

COMMENTARY

Some considerations on target estimands for health technology assessment

Antonio Remiro-Azócar 

Medical Affairs Statistics, Bayer plc, Reading, UK

Correspondence

Antonio Remiro-Azócar, Medical Affairs Statistics, Bayer plc, 400 South Oak Way, Reading RG2 6AD, UK.

Email: antonio.remiro-azocar@bayer.com

First and foremost, I would like to thank an anonymous Associate Editor and Prof. Nigel Stallard for arranging a fascinating discussion around my article, “Target estimands for population-adjusted indirect comparisons.”¹ This is based on a prior exchange with Phillippo et al.²⁻⁴ I extend my gratitude to Russek-Cohen,⁵ Schiel,⁶ Senn,⁷ Spieker,⁸ and Van Lancker et al,⁹ for their additional contributions.

This rejoinder discusses the potential development of an estimands framework in the context of evidence synthesis and health technology assessment (HTA). I consider the following base-case scenario. An evidence synthesis (eg, an indirect treatment comparison or a network meta-analysis) is required for HTA. The evidence synthesis combines the results of multiple randomized controlled trials (RCTs).¹⁰⁻¹² Each RCT has been designed for regulatory approval in the premarketing authorization setting, and has target estimands of its own. There are exceptions to this scenario; regulatory decisions are not exclusively based on RCTs and HTA decisions are not exclusively based on evidence synthesis. Nevertheless, it is a common scenario.

Two challenges arise. First, different policy decisions will require different research questions; the estimands targeted for regulatory approval will likely not align with those targeted for HTA. Second, within the evidence synthesis, there may be misalignment between the estimands targeted by individual studies. Finally, I reassert that marginal estimands should be preferred in the context of the original discussion.²⁻⁴

1 | HARMONIZATION BETWEEN DECISION-MAKING CONTEXTS

For RCTs in the regulatory setting, the use of estimands has been stimulated by the recent publication of the E9 (R1) Addendum, issued by the International Council of Harmonisation (ICH) and recognized by agencies such as the United States Food and Drug Administration.¹³ According to the ICH guidance, an estimand should be defined on the basis of five attributes: (1) treatment(s); (2) target population; (3) clinical outcome of interest; (4) population-level summary effect measure; and (5) strategy for intercurrent (post-randomization) events. My original commentary¹ places focus on the second and fourth elements. It does not intend to cover other dimensions, which are also important. The application of the ICH guidance to individual RCTs has been addressed in depth elsewhere.^{14,15}

In evidence synthesis, and more generally in HTA reporting, the term *estimand* is rarely used. This does not imply a lack of interest in clarifying the research question prior to the analysis. In HTA, the PICO framework is typically used to translate policy questions into research questions. PICO consists of four components: (1) population; (2) intervention; (3) comparator(s); and (4) outcome. To obtain reimbursement for a given treatment, the manufacturer must submit an evidence dossier that addresses the research question(s) included in the scope. Current best-practice guidelines recommend prespecifying the relevant PICO question(s) in the HTA scoping process.¹⁶⁻¹⁸ In evidence synthesis, PICO questions are often formulated prior to the analysis in order to guide the data extraction required for systematic literature reviews.¹⁹

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Author. *Statistics in Medicine* published by John Wiley & Sons Ltd.

The PICO framework is a valuable tool. Nevertheless, it needs refinement to more precisely characterize the inferential target in HTA; that is, to map the HTA research question into a formal estimand. The current framework does not discuss the population-level measure used to summarize the treatment effect. It also lacks a strategy for intercurrent events, which as stated by Van Lancker et al,⁹ “are at the heart of the ICH E9 (R1) addendum.” I would propose a PICOSI framework for describing the HTA research question: (1) population; (2) intervention; (3) comparator(s); (4) outcome; (5) summary effect measure; and (6) intercurrent events. This would make the summary measure used to compare treatments and the intercurrent event strategy explicit. It would also allow stakeholders to translate regulatory language to HTA terminology.

My original commentary¹ and prior exchange with Phillipppo et al²⁻⁴ highlight the importance of prespecifying the summary effect measure. It is necessary to make a distinction between marginal and conditional effect measures. I argue that the former are more relevant to the HTA setting.^{1,3} Another question that arises is whether relative estimands (treatment effects) or absolute estimands (mean outcomes) are more appropriate. When the goal is to estimate comparative effectiveness, relative estimands are the logical choice. Nevertheless, when the aim is to estimate comparative cost-effectiveness, absolute measures (eg, event probabilities) are often input to health economic models.⁴ The transportability of relative and absolute measures between populations depends on different conditions.²⁰ Another important matter is that the evaluation of effect modifier status, necessary to assess generalizability, transportability and external validity, is tied to the scale used to measure the relative effect.²¹ Effect modifiers cannot be selected prior to the specification of a summary effect measure of interest.

Fundamentally, the estimands required for regulatory and HTA decision-making are expected to diverge. Regulatory decisions are made prior to HTA decisions. Individual studies will likely be completed at the time of undertaking the evidence synthesis, and estimand definitions for the former will precede estimand definitions for the latter.

Moreover, regulatory and HTA decision-makers have different perspectives. The HTA paradigm places emphasis on evaluating “real-world” effectiveness, whereas regulators typically prioritize internal over external validity when assessing clinical efficacy. In terms of intercurrent events, estimands based on the “treatment policy” and the “composite” strategies seem more relevant for HTA purposes. Nevertheless, regulators may prefer other strategies in certain scenarios. In terms of comparator(s), HTA demands comparing the intervention under study versus all alternative treatments for a specific disease state. Conversely, regulators typically require a head-to-head comparison against placebo or standard of care.

In addition, there is a multiplicity of decision-makers within HTA. Different agencies will target different populations and, therefore, different estimands. HTA policy decisions typically consider external validity, but external validity depends on the population that is targeted by the payer. Given the large number of regulatory bodies and national and regional HTA agencies, I coincide with Schiel: we should avoid focusing exclusively on the viewpoint of a single stakeholder when designing studies.⁶ Traditionally, HTA objectives have played a limited role in the design of RCTs. Potential HTA research questions should be posed in the early stages of drug development, so that these can be answered at the time of HTA decision-making.^{22,23}

2 | STANDARDIZATION WITHIN THE EVIDENCE SYNTHESIS

Apart from mismatches in estimands between decision-making contexts, there may also be inconsistencies between the trial-specific estimands combined in an evidence synthesis. As expressed by Russek-Cohen, “individual studies were not planned with similar estimands nor they were necessarily planned in anticipation of a meta-analysis.”⁵ There may be mismatches in endpoint definitions, treatment implementations, study target populations, summary effect measures or intercurrent event strategies, that invalidate the results of an evidence synthesis.

The availability of standardized outcome and treatment definitions depends on the clinical indication and the interventions being evaluated. Differences in endpoints may arise if outcomes are measured using different methods, at different times, or if the trials have different duration. When a treatment is assessed in multiple studies, there may be dissimilarities between the versions of the intervention being administered, for example, in dosing formulation, delivery mechanism, co-treatment regimens, or concomitant background medications. Clinical input is critical to assess the potential impact of these sources of misalignment on the conclusions of an evidence synthesis.

The studies being synthesized may target populations that differ in terms of their baseline characteristics, due to different inclusion/exclusion criteria. Even if inclusion/exclusion criteria are identical, the covariate distribution of

the sampled patients will likely vary from study to study. These differences may give rise to heterogeneous treatment effect sizes across studies, which threaten the validity of the evidence synthesis. Random effects models have traditionally been used to account for effect heterogeneity.²⁴⁻²⁶ Nevertheless, the conventional approaches do not explicitly produce estimates that are directly interpretable in any particular population.²⁷⁻²⁹ This is problematic, given the central role of the *population* in the estimands framework, and in formulating research questions in evidence synthesis and HTA.

Here lies the value of covariate-adjusted indirect comparisons^{20,30,31} and network meta-regression approaches,^{32,33} such as those examined by Phillippo et al² in their original simulation study. These methods explicitly account for differences in patient characteristics between studies, thereby targeting estimands in specific samples or populations. As a result, they are being increasingly considered by payers.³⁴ In many practical scenarios, patient-level data will be unavailable for certain trials, and the evidence synthesis will be based on published aggregate-level data to some degree. Therefore, in the words of Russek-Cohen, “there needs to be an incentive for summarizing individual trials so they can be more easily studied in meta-analyses that utilize covariates.”⁵

Conflicts in the summary effect measures that are pooled in an evidence synthesis are also a cause for concern. When the effect measure is non-collapsible, combining marginal and conditional estimates will produce bias.¹² Non-collapsible effect measures such as (log) odds ratios and (log) hazard ratios are commonly used in network meta-analysis.^{35,36} Finally, there is a need to align within-study intercurrent event strategies prior to an evidence synthesis.

3 | CONCLUDING REMARKS

To conclude, I wish to examine some of the considerations made by Senn⁷ and Van Lancker et al⁹ about conditional and marginal estimands. Senn⁷ argues that “we usually treat individuals not populations.” I am in agreement. That being so, reimbursement decisions in HTA are typically made on populations not single patients. At the time of decision-making, the target population for the therapeutic indication is “stable” and well-defined, having been prespecified in the PICO question on the scope of the HTA.

The target population of interest is not necessarily broad. Often, it is highly selected and takes into account specific patient characteristics. For instance, the final scope of a health technology appraisal for abemaciclib (TA810), recently published by the National Institute for Health and Care Excellence, describes the target population as “adults with hormone receptor-positive, HER2-negative, node-positive early breast cancer after definitive surgery of the primary breast tumor at high-risk of recurrence.”³⁷

A preference for marginal estimands does not necessarily imply that an “all or nothing” decision is required or that “homogeneity” is assumed. When subsets of the target population are considered to benefit differently from the reimbursement of a drug, relevant sub-populations are defined during the scoping process, in which case separate PICO questions can be formulated for each. In this case, the covariates used to define the subsets are binary or categorical, and the distinction between marginal and conditional estimands is blurred. A conditional estimand at the population level can be viewed as a marginal estimand at the sub-population level.

Nevertheless, the specific context—covariate-adjusted indirect comparisons in HTA—is central to the debate. Consider the situation described by Phillippo et al in their original simulation study:² (1) covariates are continuous; (2) model-based covariate adjustment is required for transportability; (3) valid estimation relies on correct model specification. As suggested by Spieker,⁸ a “modeling stage” is necessary to select the covariate adjustment model. When the adjustment model is an outcome regression, analysts will seek goodness-of-fit to the observed data. In this setting, the conditional estimand would align to the selection of the estimator.

However, the estimands framework demands that the estimand is articulated first, prior to any estimation. Here, the definition of a conditional estimand would depend on statistical modeling assumptions. Conversely, the marginal estimand can be unambiguously interpreted without reference to a particular adjustment model. I believe that this point is implicit throughout Van Lancker et al’s discussion.⁹ Ultimately, the marginal estimand is more compatible with the estimands framework, because it is the inferential target that allows one to separate the research question from the method of statistical analysis.

ORCID

Antonio Remiro-Azócar  <https://orcid.org/0000-0002-2877-2315>

REFERENCES

1. Remiro-Azócar A. Target estimands for population-adjusted indirect comparisons. *Stat Med*. 2022.
2. Phillippo DM, Dias S, Ades A, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med*. 2020;39(30):4885-4911.
3. Remiro-Azócar A, Heath A, Baio G. Conflating marginal and conditional treatment effects: comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study". *Stat Med*. 2021;40(11):2753-2758.
4. Phillippo DM, Dias S, Ades AE, Welton NJ. Target estimands for efficient decision making: response to comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study". *Stat Med*. 2021;40(11):2759-2763.
5. Russek-Cohen E. Discussion of "Target estimands for population-adjusted indirect comparisons" by Antonio Remiro-Azocar. *Stat Med*. 2022.
6. Schiel A. Commentary to target estimands for population-adjusted indirect comparisons. *Stat Med*. 2022.
7. Senn S. Conditions for success and margins of error: estimation in clinical trials. *Stat Med*. 2022.
8. Spieker AJ. Comments on the debate between marginal and conditional estimands. *Stat Med*. 2022.
9. Van Lancker K, Vo TT, Akacha M. Discussion on "Target estimands for population-adjusted indirect comparisons". *Stat Med*. 2022.
10. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683-691.
11. Dias S, Sutton AJ, Welton NJ, Ades A. Evidence synthesis for decision making 3: heterogeneity—Subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Mak*. 2013;33(5):618-640.
12. Remiro-Azócar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Res Synth Methods*. 2021;12(6):750-775.
13. Mehrotra DV, Hemmings RJ, Russek-Cohen E, Group IEEW. Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. *Clin Trials*. 2016;13(4):456-458.
14. Akacha M, Bretz F, Ohlssen D, Rosenkranz G, Schmidli H. Estimands and their role in clinical trials. *Stat Biopharm Res*. 2017;9(3):268-271.
15. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials—broadening the perspective. *Stat Med*. 2017;36(1):5-19.
16. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value Health*. 2011;14(4):417-428.
17. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. *Value Health*. 2011;14(4):429-437.
18. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC good practice task force report. *Value Health*. 2014;17(2):157-173.
19. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak*. 2007;7(1):1-6.
20. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Mak*. 2018;38(2):200-211.
21. Brumback B, Berg A. On effect-measure modification: relationships among changes in the relative risk, odds ratio, and risk difference. *Stat Med*. 2008;27(18):3453-3465.
22. Jönsson B. Bringing in health technology assessment and cost-effectiveness considerations at an early stage of drug development. *Mol Oncol*. 2015;9(5):1025-1033.
23. Wang T, McAuslane N, Gardarsdottir H, Goettsch WG, Leufkens HG. Building HTA insights into the drug development plan: current approaches to seeking early scientific advice from HTA agencies. *Drug Discov Today*. 2022;27(1):347-353.
24. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.
25. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials*. 2007;28(2):105-114.
26. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010;29(29):3046-3067.
27. Dahabreh IJ, Petito LC, Robertson SE, Hernán MA, Steingrimsson JA. Towards causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiology*. 2020;31(3):334.
28. Sobel M, Madigan D, Wang W. Causal inference for meta-analysis and multi-level data structures, with application to randomized studies of viox. *Psychometrika*. 2017;82(2):459-474.
29. Barker DH, Dahabreh IJ, Steingrimsson JA, et al. Causally interpretable meta-analysis: application in adolescent HIV prevention. *Prev Sci*. 2022;23(3):403-414.
30. Signorovitch JE, Wu EQ, Andrew PY, et al. Comparative effectiveness without head-to-head trials. *PharmacoEconomics*. 2010;28(10):935-945.
31. Remiro-Azócar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Res Synth Methods*. 2022.
32. Phillippo DM, Dias S, Ades A, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc A Stat Soc*. 2020;183(3):1189-1210.
33. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12(1):1-16.

34. Phillippo DM, Dias S, Elstner A, Ades A, Welton NJ. Population adjustment methods for indirect comparisons: a review of national institute for health and care excellence technology appraisals. *Int J Technol Assess Health Care*. 2019;35(3):1-8.
35. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(16):2313-2324.
36. Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. *BMC Med Res Methodol*. 2010;10(1):1-9.
37. NICE. Abemaciclib with endocrine therapy for adjuvant treatment of hormone receptor-positive, HER2-negative, node-positive early breast cancer at high risk of recurrence; 2022. Accessed August 1, 2022. <https://www.nice.org.uk/guidance/ta810>

How to cite this article: Remiro-Azócar A. Some considerations on target estimands for health technology assessment. *Statistics in Medicine*. 2022;41(28):5592-5596. doi: 10.1002/sim.9566