This is the pre-peer reviewed version of the article, which has been submitted for publication in Statistics in Medicine. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Commentary on Fleming et al "A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events"

Authors: Oliver Keene[1], David Wright[2], Christine Fletcher[3]

Affiliations:

1. KeeneONStatistics, Maidenhead, UK
2. Statistical Innovation, Respiratory and Immunology Biometrics and Statistical Innovation, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK
3. Biostatistics, GlaxoSmithKline Plc, Stevenage, UK

Corresponding author: Oliver Keene, oliverkeene405@gmail.com

Conflict of interest: ONK is an independent consultant to the pharmaceutical industry. DW and CF are employed by pharmaceutical companies as declared above.

Acknowledgement: This commentary reflects the views of the authors, who are members of the Estimands Implementation Working Group (EIWG, www.eiwg.info), a collaborative effort sponsored by the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) and the European Federation of Pharmaceutical Industries and Associations (EFPIA). The views expressed here may not represent the positions of other members of that group.

Introduction

The recent article by Fleming et al titled *"A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events"* offers a strong endorsement of the treatment policy strategy for handling intercurrent events in clinical trials. The authors argue that this approach ensures both scientific rigour and real-world relevance by preserving the integrity of randomization and avoiding problematic assumptions associated with alternative estimand strategies. While we acknowledge the paper's contribution to ongoing discussions prompted by the ICH E9(R1) Addendum and its useful guidance on good clinical trial design, including the value of using a standard-of-care control arm, we believe several of its claims merit closer examination.

In this commentary, we challenge the assertion that the treatment policy strategy uniquely supports causal inference about the effect of an intervention, noting that this strategy estimates the effect of treatment assignment rather than the causal effect of the intervention itself and that alternative strategies—such as hypothetical or principal stratum—are also grounded in causal frameworks. We particularly question the suggestion that treatment policy estimation avoids strong assumptions, emphasizing that all estimand strategies, including treatment policy, require careful consideration of missing data, modelling choices, and sensitivity analyses.

We further examine the paper's characterization of integrity of randomization, arguing that preserving this integrity does not preclude alternative strategies. Moreover, we highlight that real-world relevance is not exclusive to treatment policy estimands; other strategies may better align with patient-centred decisions. Finally, we caution against the view that hypothetical, principal stratum and while-on-treatment strategies mislead patients and prescribers. Different strategies can be scientifically rigorous and clinically justified, depending on the specific context and the nature of the questions being asked.

Through this commentary, we aim to promote a more nuanced understanding of estimand strategies and encourage balanced consideration of their respective strengths and limitations in clinical trial design.

1. **Causal effects**

A key claim of the paper [1] is that use of the treatment policy strategy "allows any differences observed between arms to be causally attributed to the effect of the intervention". In fact, using the treatment policy strategy only assesses the causal effect of *initial assignment* to the randomised intervention, and therefore does not assess the causal effect of receiving the intervention itself. This is an important distinction. If a participant discontinues their randomised intervention because it is ineffective and then receives an alternative intervention which is effective for them, under the treatment policy strategy the outcome from use of the effective alternative intervention is attributed to the randomised intervention.

When causal inference is applied to assessment of interventions in clinical trials, two separate causal effects are typically identified [2, 3]:

(1) the effect of assignment to the interventions at baseline (regardless of whether the interventions are received during follow-up, sometimes known as the 'intention-to-treat effect');
(2) the effect of adhering to intervention as specified in the trial protocol (sometimes known as the 'per-protocol effect'.

The Fleming et al paper [1] states that "*Implementation of the estimand framework using while on treatment, hypothetical, or principal stratum strategies for handling intercurrent events is problematic … regarding causal effects of interventions*". While the treatment policy strategy corresponds to assessing the effect of assignment, the while on treatment, hypothetical, and principal stratum strategies correspond to approaches to assessing the effect of adhering to the intervention. All of these strategies can be expressed as causal effects as shown by Drury et al [4].

2. Estimation

One of the arguments used in the paper for use of treatment policy strategy is that "*Use of a hypothetical estimand leads to estimating effects in counterfactual settings, often requiring reliance on strong, untestable assumptions*." In the presence of missing data, estimation of effects using the treatment policy strategy is also difficult and also typically relies on strong, untestable assumptions.

We agree with Fleming et al [1] that fundamental to estimands which use a treatment policy strategy is that every attempt should be made to collect data post the occurrence of intercurrent events to minimise issues relating to missing data. This includes a rigorous approach to request participants who no longer wish to remain on their assigned treatment to stay in the trial until the follow up of key outcomes has been achieved.

However, even with these measures, nearly all clinical trials will continue to encounter missing data. In terms of estimating a treatment policy estimand, the paper states: "*if some data are missing, assumptions about missing data …, should be centered around the best projections for the participants' outcomes had they been captured".* It is unclear how to determine the "best projection" for a participant's outcome given that the participant could receive a variety of potential alternative interventions once they have left the trial. This is a considerably more complicated problem that determining the best projection had they continued to take the assigned treatment and there is likely to be a shortage of observed data within the trial to allow estimation that conditions fully on all interventions taken.

Bell et al [5] conducted a comprehensive review of statistical methods for estimating treatment policy estimands. Their conclusions include the findings that "*Handling IEs (intercurrent events) via treatment policy is easy to specify at the estimand level, but hard to reliably estimate*" and that "*Whenever pre- specifying treatment policy based estimation, a statistician is thus faced with a dilemma; either to make very strong assumptions directly about the magnitude of the treatment effect that could lead to bias if false, or weaker assumptions where insufficient observation of post- IE data will lead to variance inflation and greatly reduced power*."

The paper by Fleming et al refers to "*incorporating other meaningful intercurrent events, such as death, into the primary endpoint applying a composite strategy*". When using a composite strategy for a continuous variable, a numerical (failure) value is required to appropriately reflect the poor outcome for the participant. When the mean is used as a summary statistic, the choice of this value becomes a key part of the definition of the estimand as it is potentially influential on the estimated treatment effect. Importantly, the appropriate choice is not obvious in many cases and recent work has shown that failure values far from the distributional mean cause large, predictable, variance inflation [6].

Estimation approaches for all of the strategies, including treatment policy, therefore require strong assumptions and the plausibility of these assumptions needs to be assessed based on the context of

the clinical trial. There is typically a need for sensitivity analyses whichever estimand strategy is chosen.

3. Integrity of randomization

The paper claims in a number of places that only use of a treatment policy strategy preserves the "integrity of the randomization". In a simple sense, the integrity of the randomization is preserved if all randomized participants are included in the analysis and estimators addressing hypothetical, composite and while on-treatment estimands will typically do this.

For regulatory decision making, we agree that a primary analysis of a pivotal trial should include all randomised participants, but the analysis doesn't necessarily need to include all data from these participants as the data to be included should align to the clinical question of interest. Composite and/or hypothetical approaches to handle ICEs are potentially appropriate alternatives.

Later the paper refers to hypothetical and while on-treatment strategies as using "post-randomization outcomes to exclude information from randomized participants, thus not preserving the integrity of randomization".

If the "integrity of randomization" refers to analysis that uses all data for all randomized participants, then a treatment policy approach will also typically fail to "preserve the integrity of the randomization" as this approach excludes important, highly relevant information i.e. that the participant has experienced an intercurrent event.

We agree with the statement in the paper that comparing treatment arms with respect to occurrence of intercurrent events can provide relevant additional insight. The paper provides an example for where it is claimed a "while on-treatment" strategy would mislead i.e. when the experimental regimen's only causal effects are its unfavorable effects on occurrence of intercurrent events. A simple comparison of intercurrent events would avoid any unwarranted conclusions in this case.

4. Relevance to Real-World Settings

The article argues that estimands based on the treatment policy strategy are uniquely relevant to real-world clinical practice, citing the prevalence of intercurrent events (ICEs) in such settings. However, this claim warrants more consideration. While estimands using the treatment policy strategy incorporate all post-randomization data irrespective of intercurrent events, their real-world applicability depends critically on how closely the intercurrent events observed in the trial reflect those that would occur in routine practice. In many cases, this alignment is uncertain.

Indeed, treatment policy estimands can be *more* sensitive to such differences, particularly because the strategy attributes outcome differences to the randomized treatment even when subsequent interventions following treatment discontinuation vary across settings or patient populations. These follow-on treatments are typically not standardized, may evolve over time, and often differ across geographical regions. As such, the post-randomization care pathways embedded within a trial may not constitute a coherent or well-defined "treatment policy" in any practical or reproducible sense. In these cases, describing such estimands as reflecting a "real-world treatment policy" risks overstating their generalizability.

Furthermore, the assumption that hypothetical strategies lack real-world relevance also deserves reconsideration. In some scenarios, such as when estimating the efficacy of a treatment had the

intercurrent event not occurred, hypothetical estimands may better isolate the effect of the intervention itself. This information can be crucial for regulators, clinicians, and patients seeking to understand the efficacy profile of a drug, independent of rescue treatments, treatment switching, or early discontinuation. Indeed, the CHMP guideline on the clinical investigation of medicinal products for the treatment or prevention of diabetes mellitus proposes the use of a hypothetical strategy for subjects who receive rescue medication [7].

The paper acknowledges that "*Incorporating intercurrent events of compelling clinical relevance into a composite primary endpoint (i.e., using a composite strategy), may sometimes be desirable*". The only examples given here are for events such as death, surgery or severe CV events. Later, the paper expresses concern on the use of the composite strategy for treatment discontinuation stating that "*considering treatment discontinuation to be an event in a composite endpoint would meaningfully weaken the clinical relevance of that endpoint*." When treatment discontinuation occurs due to lack of efficacy, it provides important evidence that the intervention is failing to deliver its intended benefit. Focusing only on later outcomes, which may be influenced by alternative medications, ignores information that is highly relevant to patients and prescribers.

In reality, the composite strategy may be useful more broadly than for death, surgery or severe CV events, for example for the common intercurrent event of use of a new medication, either as a rescue medication or as an alternative/additional medication to the randomized treatment. From a clinical perspective, it may be considered that this is a compelling event in terms of understanding the efficacy of a product and that the estimand chosen needs to reflect the negative outcome for patients of the need to use rescue or alternative medication. A composite strategy for the intercurrent event which changes the definition of the endpoint explicitly recognises the occurrence of the event whereas a treatment policy strategy disregards the event [8].

Therefore, the notion that treatment policy is inherently more relevant to clinical practice than other estimand strategies oversimplifies the complexity of real-world decision-making and patient care. All estimand strategies have limitations in generalizability, and the choice should depend on the specific clinical and regulatory questions being asked.

5. Scientific Questions of Interest

One of the addendum's strengths is to emphasise that clinical trials should clearly define the clinical question of interest and that this should be considered before defining the analysis method. The guideline is broader than simply "*providing a common language to describe how handling of missing data modifies the target of scientific investigation*" and the intent of the addendum is for the clinical question of interest to lead to the analysis method.

A central position in the Fleming et al article [1] is that clinical trials should be designed to answer a single primary question—namely, the effect of assignment to an intervention —and that other estimand strategies may mislead clinicians and patients. While the goal of clarity is commendable, this position appears overly restrictive.

There is no single scientific question of interest that universally applies across all clinical contexts. Clinical trials often serve diverse stakeholders—regulators, prescribers, payers, and patients—each of whom may prioritize different aspects of a treatment's effect. While treatment policy estimands may align with certain public health objectives, other strategies, such as hypothetical estimands, can provide valuable insights into the causal efficacy of the intervention itself—i.e., what the drug can achieve when taken as prescribed [9].

For example, for non-inferiority trials, the Fleming paper [1] recommends use of a treatment policy strategy. When a treatment policy strategy is used for use of rescue and/or alternative medication, the intercurrent event will be an essential part of the treatment strategy. The clinical question then becomes one of assessing the non- inferiority of treatment strategies including rescue and/or alternative medications not one of non- inferiority of the randomised treatments themselves [10].

Moreover, while the term "hypothetical" may raise concerns for some non-statisticians, this should not be equated with irrelevance or speculation. When well defined and transparently motivated, hypothetical strategies offer an explicit framework to estimate the causal effect of the intervention in the absence of ICEs—precisely the kind of information many clinicians and patients seek [11]. In contrast, treatment policy estimands may reflect a composite effect of multiple interventions and patient behaviours post-randomization, complicating interpretation.

Thus, dismissing alternative estimand strategies as scientifically unsound overlooks their value in illuminating different, yet important, aspects of treatment effect. A more pluralistic approach—recognizing that different scientific questions may require different estimands—is both statistically principled and clinically appropriate.

Conclusion

The paper by Fleming et al. offers a strong endorsement of the treatment policy strategy for handling intercurrent events in clinical trials (with potentially a composite strategy for compelling clinical events), positioning this approach as the only reliable basis for causal inference and the sole estimand strategy aligned with real-world practice. While their emphasis on clear, interpretable results is appreciated, the arguments presented often overlook both the conceptual limitations of the treatment policy strategy and the legitimate scientific value of alternative estimand approaches.

Our commentary highlights that treatment policy estimands do not estimate the causal effect of receiving the treatment itself, and in fact, can obscure this effect by including outcomes influenced by alternative interventions. Furthermore, while presented as straightforward, the estimation of treatment policy estimands still requires strong, often unverifiable assumptions in the presence of missing data.

Crucially, clinical relevance is not determined solely by the treatment policy approach. In many real-world contexts strategies such as hypothetical may provide more interpretable and appropriate assessments of treatment benefit. To claim that estimands addressing other questions are inherently misleading risks constraining scientific inquiry and ignoring stakeholder diversity.

Rather than endorsing a single strategy, we advocate for a principled, context-dependent approach to estimand selection, recognizing that different clinical questions require different estimand strategies, and that dismissing alternative approaches undermines the potential for trials to inform nuanced but clinically relevant decisions.

References

1. Fleming TR, Carroll KJ, Wittes JT, Emerson SS, Rothmann MD, Collins S, Levin G. A Perspective on the Appropriate Implementation of ICH E9 (R1) Addendum Strategies for Handling Intercurrent Events. Statistics in Medicine. 2025 May;44(10-12):e70104.
2. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020 (revised Mar 2023).
3. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, Emberson JR. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;28;366.
4. Drury T, Bartlett JW, Wright D, Keene ON. The Estimand Framework and Causal Inference: Complementary not Competing Paradigms. arXiv preprint arXiv:2412.12380. 2024 Dec 16.
5. Bell J, Drury T, Mütze T, Pipper CB, Guizzaro L, Mitroiu M, Rantell KR, Wolbers M, Wright D. Estimation methods for estimands using the treatment policy strategy; a simulation study based on the PIONEER 1 Trial. Pharmaceutical Statistics. 2025 Mar;24(2):e2472.
6. Bell J. Continuous Composite Endpoints: How Bad Is Too Bad? Presentation at PSI conference 2025.
7. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. EMA/CHMP/51351/2023. 2023. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-or-prevention-diabetes-mellitus-revision-2_en.pdf
8. Keene ON. Strategies for composite estimands in confirmatory clinical trials: examples from trials in nasal polyps and steroid reduction. Pharmaceutical Statistics. 2019 Jan;18(1):78-84.
9. Keene ON. Intent-to-treat analysis in the presence of off-treatment or missing data. Pharmaceutical Statistics. 2011 May;10(3):191-5.
10. Lynggaard H, Keene ON, Mütze T, Rehal S. Applying the Estimand Framework to Non-Inferiority Trials. Pharmaceutical Statistics. 2024 Nov;23(6):1156-65.
11. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials – broadening the perspective. Statist. Med. 2017;36:5–19.