research question(s) is relevant, from an HTA perspective, when it comes to the design of confirmatory trials, and what strategies for addressing ICEs could be acceptable for HTA purposes.

Results

The IQWiG and NICE perspectives on the ICH E9 (R1) addendum and on the ITT principle

To date, there is scarce evidence that HTA agencies have taken an explicit position on the addendum and its implications. However, some initial considerations on how IQWiG and NICE may respond to analyses that are in line with the ICH E9 (R1) addendum can be made by looking at examples where the Sponsor proposed an alternative approach to the ITT principle when

accounting for treatment switching in interventional, oncology trials. Interestingly, these two Agencies appear to have diverging perspectives around the relevance of the ITT principle and, in particular, the acceptance of hypothetical strategies for estimating treatment effects.

The IQWiG perspective

The most notable promotor of the ITT principle and, therefore, of the treatment policy strategy in the addendum, is the German assessor IQWiG, who assess the level of certainty and extent of added benefit of newly marketed drugs versus the appropriate comparative therapy, as part of the AMNOG process.[10] In consultation comments on the addendum, IQWiG stated[11] that estimates of treatment effects should consider the entire treatment strategy, regardless of treatment discontinuation or switching to alternative treatments. In addition, IQWiG stressed that, of the proposed strategies for ICEs, only two (the treatment policy strategy and the composite strategy) should be used in general as the main analysis, with the remaining three (hypothetical, principal stratum, while on treatment strategies) only useful for supplementary or sensitivity analyses. On the hypothetical strategy, IQWiG notes that this strategy would produce estimates that, ultimately, do not reflect a clinical practice where ICEs have occurred, and that these estimates would often be based on untestable modelling assumptions. An example of IQWiG not accepting a hypothetical strategy in the context of cross-over comes from the assessment of dabrafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma.[12] IQWiG noted that the company's conclusions were "based on a summary of different analyses (intention-to-treat analysis and analyses for the adjustment of the crossover effect). However, the crossover adjustments "[..] were not relevant for the benefit assessment because they were based on strong assumptions, the fulfilment of which cannot be checked with the available data. In addition, the

results of these analyses presented by the company showed no statistically significant advantage of dabrafenib with regards to overall survival".[12]

IQWiG separately noted[12] that estimates of treatment effects should consider the entire treatment strategy regardless of treatment discontinuation or switching to alternative treatments, thus placing great emphasis on adequate follow-up of subjects that extends beyond the end of treatment. This is also stated in version 6.0 of their general methods guidance[13] for the analysis of adverse events, where they argue that the common practice of recording adverse events only up to a maximum time following treatment discontinuation or switching is insufficient for their benefit assessment, as evaluations based on the treatment policy estimand and the ITT principle are needed. Consequently, IQWiG advocates that complete data collection of adverse events be performed, even after treatment discontinuation or switch to alternatives. This is often in stark contrast to adverse event reporting seen in the benefit-risk assessments to obtain marketing authorisation, which often focuses on adverse events that occur under treatment (including several days following discontinuation). To compound the misalliance with IQWiG, analyses conducted for marketing authorisation are subsequently performed on events occurring during a "treatment emergent" window, regardless of when they were reported during the trial.

IQWiG also argues[13] that appropriate methods for time to event data are needed for treatment comparisons for which the trial arms have different durations of observation; this advice applies to all endpoints, including safety, which is why many dossiers currently being submitted now contain extensive survival analyses of adverse events. IQWiG also warns[10] of the case where different observation durations are due to incomplete data collection (censoring in case of treatment discontinuation or change), which would cause informative censoring and may result in a high potential for bias.

IQWiG has always stipulated the necessity for adequate application of the ITT principle, which until its revision, was anchored in ICH E9. The recent additions to IQWiG's guidance highlighted above reinforce this position. This has the potential consequence that analyses submitted for marketing authorisation and conducted using estimands which veer from an ITT principle will be dismissed as inappropriate by IQWiG for the purpose of demonstrating added benefit. Under such a scenario, and if the trial data allow, additional analyses that consider the appropriate application of the ITT principle will need to be conducted. If, on the other hand, the data required for an ITT analysis have not been collected as part of the trial design, an ITT analysis may be impossible to perform. The consequence of not providing a correct analysis would then almost certainly have a negative impact on the evaluation of added benefit by IQWiG.

NICE perspective

Despite not explicitly referring to estimands and hypothetical strategies, NICE was considering their importance when addressing treatment switching – as an alternative to the ITT analyses – more than 10 years ago. The technology appraisal of sunitinib for the treatment of gastrointestinal stromal tumours,[14] published in 2009, contained analyses conducted by the manufacturer that adjusted for treatment switching in the pivotal RCT, which the NICE appraisal committee agreed were acceptable. Subsequently, making adjustments for treatment switching using statistical methods was explicitly mentioned in the NICE methods guide in 2013,[15] followed by a detailed technical support document on the topic published by NICE's Decision Support Unit in 2014.[16]

NICE is not alone: in 2016 Australia's HTA agency, the Pharmaceutical Benefits Advisory Committee (PBAC), published its own recommendations on how to deal with treatment switching when treatment patterns observed in RCTs do not reflect those expected in clinical practice.[17] In both instances, the motivation for moving beyond the ITT analysis is a desire to estimate outcomes that would be observed if treatment pathways that are likely to occur in practice had been followed in the RCT. This reflects a situation where long-term outcomes are important – for example, overall survival. HTA Agencies, particularly those basing their assessment on cost-effectiveness criteria, typically make decisions at a population level, and, therefore, rely on *mean* estimates of outcomes, rather than medians. A lifetime period is analysed, rather than a period restricted to the trial duration. For treatments that affect survival, a key aim is to accurately estimate the mean overall survival associated with each treatment option. Therefore, it makes sense to consider what post-study treatments were received in an RCT, in order to explore the potential impact of these on long-term outcomes – and whether these are representative of what would happen in clinical practice.

In its recent health technology evaluations manual,[18] NICE suggests that treatment switching could result in an ITT analysis being considered inappropriate. This does not only include situations where trial participants randomised to the control group switch onto the new intervention – it also applies to patients randomised to either group switching onto any other treatment that is not part of the standard treatment pathway. In essence, a detailed investigation into treatments received after cessation of randomised treatments is required and then, where possible, adjustment analyses should be undertaken to account for any non-standard treatments, providing estimates of long-term outcomes that would be expected to be observed if appropriate treatment pathways were followed. These pathways could vary in different jurisdictions, because

different treatments are available around the world. Therefore, in general the required analyses could differ depending upon the jurisdiction it is for.

Practically, it may not be possible to adjust for all treatment switches that are deemed "non-standard". However, for HTA Agencies relying on cost-effectiveness criteria in their assessments, there is a willingness to consider hypothetical estimands when ITT analyses do not allow an assessment of the decision options under consideration.

Discussion

Estimands and decision making

The potential for divergence of the views of NICE and IQWiG on the ITT principle reflects the broader issue that different stakeholders (patients, regulators, reimbursement agencies, clinicians) all have their own decision problem to address,[6,19] expressed through different research questions of interest.

Stakeholders may differ according to whether they are interested in *inference* or *estimation*: Regulators would be interested in the benefit/risk profile of a treatment, whilst Payers in assessing the magnitude of the incremental benefit a treatment offers. Stakeholders also vary according to their risk preferences: individual patients may be risk seeking or risk averse and be naturally interested in the efficacy of a treatment while it is being received. Clinicians may be interested in the treatment effects under different scenarios of treatment adherence.[20]

A trial may represent a direct experimental model of the clinical question and an estimand may be sufficient to directly inform a decision; alternatively, it may be a source of parameter

estimates for inclusion in a synthesis of a wider evidence (e.g., network meta-analyses or cost-effectiveness models).

To inform different stakeholders' decision making, Sponsors may need to specify more than one estimand for a given endpoint, as in the PIONEER trial[21] (Table 2).

By acknowledging that different research questions may be most effectively addressed by constructing different estimands, the addendum makes a valuable contribution in maximising the value of trials in supporting decision-making across a range of stakeholders.

Estimands and implementation considerations

Each ICEs strategy rests on a different set of assumptions and limitations, affecting the circumstances in which a given approach may be most appropriate, as detailed in Table 3. Analyses for hypothetical and principal stratum strategies may be prone to bias, if important assumptions of the analytical techniques do not hold. Ultimately, HTA Agencies will have to assess whether strategies that depart from the ITT principle are acceptable, balancing the potential risk of bias against the need to address a research question of interest.[34] Sensitivity analyses are likely to be important, especially for Agencies such as NICE, who use quantitative approaches to estimate clinical and cost-effectiveness.

Conclusions

The ICH E9 (R1) addendum, whilst reflecting legitimate research questions for some stakeholders, for others appears to undermine the very cornerstone of the ITT principle, by

permitting the use of methods that depart from the principle of following up, assessing and analysing subjects irrespective of potential ICEs. To a large extent, the relevance of different estimands will depend on the perspectives and objectives of the stakeholders using them. Often, HTA Agencies require analyses that extrapolate beyond trial periods, and therefore considerations of treatment pathways received during and beyond trial follow-up periods become crucial. Almost by definition, extrapolation involves hypothetical scenarios, and therefore decision makers who rely on extrapolation are likely to be more comfortable with the hypothetical strategy. Understanding the implications of the addendum for indirect-treatment comparisons (e.g., when assessing the heterogeneity across trials) will also be important. Given that Regulators and HTA Agencies both rely heavily on RCTs for their decision making, it may be helpful for there to be a greater degree of convergence between them when formulating guidance such as the addendum. In addition, the language used in the world of HTA differs to that used by Regulators: HTA Agencies typically refer to "decision problems", whilst Regulators refer to "estimands": a harmonisation of language may also be valuable.

It is important to recognize that it is challenging, if not impossible, to design a trial that meets all stakeholders´ research questions of interest. Different estimands may indeed serve to answer different relevant questions or decision problems, so that HTA-specific research questions can potentially be addressed by defining additional estimands, in other words additional treatment effects of interest. Sponsors' engagements through early dialogues or scientific advice (with both Regulators and HTA Agencies) will be important in clarifying stakeholders' views on the potential estimands.

Furthermore, the statistical methods used to assess the different estimands referred to in the addendum often rely on comprehensive data collection (e.g., adequate follow-up, off treatment

data, measured confounders of switching and prognosis). In the absence of such data, analyses of hypothetical estimands will be unreliable. The addendum states that all data required to support estimation of the different estimands should be collected. Here, adoption of the addendum may be particularly useful in the world of HTA, as improved data collection in trials may result in more reliable estimation of hypothetical estimands for future decision making. Although there is a desire and need for long-term data, a pragmatic way forward may see the trial data collection limited to the most important clinical period, and the duration of assessment limited to a time period that is both feasible and interpretable (e.g., allowing long-term data collections through the use of RWE registries). As the addendum continues to be implemented, the interpretation of its principles by HTA Agencies will also mature and evolve: future research should consider assessing the addendum looking at Agencies beyond NICE and IQWiG, as well as using different methodological approaches (e.g., semi-structured interviews with representatives of individual organisations).

References

1. Unnebrink K. and Windeler J., Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases, *Stat Med* 2001 Dec 30; 20(24):3931-46.
2. Hill AB., *Principles of Medical Statistics*, 7th edn, The Lancet Limited, London, 1961.
3. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 guideline on the "Statistical Principles for Clinical Trials", 1998. Accessed November 2021. https://www.ich.org/page/efficacy-guidelines
4. ICH E9 (R1) "Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials", 2019. Accessed January 2022. https://www.ich.org/page/efficacy-guidelines.
5. ICH Mission, ICH official website. Accessed January 2022. https://www.ich.org/page/mission.
6. Mallinckrod C, Molengerghs G, LipKovich I, Ratitch B, *Estimands, Estimators and Sensitivity Analysis in Clinical Trials*, Chapman & Hall/CRC Biostatistics series, 2020
7. Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH, Intention-to-treat analysis: Who is in. Who is out? *J Fam Pract.* 2002; 51; 969–71.
8. Gupta, S K, Intention-to-treat concept: A review, *Perspect Clin Res* 2011; 2(3); 109–112
9. National Research Council (US) Panel on Handling Missing Data in Clinical Trials, *The prevention and treatment of missing data in clinical trials*, National Academies Press, 2010
10. The AMNOG procedure: more than just cost control. IQWiG website. Accessed January 2022 https://www.iqwig.de/presse/im-fokus/neue-arzneimittel-zulassung-nutzenbewertung-erstattung/3-das-amnog-verfahren-mehr-als-nur-kostenkontrolle/.

11. 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). Accessed December 2021. https://www.ema.europa.eu/en/documents/comments/overview-comments-received-draft-ich-e9r1-addendum-estimands-sensitivity-analysis-clinical-trials/chmp/ich/436221/2017_en.pdf.
12. IQWiG dossier evaluation nr 203 - Dabrafenib – Nutzenbewertung gemäß § 35a SGB V. Accessed November 2021. https://www.g-ba.de/downloads/92-975-393/2013-12-23_A13-35_Dabrafenib_Nutzenbewertung-35a-SGB-V.pdf
13. IQWiG General Methods guidance. Accessed November 2021. https://www.iqwig.de/methoden/allgemeine-methoden_entwurf-fuer-version-6-1.pdf.
14. Sunitinib for the treatment of gastrointestinal stromal tumours, NICE Technology appraisal guidance, TA179, 2009. Accessed December 2021. https://www.nice.org.uk/guidance/ta179
15. Guide to the methods of technology appraisal 2013, NICE. Accessed November 2021. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781.
16. Latimer NR, Abrams KR, NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching, report by the Decision Support Unit, 2014. Accessed November 2021. http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD16_Treatment_Switching.pdf.
17. PBAC Guidelines, Adjustments for treatment switching, subsection 2.6.4. 2016. Accessed November 2021. https://pbac.pbs.gov.au/section-2/2-6-trial-results-additional-analyses.html.

18. NICE health technology evaluations: the manual - Process and methods, 2022. Accessed April 2022. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741 .
19. Keene, O. N., Ruberg S, Schacht A, Akacha M, Lawrance R, Berglind A, Wright D, What matters most? Different stakeholder perspectives on estimands for an invented case study in COPD, *Pharmaceutical Statistics* 2020; 19(4); 370-387
20. Qu Y, Luo J, Ruberg SJ, Implementation of tripartite estimands using adherence causal estimators under the causal inference framework, *Pharmaceutical Statistics* 2021; 20:55–67
21. Aroda VR, Rosenstock J, Terauchi Y. et al., PIONEER 1 Investigators, PIONEER 1: Randomized Clinical Trial of the Efficacy and Safety of Oral Semaglutide Monotherapy in Comparison With Placebo in Patients With Type 2 Diabetes, *Diabetes Care* 2019; 42(9):1724-1732.
22. Halpin DMG, Worsley S, Ismaila AS, et al., INTREPID: single- versus multiple inhaler triple therapy for COPD in usual clinical practice. *ERJ Open Res* 2021; 7: 00950-2020.
23. INTREPID: Investigation of TRELEGY Effectiveness: Usual Practice Design, ClinicalTrials.gov Identifier: NCT03467425, Clinicaltrial.gov trial record. Accessed April 2022. https://clinicaltrials.gov/ct2/show/NCT03467425?cond=intrepid+copd&draw=2&rank=1.
24. Armstrong PW, Westerhout CM, Composite End Points in Clinical Research: A Time for Reappraisal, *Circulation* 2017; 135; 2299–2307
25. Cordoba G, Shwartz L, Woloshin S, et al., Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review, *BMJ* 2010;341:c3920

26. Kahan, BC, Morris, TP, White, IR, et al., Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions. *BMC Med* 2020; 18: 286
27. Bridge, H and Schindler, TM, Estimands – closing the gap between study design and analysis, *Medical Writing* 2018; 25(4); 52-56.
28. Saha C and Jones MP, Bias in the last observation carried forward method under informative dropout, Journal of Statistical Planning and Inference 2009; 139(2): 246-255
29. Dematteo RP, Ballman KV, Antonescu CR, et al., Adjuvant imatinib mesylate after resection of localised, primary gastrointestinal stromal tumour: a randomised, double-blind, placebo-controlled trial [published correction in *Lancet* 2009; 374(9688):450]. *Lancet* 2009; 373(9669):1097-1104.
30. NICE Final appraisal determination – Imatinib for the adjuvant treatment of gastrointestinal stromal tumours (review of NICE technology appraisal guidance 196), 2014. Accessed April 2022. NICE_gastrointestinal-stromal-tumours-imatinib-adjuvant-rev-ta196-id696-final-appraisal-determination-document2.pdf.
31. Latimer NR, Abrams KR, Lambert PC, et al., Adjusting for treatment switching in randomised controlled trials - A simulation study and a simplified two-stage method. *Stat Methods Med Res* 2017; 26(2): 724-751.
32. Cree BA, Magnusson B, Rouyrre N, et al., Siponimod: Disentangling disability and relapses in secondary progressive multiple sclerosis. *Mult Scler* 2021; 27(10): 1564-1576.
33. Page LC, Feller A, Grindal T, Miratrix L, Somers M-A. Principal Stratification: A Tool for Understanding Variation in Program Effects Across Endogenous Subgroups. *American Journal of Evaluation* 2015; 36(4): 514-531.

34. Keene ON, Wright D, Phillips A, Wright M., Why ITT analysis is not always the answer for estimating treatment effects in clinical trials. *Contemp Clin Trials*, 2021;108:106494.

Figure 1: Intention-to-treat principle in a clinical trial

The intention-to-treat principle calls for subjects to be followed-up, and for the relevant variable or endpoint of interest to be analysed, up to the planned point of data collection for the endpoint or variable, irrespective of whether subjects have experienced intercurrent events (such as, discontinuation, adherence to treatment, rescue medication) following randomisation. It requires the collection of all data necessary to the comparison between the relevant treatments. This in turn led, historically, to associate a strict interpretation of the ITT principle with issues of missing data.

Figure 2: The Estimand framework in the ICH E9 (R1) addendum

An estimand as a "precise description of the treatment effect, reflecting the clinical question posed by a given clinical trial objective" .[4] Compared with traditionally used endpoints, estimands are more detailed definitions of the quantity to be estimated[6]. The addendum identifies five specific attributes to consider when designing estimands, represented by treatment, population, variable or endpoint, population level summary and intercurrent events. For clinical trial designs, it is critical to align these attributes with the treatment investigated.[4] Each estimand calls for the description of the possible ICEs of interest that can affect the endpoint or variable. In correspondence to each of these ICEs, the strategy for handling them is expected to be clearly formulated at protocol design stage. The addendum identifies at the least five potential strategies. The choice of the relevant strategy for each ICE will depend on, and reflect, the specific trial objectives, as well as additional considerations (e.g., the specific perspective of the relevant decision makers).

Figure 3: Applying the estimand framework in a hypothetical oncology trial

The estimand framework in action: following a clear definition of the trial objectives, the primary and secondary estimands are constructed, through the specification of the relevant attributes, including the possible ICEs. Depending on the research question of interest addressed by the trial, the relevant strategies for handling the different ICE are selected.

1= the secondary comparison will be made up to the point of tumour progression or death; NAT= novel anticancer therapy; SOC=Standard of care.

Table 1

ICH E9 (R1) addendum - Implementation status

Implemented

In the process of implementation

ICH member	**Country or Region**	**Status (as of May 2022)**
ANVISA	Brazil	
EC	Europe	
FDA	United States	
HSA	Singapore	
MFDS	Republic of Korea	
Health Canada	Canada	
MHLW/PMDA	Japan	
NMPA	China	
Swissmedic	Switzerland	
TFDA	Chinese Taipei	

Source: https://www.ich.org/page/efficacy-guidelines (accessed 4th May, 2022).
ANVISA=Agência Nacional de Vigilância Sanitária; EC= European Commission
FDA= Food and Drug Administration; HAS = Health Sciences Authority; MFDS=Ministry of Food and Drug Safety; MHLW= Ministry of Health, Labour and Welfare;

PMDA=Pharmaceuticals and Medical Devices Agency; NMPA= National Medical Products Administration; TFDA=Taiwan Food and Drug Administration

Table 2

Estimands in PIONEER trial from the perspectives of different stakeholders

	Estimands in PIONEER trial	
	"Treatment policy" Estimand: The treatment policy estimand provides estimates of treatment effect in trial patients regardless of trial product discontinuation or use of rescue medication.	"Trial product" Estimand: The trial product estimand provides estimates of treatment effect if patients had continued trial product and had not used rescue medication.
Relevant Stakeholder	The treatment policy estimand may be of interest to regulators interested in making inference about the efficacy of the semaglutide or HTA agencies interested in the expected effectiveness in a population accounting for discontinuation and use of rescue medication.	The "trial product" estimand may be of helpful to patients who are risk-seeking or apply minimax regret criteria and are interested in the potential outcomes if they do not discontinue treatment or require rescue medication. It may also be of interest to agents developing models where the causals effects of rescue medication and post-discontinuation treatments are explicitly modelled. This would allow predictions of effectiveness to be estimated conditional on patterns of discontinuation and rescue medication that differ from those observed in the PIONEER trial.

Table 3

Table 3 Context, assumptions and limitations of strategies for ICEs

ICE strategy	Appropriate when	Examples	Estimation methods	Key assumptions and limitations
Treatment policy strategy	Treatment policies in the trial reflect the decision problem and treatment strategies in clinical practice (generalisability); relevant ICE (e.g., rescue treatment) has relatively limited impact on the estimated endpoint	In the INTREPID trial[22,23] (COPD disease area), a secondary outcome measured the change from baseline in forced expiratory volume in 1 Second (FEV1) at Week 24. A treatment policy strategy was used for the ICEs of randomized treatment discontinuation, randomized treatment modification, change of pulmonary rehabilitation status and start of oxygen therapy.	In the INTREPID trial,[22,23] the ITT population is used for the analysis of primary and secondary outcomes	Requires patients' follow-up irrespective of occurrence of ICE (e.g., even after treatment discontinuation); burden of data collection on patients should be clearly foreseen. Not feasible for terminal ICE (since the variable cannot be observed nor measured following the ICE in these instances, unless the terminal event itself is the observed outcome).[6]
Composite strategy	The composite endpoint (which incorporates the relevant ICE) needs to have clinical plausibility and interpretability. Useful when ICEs are terminal events (e.g., death)[6], and when designing secondary	In the INTREPID trial[22,23] (as above), the primary outcome assessed the number of Responders and Non-responders based on the Chronic Obstructive Pulmonary Disease Assessment Test (CAT) at Week 24 and number of participants with Imputed		Risk of bias when the relevant ICE depends on the clinician or investigator assessment and reporting (e.g., admission to hospital), particularly in non-blinded trials, as already noted in the composite endpoints literature.[24, 25]

	estimands (e.g., patients reported outcomes).	CAT Score at Week 24. A composite strategy was applied for ICEs of randomized treatment modification, change in pulmonary rehabilitation or start of oxygen therapy, otherwise a treatment policy strategy was applied.		
While on treatment strategy	The rate of an event or outcome is constant over time and, in general, when the duration of the intervention is not important[6] in the estimate of treatment effects	A while-alive estimand has been proposed for clinical trials of treatments for patients hospitalised for COVID-19, to address objectives related to the healthcare systems perspective which focus on estimating real-world resource savings due to treatment, informing planning and healthcare capacity.[26]	The variable of interest is measured up to the occurrence of the ICE. Usually, the endpoint is measured multiple times during the course of the study. In other instances, estimates are based on the last recorded observation of the variable or endpoint, prior to the occurrence of the ICE[25].	Risk of bias if estimates based on the last-observation-carried-forward analysis method.[27, 28]

Hypothetical strategy	The hypothetical scenarios linked to the occurrence of the ICEs are clinically plausible and relevant to decision making	The ACOSOG Z9001[29] study was a double-blind, placebo-controlled phase III trial of adjuvant imatinib in adults at any level of risk of recurrence after complete surgical removal of KIT (CD117)-positive Gastrointestinal Stromal Tumors (GISTs). Patients randomised to placebo who had not experienced disease recurrence were allowed to crossover to treatment with imatinib for 1 year.	In the submission[30] to NICE for the appraisal of imatinib in GISTs, the sponsor used different methods for adjusting for treatment crossover in ACOSOG Z9001, namely: a rank-preserving structural failure time model (RPSFTM), the iterative parameter estimation (IPE) algorithm, inverse probability of censoring weights (IPCW) and per-protocol analyses that censored crossovers at the time of switching or excluded them altogether. The Evidence Review Group (ERG) concluded that the IPCW method was the most reliable for estimating recurrence-free survival and overall survival.	Methods that exclude or censor crossover patients from the analysis are affected by selection bias. Both the RPSFTM and IPE assume that the relative treatment effect for crossover subjects is the same as that of subjects initially randomized to the experimental group. These randomization-based methods call for treatment effects to be not time-dependent in order to generate unbiased results. The IPCW and the Structural Nested Model (SNM) with g-estimation methods assume no unmeasured confounders and require data on all prognostic covariates. These observational based methods are sensitive to the proportion of control group patients that crossover.[31]

Principal stratum strategy	The principal strata are clinically meaningful and relevant to decision making.	The EXPAND trial[32] (multiple sclerosis disease area) investigated the extent to which the efficacy of Siponimod on 3 and 6-month confirmed disability progression were independent of on-study relapses. Treatment effects were estimated in the principal stratum of non-relapsing patients, defined as individuals who would not relapse regardless of treatment assignment.	Methods for estimating stratum-specific treatment effects can be broadly divided into moment-based methods (nonparametric), which use sample-level information, and model-based methods, which are based on individual-level information.[33]	Moment-based methods rely on weaker assumptions, but lack of structure makes disentangling many strata more difficult. Model-based methods may enable separating multiple strata, but treatment effect estimates can be sensitive to modeling choices.[33]