Different Clinical Questions Need Different Estimands

Oliver Keene[1]. Helle Lynggaard[2], Christine Fletcher[3], David Wright[4], Paul Terrill[5]

Affiliations:
1. KeeneONStatistics, Maidenhead, UK
2. Biostatistics, Novo Nordisk A/S, Denmark
3. Biostatistics, GlaxoSmithKline Plc, Stevenage, UK
4. Statistical Innovation, Respiratory and Immunology Biometrics and Statistical Innovation, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK
5. PT Stat Consulting Ltd, Church Stretton, UK

Corresponding author: Oliver Keene, oliverkeene405@gmail.com

Keywords: Estimands; treatment policy; hypothetical strategy; causal inference; intercurrent events.

Introduction

The ICH E9(R1) estimands framework requires precise specification of the treatment effects (estimands) a trial is designed to estimate. Recent articles[1,2], have advanced a narrow view that only estimands using the treatment-policy strategy are scientifically defensible. In particular, Troxel et al[1] recommend that journals adopt a new policy with regard to reporting results from clinical trials, advocating that only results based on the treatment policy strategy should appear in the main body of the paper with estimates based on other strategies relegated to supplementary materials. Treatment-policy estimands target the effect of assignment to treatment and are defined to include outcomes after post-randomization events, such as treatment discontinuation or initiation of alternative or rescue medications. While appropriate for certain purposes, they do not address all clinically relevant questions. Results based on estimands using alternative strategies should therefore also be presented in the main body of the published paper when they address important clinical questions that are directly relevant to patient care or decision-making.

Different questions, different effects

The PIONEER-1 trial,[3] which compared semaglutide with placebo for glycemic control in Type 2 diabetes, illustrates this point. Rescue medication is routinely given to patients with persistent hyperglycemia, and additional glucose-lowering therapies could also be prescribed after discontinuation of randomized treatment. Use of such therapies was substantially more frequent among placebo recipients than semaglutide recipients: 19.7% in the placebo group compared with 9.1%, 4.6% and 4.0% in the semaglutide 3-mg, 7-mg and 14-mg groups, respectively.[3] These additional therapies comprised multiple drug classes, including biguanides, sulfonylureas, insulin, DPP-4 inhibitors, and thiazolidinediones.

Several distinct questions arise. Is the treatment effect of interest the effect of initiating semaglutide, potentially followed by different glucose-lowering medications, compared to placebo potentially followed by different glucose-lowering? Or is it the effect of semaglutide itself versus placebo, absent use of these additional medications? Similarly, is the estimand of interest the effect of assignment to

semaglutide ignoring whether the course of treatment was discontinued, or the effect that would be observed if patients adhered to treatment as prescribed?

Each of these important clinical questions corresponds to different causal estimands. Addressing these important clinical questions may require readily available estimates of treatment effect for the different estimands in publications reporting trial results.

Beyond the ITT / per-protocol dichotomy

Historically, analyses have often been framed as a dichotomy between ITT and per-protocol analyses. These are sometimes expressed as estimating either the effect of assignment to intervention or the effect of adhering to the intervention as specified in the protocol (per-protocol effect). For example, the Cochrane Risk of Bias tool[4] describes these effects as: "(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')" However, there are important clinical questions which are not addressed by either of these analyses.[5]

Reference to per-protocol effects can easily be confused with estimation using a per-protocol analysis set, which is known to be a biased approach to estimation. Moreover, adherence to an intervention is an ambiguous concept as it may refer to compliance with the randomized treatment, adherence to a broader treatment strategy, and/or adherence to the protocol procedures.

The ICH E9(R1) estimands framework resolves much of this ambiguity by requiring explicit identification of intercurrent events and specification of strategies for handling them, thereby allowing the clinical questions of interest to be precisely defined. Hernán and Robins[6] argued that randomized trials are not limited to estimating ITT effects and that restricting inference to only the treatment effect of randomisation assignment risks answering a less relevant clinical question.

As well as the treatment policy strategy, the ICH E9(R1) estimands framework describes several alternative strategies. Hypothetical, principal-stratum, and while-on-treatment strategies can all be used to estimate well-defined causal effects of treatment adherence, provided that the required assumptions are clearly stated and justified.[7]

The composite strategy can provide an alternative to treatment policy by incorporating events such as initiation of rescue or additional medication directly into the endpoint definition. Outcomes cannot be observed after death, making a treatment-policy strategy impossible for that intercurrent event. The composite approach explicitly recognises these events as clinically meaningful indicators of treatment failure. By contrast, treatment-policy includes outcomes after initiation of rescue or additional medication without accounting for the lack of efficacy that prompted their use.[8]

Hypothetical strategy and reality

Troxel et al. [1] state that an effect estimate based on a hypothetical strategy is "often of limited value because the scenario didn't occur in reality." This criticism conflates *counterfactual* with *irrelevant*.

The fact that a clearly articulated hypothetical scenario did not fully occur during the trial does not diminish its scientific or clinical value. Rather, such estimands explicitly address questions that trials are often designed to inform but cannot directly observe. By contrast, treatment-policy estimands may reflect a composite of downstream interventions and patient behaviors that obscure the effect attributable to the randomized treatment itself. Estimands targeting the causal effect of adhering to

treatment therefore often provide information that is more directly relevant to patients and prescribers and may also inform regulatory decision making when understanding the efficacy attributable to a specific medicine rather than assignment alone, while facilitating more meaningful comparisons across trials. For example, the European regulatory (CHMP) guideline for diabetes describes using the hypothetical strategy that rescue medication was not initiated for participants who take it.[9]

Assumptions are unavoidable

Troxel et al. [1] express concern that estimation use of hypothetical estimands requires assumptions that "*generally cannot be verified or refuted with available data*". In the presence of missing data, estimation of effects using the treatment policy strategy is also difficult and also typically relies on assumptions that cannot be verified with available data.

For a treatment policy estimand, every attempt should be made to collect data after intercurrent events to minimise issues relating to missing data. However, nearly all clinical trials will continue to encounter missing data and in some therapeutic areas—such as psychiatry—retention is particularly challenging, leading to extensive post-discontinuation missing data that must be imputed. Troxell et al state that estimation of treatment-policy estimands must therefore proceed under "*appropriate context-specific assumptions".* It is unclear how to determine the "appropriate context-specific assumptions" for a participant's outcome given that once they have left the trial the participant could receive interventions which are often heterogeneous, poorly characterized, and influenced by clinical judgment. This is a considerably more complicated problem than determining the best projection had they continued to take the assigned treatment and there may be a shortage of observed data within the trial to allow estimation that conditions fully on all interventions taken.

Bell and colleagues[10] note that while treatment policy estimands are easy to specify they are hard to estimate reliably when the primary outcomes of interest are not observed for all trial participants. Clinical researchers face a dilemma between making strong assumptions that risk bias or weaker assumptions that lead to imprecision and loss of power.

The relevant comparison, therefore, is not between assumption-free and assumption-laden estimands, but between assumptions that are explicit and aligned with the clinical question, and assumptions that are implicit and poorly understood.

Clinical relevance and pluralism

Troxel et al. caution against misinterpretation of non-ITT estimands and recommend de-emphasizing them in primary publications. We agree that clear definition and clinical explanation are essential. However, restricting the main body of the publication to ITT estimands may not align with trial objectives. For example, sponsors, in discussion with regulatory agencies and following clinical guidance documents, may prespecify estimands other than treatment policy as primary estimands to address key regulatory questions particularly when the goal is to understand the efficacy attributable to the intervention itself rather than the effect of assignment alone. If the recommendation of Troxel et al[2] that journals publish only results from the treatment policy estimand in the main body of the paper, then a pre-specified primary estimand and its associated analysis, agreed in advance with the European regulatory body, could be relegated to the supplementary materials.

Different stakeholders reasonably prioritize different questions, and the estimands framework was designed to accommodate this diversity. A pluralistic approach, reporting multiple well-justified estimands with transparent assumptions, can provide additional insights.

Conclusion

Randomized clinical trials can answer multiple causal questions, provided those questions are explicitly and precisely defined, an aligned estimator can be applied and the assumptions required to estimate them are made transparent. Randomization identifies the causal effect of assignment, but it does not restrict investigators to that effect alone.

Rather than treating estimands other than treatment policy as secondary or misleading, they should be recognized as legitimate—and often essential—complements to treatment policy analyses, particularly for clinicians and patients deciding whether, and how, to use a medicine and for regulators to decide whether to license a medicine. Consequently, there should be no requirement for journals to limit the main text of publications to only results based on the treatment policy strategy.

Declaration of interests: ONK and PT are independent consultants to the pharmaceutical industry. HL, CF and DW are employed by pharmaceutical companies as declared above and hold shares in these companies.

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 are those of the authors and may not represent those of other members of the group or of the companies with which the authors are affiliated.

References

1. Troxel AB, Gatsonis CA, Hogan JW, Hubbard RA, Hunter DJ, Normand SL. Statistics in Medicine—What's in an Estimand?. New England Journal of Medicine. 2026 Jan 1;394(1):6-9.
2. 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.
3. Aroda VR, Rosenstock J, Terauchi Y, Altuntas Y, Lalic NM, Morales Villegas EC, Jeppesen OK, Christiansen E, Hertz CL, Haluzík M. 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 Sep 1;42(9):1724-32.
4. 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.
5. Keene ON, Lynggaard H, Englert S, Lanius V, Wright D. Why estimands are needed to define treatment effects in clinical trials. BMC medicine. 2023 Jul 27;21(1):276.
6. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med. 2017 Oct 5;377(14):1391-8.
7. Drury T, Bartlett JW, Wright D, Keene ON. The estimand framework and causal inference: Complementary not competing paradigms. Pharmaceutical statistics. 2025 Sep;24(5):e70035.
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. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus - Revision 2. CPMP/EWP/1080/00 Rev. 2. Published Jun 28, 2023. Effective Jan 1, 2024. Accessed Feb 19, 2026.
10. 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.