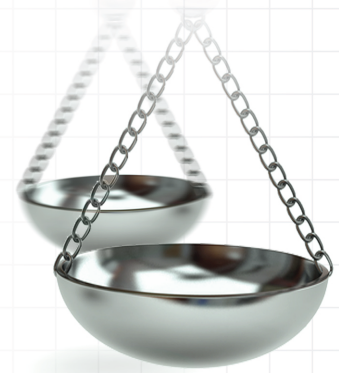




# Estimating per-protocol effects in external comparator analyses using real-world data

Alind Gupta<sup>\*,1,2</sup> , Evie Merinopoulou<sup>3</sup>, Stephen J Duffield<sup>4</sup>, Manuel Gomes<sup>5</sup> ,  
Seamus Kent<sup>6</sup> , Nicolas Scheuer<sup>7</sup>, Gerardo Machnicki<sup>8</sup>  & Thibaut Sanglier<sup>8</sup> 

<sup>1</sup>Fifty1 AI Labs Vancouver, British Columbia. V6H 3Y4 Canada

<sup>2</sup>Dalla Lana School of Public Health, Department of Epidemiology, University of Toronto, Canada

<sup>3</sup>Cytel Inc., London, UK

<sup>4</sup>National Institute for Health & Care Excellence, Manchester, UK

<sup>5</sup>University College London, Primary Care & Population Health, UK

<sup>6</sup>Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Netherlands

<sup>7</sup>Roche Products Ltd, Welwyn Garden City, UK

<sup>8</sup>F. Hoffmann-La Roche Ltd., Basel, Switzerland

\*Author for correspondence: [alind.gupta@utoronto.ca](mailto:alind.gupta@utoronto.ca)

Journal of **Comparative  
Effectiveness Research**

Analysis of single arm trials complemented with external comparator arms (ECAs) may be used to support evidence of effectiveness of novel therapies in oncology research when a randomized trial is unavailable or unfeasible. However, the intention-to-treat effect, which is a common target of estimation in ECA studies, is difficult to interpret when there are differences in adherence between the trial and ECA. This paper describes an approach to estimation of per-protocol effects in ECA studies using the target trial emulation framework for study design and analysis based on the results from an exploratory case study (TBASEL). We highlight challenges, potential solutions and future opportunities from the perspectives of protocol specification, data suitability and analysis, to help guide future implementations of per-protocol effects in ECAs.

First draft submitted: 2 April 2025; Accepted for publication: 13 August 2025; Published online: 6 October 2025

**Keywords:** adherence • comparative effectiveness • external controls • per-protocol effects • target trial emulation

Randomized clinical trials (RCTs) are the gold standard for generating evidence on comparative effectiveness of novel therapies, and the foundation for decision-making in many settings including health technology assessments (HTA). HTA involves systematically evaluating evidence to guide healthcare decisions regarding new technologies like drugs and medical devices, including decisions about reimbursement [1]. The value of large RCTs stems from their ability to eliminate confounding, on average, through randomization of treatment assignment at baseline. This provides an idealized and carefully monitored setting for estimating treatment effectiveness, usually under high, but rarely perfect, treatment adherence.

The most commonly estimated treatment effect in RCTs is known as the intention-to-treat (ITT) effect. The ITT effect aims to estimate the effect of treatment assignment at baseline regardless of subsequent adherence. The ITT effect is often preferred because it is easy to estimate in RCTs with baseline randomization and low study dropout rate. It also incorporates a measure of the imperfect adherence that would be expected in routine clinical care and is generally considered to be a more conservative estimate of the treatment effect because it combines the treatment effect across both adherent and nonadherent individuals, which is of prime interest to decision makers to inform system-wide decisions. However, an exclusive reliance on the ITT effect for understanding treatment effectiveness can be problematic [2]. For one, defining treatment adherence as a simple dichotomy of whether or not a randomized participant initiates the assignment treatment overlooks the complexity of adherence as a time-varying phenomenon. Adherence is the clinical context for a drug's use following treatment assignment, including when it should be used, modified or discontinued, and the precise manner of its administration is dependent on a patient's time-evolving health status, i.e., the 'protocol'. In HTA, particularly in oncology, the impact of the use of concomitant medications, dosage changes, subsequent lines of therapy and discontinuation, especially if they differ

between treatment arms, are important considerations for an understanding of the real-world benefit, comparative effectiveness and costs and the ITT effect alone may be insufficient for decoupling these.

In the absence of an RCT, HTA bodies are sometimes willing to accept evidence incorporating real-world data in the form of an external comparator arm (ECA) to contextualize single-arm clinical trials [3,4]. The use of ECAs is increasingly common in some advanced cancers and rare diseases where conducting large well-powered RCTs is infeasible [5,6]. A major concern cited with the use of ECAs to estimate the ITT effect is the potential risk of bias due to confounding at baseline, which can be addressed for measured confounders using statistical techniques like matching [7–9]. However, important differences can exist post-baseline in the standards of care between clinical trials and the real world such that the ITT effect does not clearly estimate the clinical benefit of the novel therapy. There may be ‘contamination’ due to crossing-overs, as well as differences in dose adjustments, concomitant therapies, timing and reasons for treatment discontinuations and switching between the treatment groups, partly because patients in clinical trials are monitored better and more directly than in routine care.

Finally, study discontinuation may occur differentially between the trial arm and ECA. These problems with the ITT effect have stimulated the simultaneous estimation of the per-protocol (PP) effect in many target trial emulation studies in recent years [10–13]. In parallel, the estimand framework for RCTs has also been developed to help address post-baseline factors (‘intercurrent events’) that affect clinical outcomes and modify the interpretation of the treatment effect [14], although it has been criticized for including estimands that are not clinically relevant, among other issues [15]. Although existing studies have provided detailed technical descriptions of the PP effect and its estimation, how it differs from the ‘naïve’ PP analysis reported by some RCTs, as well its application to observational analysis of real-world databases, we found only one prior study [16] so far that has attempted to operationalize its use in ECA studies [17–19].

In this study, we attempted to emulate a trial using an ECA, with the goal of illustrating the practical challenges with the specification, estimation and benchmarking of the PP effect. The study was stopped due to the lack of measurement of important data elements necessary for estimation of the PP effect. Here we describe these missing elements and learnings with the view of supporting future ECA analyses, and assessments of their feasibility.

### The TBASEL study: OAK trial emulation for estimating PP effects

We used data from the OAK trial and emulated the control arm using an ECA. We used data from an existing RCT instead of a single arm trial for this emulation, so that we could benchmark the target trial specification and findings of the emulated trial against the actual RCT. The OAK trial was a pivotal phase III clinical study that investigated the efficacy of atezolizumab, an anti-PD-L1 immunotherapy, in patients with non-small cell lung cancer (NSCLC) [20]. The trial enrolled eligible patients who had previously progressed on at least one line of platinum-based chemotherapy. Patients were randomized to receive either atezolizumab or docetaxel, a standard chemotherapy drug. The results showed that atezolizumab significantly improved overall survival compared with docetaxel, leading to its approval for second-line NSCLC and establishing immunotherapy as a key treatment option in lung cancer care.

The OAK trial reported an intent-to-treat (ITT) effect but did not estimate a PP effect. To emulate an ECA, we selected an ECA cohort of docetaxel initiators who met eligibility criteria of the OAK trial that could be operationalized using data from Flatiron Health – a large United States nationwide electronic health records database [21,22]. The target trial specification for the trial and its emulation is documented in Table 1. In this study, we attempted to estimate both the observational analog of the ITT effect from the ECA analysis, which could be benchmarked against the results from the trial, as well as the PP effect, which could not. If the considerations for post-baseline events described previously were important for this specific setting, for example, because they occurred differentially between the treatment groups, the PP effect estimate would be expected to be different from the ITT effect.

In the target trial emulation framework for the design of observational studies, ‘target trial’ represents the RCT that ideally would have been conducted to estimate the causal effect of an intervention and which we are trying to emulate using observational data. In TBASEL, the eligibility criteria and treatment strategies, follow-up and outcomes were aligned to be as close to the OAK trial using measured data. Therefore, the target trial represented OAK, with minimal modifications to allow for estimation of the PP effect. For example, we used the initiated treatment group rather than the randomized treatment group (i.e., excluded those randomized who never started treatment) for comparison with docetaxel initiators in the ECA to estimate the PP effect, and the treatment strategy was defined with respect to initiation rather than assignment. The most important consideration was the

Table 1. Target trial emulation plan summary for the TBASEL study.

| Component                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Target trial protocol (OAK)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Emulated trial using FI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Eligibility criteria</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <ol style="list-style-type: none"> <li>1. Signed informed consent form</li> <li>2. Ability to comply with protocol</li> <li>3. Age <math>\geq 18</math> years</li> <li>4. Histologically or cytologically documented locally advanced or metastatic NSCLC</li> <li>5. Representative FFPE tumor specimens or slides for PD-L1 testing before enrollment</li> <li>6. Disease progression during/following prior platinum-containing regimen, or recurrence within 6 months of platinum-based adjuvant/neoadjuvant/combined modality therapy</li> <li>7. Measurable disease by RECIST v1.1</li> <li>8. ECOG performance status of 0 or 1</li> <li>9. Life expectancy <math>\geq 12</math> weeks</li> <li>10. Adequate hematologic and end organ function within 14 days before treatment</li> <li>11. No active or untreated CNS metastases</li> <li>12. No leptomeningeal disease</li> <li>13. No uncontrolled pleural/pericardial effusion or ascites</li> <li>14. No uncontrolled tumor-related pain</li> <li>15. No uncontrolled hypercalcemia</li> <li>16. No significant cardiovascular disease</li> <li>17. No severe infections within 4 weeks before randomization</li> <li>18. No peripheral neuropathy</li> <li>19. No severe allergic, anaphylactic or hypersensitivity reactions to chimeric/humanized antibodies or fusion proteins</li> <li>20. No history of autoimmune diseases (e.g., myasthenia gravis, SLE, RA, IBD and vasculitis)</li> <li>21. No prior allogeneic bone marrow or solid organ transplantation</li> </ol> | <p>In order:</p> <ol style="list-style-type: none"> <li>1. None (consent)</li> <li>2. None, assumed from physician's decision to treat (protocol compliance)</li> <li>3. Same (age)</li> <li>4. Included patients with prior locally advanced/metastatic NSCLC diagnosed in/after Jan 2011</li> <li>5. Based on recorded PD-L1 expression</li> <li>6. Included patients previously exposed to platinum-based chemotherapy (max 2 prior systemic therapies)</li> <li>7. None, assumed from physician's decision to treat (measurable disease)</li> <li>8. Same (ECOG)</li> <li>9. Not applied but considered in QBA (life expectancy)</li> <li>10. Applied lab cutoffs from CSR within 14 days before time zero (specific labs: neutrophils, WBC, LCA, albumin, platelets, Hb, AST, ALT)</li> <li>11. Excluded if CNS mets diagnosed within 30 days before time zero (ICD: 'secondary malignant neoplasm AND brain')</li> <li>12. Same, using ICD within 30 days before time zero ('leptomeningeal')</li> <li>13. Same, using ICD within 30 days before time zero ('leptomeningeal')</li> <li>14. Same, using ICD within 30 days before time zero ('pain')</li> <li>15. Same, using ICD within 30 days before time zero ('hypercalcemia')</li> <li>16. Same, using ICD codes (cardiovascular disease)</li> <li>17. Same, using ICD within 30 days before time zero ('infection')</li> <li>18. Same, using ICD within 6 months before time zero ('neuropathy')</li> <li>19. Not applied (not available via ICD)</li> <li>20. Same, using ICD within 6 months before time zero ('autoimmune' or listed diseases)</li> <li>21. Not applied (not available via ICD)</li> </ol> |
| <b>Treatment Strategies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <p><b>Initial treatment strategy:</b></p> <p>Arm A: Atezolizumab (1200 mg intravenous) is administered on Day 1 of a 21-day cycle (+/- 3 days).</p> <p>Arm B: Docetaxel (75 mg/m<sup>2</sup> intravenous). Docetaxel is administered on Day 1 of a 21-day cycle (+/- 3 days), as per the locally approved label.</p> <p><b>Subsequent therapies:</b></p> <p>From initial treatment, all subsequent therapies are considered permissible, including noncancer therapies.</p> <p><b>Justifiable treatment termination:</b></p> <ul style="list-style-type: none"> <li>• Event of progression</li> <li>• Unmanageable toxicity</li> </ul> <p>If termination occurred with any documented progression or unmanageable toxicity this is not considered a protocol violation.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <p><b>Initial treatment strategy for comparator arm only:</b></p> <p>Docetaxel initiated within 7-days of meeting eligibility criteria (around 75 mg/m<sup>2</sup> IV, with some permissible leeway). Docetaxel is administered on Day 1 of a 21-day cycle (+7 days).</p> <p>NB: Descriptive results on observed dosing in FI were generated. Based on clinical feedback, patients in FI that were given a dosage much lower than that given in OAK (i.e., 30 mg/m<sup>2</sup>) were dropped from the analyses.</p> <p>The dosing schedule may be relaxed for FI patients, should the resulting ECA be of an insufficient size.</p> <p><b>Subsequent therapies:</b></p> <p>Permissible subsequent therapies matching those given in the target trial, including noncancer concomitant therapies. Given differences in recording or under recording of therapy lines, the maximum leeway allowed is 60 days between end of treatment and start of subsequent treatment. In this event, the relative justification should be documented, and any resulting tradeoffs explored.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Treatment Assignment</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Participants are randomly assigned to a treatment arm at baseline, and treatments are unblinded.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Randomization is assumed conditional on baseline covariates (age, sex, ECOG status, race, smoking history, time since diagnosis, index year, select lab measurements, metastatic sites, PD-L1 expression)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Follow-up Period</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <p>Begins at treatment initiation (index date) between 11 March 2014 and 29 April 2015. The cut-off date is 9 January 2019.</p> <ul style="list-style-type: none"> <li>• It is permitted that the date of randomization be used for the start of follow-up if necessary if initiation occurs shortly (&lt;7 days) after randomization.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <p>Begins from record of treatment initiation (index date) within the observation period of 11 March 2014, through 29 April 2015, and ends at the date of death, last record last record, or 9 January 2019 (whichever comes first).</p> <ul style="list-style-type: none"> <li>• Note: For the trial arm, date of randomization was used as the start of follow-up. Sensitivity analyses for this were performed.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Outcomes</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <p>OS in the trial is defined as time from index date until death due to any cause.</p> <p>Original primary outcomes:</p> <ul style="list-style-type: none"> <li>• OS in the ITT population</li> </ul> <p>Emulated trial outcome:</p> <ul style="list-style-type: none"> <li>• OS in the PP population</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <p>OS in FI (time from index date) until death from any cause.</p> <p>Mortality in FI is based on a composite end point from linked sources and shown to be of high accuracy and reliability [23].</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <p>FI is the abbreviation for data from Flatiron Health used for the ECA.</p> <p>ATE: Average treatment effect; ATT: Average treatment effect on the treated; CNS: Central nervous system; ECA: External comparator arm; ECOG: Eastern Cooperative Oncology Group; ICD: International Classification of Diseases; ITT: Intention-to-treat; NSCLC: Non-small cell lung cancer; OS: Overall survival; PP: Per-protocol; QBA: Quantitative bias analysis; WBC: White blood cell.</p> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

| Table 1. Target trial emulation plan summary for the TBASEL study (cont.). |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Component                                                                  | Target trial protocol (OAK)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Emulated trial using FI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Causal contrasts                                                           | Intention-to-treat effect and the PP effect.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Observational analogs of the intention-to-treat (effect among those who received atezolizumab, that is, treatment effect among the 'treated' or ATT) and PP effects (average treatment effect, or ATE).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Statistical analysis                                                       | ITT and PP effects on mortality will be estimated under both the trial and its emulation. For the trial, hazard ratios comparing assigned treatment strategies will be estimated using a Cox proportional hazards model with time since randomization as the time scale, consistent with the ITT principle.<br>For the PP analysis in the trial, individuals will be censored at the time of protocol deviation, and a Cox model will be used for estimation with artificial censoring at the time of protocol deviation adjusted using inverse probability of censoring weighting. | For the target trial emulation, pooled logistic regression with time since treatment assignment as the time scale will be used to approximate hazard ratios. For the emulated ITT analysis, individuals will be included according to the treatment strategy assigned at baseline, regardless of subsequent adherence, with adjustment for baseline covariates to address confounding to estimate the ATT. For the emulated PP analysis, individuals will be censored at the time of deviation from the assigned treatment strategy, and inverse probability of treatment and censoring weights will be applied to adjust for baseline and time-varying confounding. Weighted pooled logistic models will be used to estimate hazard ratios under the PP effect. |

FI is the abbreviation for data from Flatiron Health used for the ECA.  
 ATE: Average treatment effect; ATT: Average treatment effect on the treated; CNS: Central nervous system; ECA: External comparator arm; ECOG: Eastern Cooperative Oncology Group; ICD: International Classification of Diseases; ITT: Intention-to-treat; NSCLC: Non-small cell lung cancer; OS: Overall survival; PP: Per-protocol; QBA: Quantitative bias analysis; WBC: White blood cell.

specification of the treatment strategy, where we made some assumptions due to the lack of specification about what constituted a protocol violation in the OAK trial.

The major assumption concerned the specification of adherence with respect to treatment discontinuation and switching to a different therapy. In the OAK trial, clinically valid reasons for treatment discontinuation were not explicitly specified, presumably because all reasons for treatment discontinuation in the context of an RCT would be expected to occur for clinically justified reasons, i.e., either progression on protocol therapy, or unmanageable toxicity. For the target trial, we explicitly specified these as clinically valid reasons for treatment discontinuation, and therefore any other reasons for discontinuation, such as loss of coverage, would be considered protocol violations. Similarly, all post-protocol therapies were permissible in the OAK trial. For the emulation, we sought to align the distribution of postdocetaxel therapies received in the ECA to those that patients would have received in OAK. Therefore, for the emulation, permissible subsequent therapies were the set of therapies that were administered to at least one patient in the OAK trial. Although this represents strategies followed in the OAK trial and may not be consistent with common strategies seen in routine clinical care (e.g., there may be additional therapies not given in the OAK trial), this was done to enable (marginal) positivity within the trial arm, that is, that no patient within the trial arm should have had a probability of zero of initiating these permissible subsequent therapies.

For statistical analysis, we artificially censored initiators of atezolizumab or docetaxel in the trial arm or ECA, respectively, at the time that they deviated from the assigned treatment regimen. Any patients who were lost to follow-up were also censored at the last date of study contact. We used inverse probability of treatment weighting to adjust for all measured prognostic baseline variables described in Table 1, and for time-varying predictors of nonadherence as defined in the study protocol, as well as censoring due to losses to follow-up separately by treatment group, to attempt to estimate the PP effect (i.e., hazard ratios for all-cause mortality over the follow-up) within the population of patients who received either atezolizumab or docetaxel. Predictors of treatment discontinuation or switching, were assumed to include tumour progression, adverse events and diagnostic and physiological measures of time-evolving patient health or performance, such as laboratory tests for serum creatinine clearance and time-varying performance scores measures during the follow-up. We also attempted to estimate standardized survival curves adjusted for the same variables. However, primarily due to the lack of measured elements for time-varying variables, such as progression, ECOG scores and laboratory data, throughout follow-up in either the trial or real-world data or both, we were unable to adjust for time-varying predictors of nonadherence and therefore, could not complete the estimation of the PP effect.

In the following sections, we highlight in more detail aspects related to key challenges in the TBASEL study: specifying the target trial protocol and measured data elements. These challenges and potential strategies to address them are summarized in Table 2.

**Table 2. Challenges uncovered in OAK emulation for per-protocol effect estimation.**

|                                                                                                                                                                     | Main challenge with operationalization of PP effect                                                                                                                                                                               | Potential strategy                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Definition of protocol                                                                                                                                              | <ul style="list-style-type: none"> <li>Multiple estimands (choice between idealized estimand or pragmatic)</li> <li>Choice between ATE or ATT estimand</li> </ul>                                                                 | <ul style="list-style-type: none"> <li>Advisable to engage in a multidisciplinary team to define the estimand of interest (depends on decision problem at hand)</li> <li>From an HTA perspective, pragmatic estimands reflecting real-world scenarios may be preferred, seek early scientific advice</li> </ul>                                                                                                                                                                                                             |
| <b>Data challenges</b>                                                                                                                                              |                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Variable measurement in OAK trial (applicable to most trials)                                                                                                       | <ul style="list-style-type: none"> <li>Absence of data capture on patient covariates, such as adverse events and lab values, after drug discontinuation</li> <li>Progression only measured until study discontinuation</li> </ul> | <ul style="list-style-type: none"> <li>Rely on assumptions for treatment discontinuation backed by clinical experts</li> <li>Use published evidence or conduct analyses to establish reasons for discontinuation where data is available</li> </ul> <p>Further perspectives:</p> <ul style="list-style-type: none"> <li>Single arm trial design should facilitate ECA analyses</li> <li>Leverage novel data enhancements with follow-up data collection, e.g., tokenization and linkage of trial patients to RWD</li> </ul> |
| Variable measurement in the ECA                                                                                                                                     | <ul style="list-style-type: none"> <li>Incomplete data on metastases</li> <li>Limited data on reasons for discontinuation</li> </ul>                                                                                              | <ul style="list-style-type: none"> <li>Extract information from available unstructured notes (requires extensive efforts, more reliable)</li> <li>Use of proxies for broad concepts such as safety events recorded using diagnosis codes (harder to operationalize, less reliable)</li> <li>Pre-trained large language models may be useful for both assisting with and validating data extraction, as well as operationalization of proxies for adverse events in clinical notes</li> </ul>                                |
| ATE: Average treatment effect; ATT: Average treatment effect on the treated; ECA: External comparator arm; HTA: Health technology assessment; RWD: Real-world data. |                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

## Defining estimands & aligning on protocol requirements

In the TBASEL study, a considerable portion of time was spent on specification of the estimand of interest, specifically, whether to tailor it seeking an estimate of drug efficacy under idealized conditions, or alternatively under a pragmatic regimen reflecting real-world practice. For instance, because switching to a subsequent therapy can make it difficult to interpret the effect of the study drugs on patient outcomes, artificially censoring them at the time of switching, and then adjusting for the potential selection bias introduced by such censoring using inverse probability weighting, could be useful in a confirmatory clinical trial aimed at demonstrating drug efficacy under sustained treatment throughout the follow-up.

In the TBASEL study, we engaged in lengthy discussions among the multidisciplinary study team composed of pharmacoepidemiologists, statisticians, causal inference experts, clinicians and representatives of a HTA decision maker perspective to define the estimand of interest. Ultimately, the treatment strategies and definitions of protocol violations for the target trial (Table 1) were not derived as such from the protocol for the OAK study. Instead, the protocol was adapted to include additional specifications of nonpermissible subsequent therapies and protocol violations to align subsequent therapies received in the ECA with those observed in the OAK trial. As a pragmatic approach, patients were allowed to initiate any effective drug following progression or unmanageable toxicity. In this study, all subsequent therapies were allowed to reflect a more widely applicable scenario from an HTA perspective. However, it is important to recognize that approved or commonly used therapies may vary across countries, and future applications should be adapted to fit local contexts.

Secondly, the average treatment effect among the treated, or the ATT estimand that results from matching to the trial, which essentially focuses on the treatment effect to the treated (i.e., the trial population), is generally undefined for estimating the PP effect of a time-varying treatment regimen [24]. Adherence is a function of time-varying factors, such as time-evolving health state and toxicity, and in a time-varying setting, the subset of individuals who are adherent at any time may change over time and therefore there is no definite 'adherent' or 'treated' group. Because we cannot estimate the ATT per protocol effect, the average treatment effect in the population was targeted in the TBASEL study.

## Data elements

For the TBASEL study, an important barrier to estimating PP effects was related to limitations of data measurement of key elements in the trial as well as the real-world database [25].

Limitations in OAK clinical trial data, which we expect apply to most clinical trials datasets, encompassed the absence of information about the reasons for discontinuation of the relevant treatment strategies, such as adverse events and lab values. As RCTs traditionally tend to focus on the ITT effect, collecting information about the reasons for treatment discontinuation is often not a priority for the study investigators. This lack of granularity made it difficult to operationalize the PP effect in TBASEL. Our scoping review and discussion with clinical experts suggested that in patients with advanced/metastatic NSCLC, treatment discontinuations for taxanes and immunotherapies occur overwhelmingly for clinically valid reasons including disease progression and toxicity (unpublished results from targeted review of the literature).

In the TBASEL study, although we could not accurately establish the reasons for treatment discontinuation, expert input and published evidence suggested that we might assume treatment discontinuation should not be considered a protocol violation. However, we felt this approach was less than ideal and ultimately chose not to proceed with the emulation based on this assumption. Progression was also only measured until study discontinuation and could not be ascertained for subsequent on-protocol therapies received by the patient.

We also encountered challenges in variable measurement in the ECA. In the Flatiron real-world data used in TBASEL information on metastases was incomplete due to under recording of diagnoses via ICD codes. Additionally, there was limited data on reasons for treatment discontinuation, which is crucial for assessing adherence to a treatment strategy. Although such information may be available in unstructured notes, it was not fully extracted in a structured way in the data available, complicating the determination of clinically valid discontinuation reasons. Furthermore, we considered the use of proxies for broad concepts such as safety events recorded using ICD codes. Although safety events have specific definitions in clinical trials, they often differ significantly from the use of ICD codes, lab values, and physicians' notes, making them difficult to harmonize, and therefore we determined that it is not possible to reliably operationalize those in TBASEL. Finally, while progression was measured, it may not be complete or as reliably measured as in OAK, making it difficult to harmonize. Indeed, progression is typically assessed using protocol-defined imaging schedules in RCTs, while in real-world data, progression is recorded based on clinical judgment during routine visits, leading to differences in assessment frequency and timing.

## Discussion

There is interest in the adoption of PP effect estimates to complement the ITT effect from researchers and decision-makers, which may remain the primary measure of interest for both regulatory and HTA decision-making. While there is a body of literature emphasizing the importance of PP effects and their interpretation for decision-making, especially concerning adherence, emerging literature on estimating PP effects in settings with fit-for-purpose longitudinal data has primarily focused on analyses of real-world databases, not ECA analyses.

In the TBASEL project, we encountered challenges related to the protocol design and availability of measured elements for analysis. Consistency, no interference, positivity and exchangeability (no unmeasured confounding) are assumed for identification of causal effects; in this study, conditional exchangeability was a highly tenuous assumption. For PP analyses, correct model specification for IPTW and handling of censoring at protocol deviations were additionally assumed. The challenges discussed in this paper and illustrated in the TBASEL case study are not insurmountable, but they require additional efforts. From the perspective of single-arm trials, they should be designed to facilitate comparative analyses with ECAs using target trial emulation at as early a step as possible. This entails planned measurements for any elements that can help alleviate concerns about bias, such as measuring important predictors of treatment assignment and drop out over time. It also involves anticipatory steps to enhance the emulation of the target trial, including the use of eligibility criteria that can be feasibly emulated with real-world data, as well as what constitutes protocol violations and measuring factors related to protocol violations, such as patient health and progression beyond treatment discontinuation. As operationally additional data enhancements may not be realistic, advances with data tokenization and linkages of trial patients to real-world databases could be leveraged for further follow-up information.

With regards to real-world data used for selecting ECAs, naturally all databases will have some limitations around availability or sufficient capture or accurate measurement. Addressing this may require linking multiple databases and performing sensitivity analyses for residual sources of bias. For example, limitations in cancer registry data for capture of detailed treatment pathways or clinical variables could be addressed by linking subjects to data from administrative claims or hospital electronic health record databases. Second, for PP estimation, there may additionally be a need to extract reasons for treatment discontinuation and their relevance to protocol violations from clinical notes. Although this can be a substantial undertaking, one way to reduce the scope of this exercise may

be to check assumptions encoded in the target trial specification for a random subset of patients in the real-world cohort. This should be done in consideration to guidance from decision-making bodies on the choice of a fit for purpose dataset used for data scoping and collection, and the analytical methods [26].

Based on our experience, there will usually be sources of bias that cannot be addressed completely given the available data at hand. If the residual bias is deemed too great to be addressed analytically, it may be possible to change the causal question to be more compatible with measured data, or to collect new data. If not, one should use statistical methods (e.g., quantitative bias analysis) to quantify the risk of residual sources of bias or uncertainty instead of merely stating limitations of measured data qualitatively to help contextualize study results [27]. Care must be taken to incorporate uncertainty associated with differences between the setting of the ECA study and external information, and alignment of study parameters from those estimated from published data [28]. External information can also inform the reasons for treatment discontinuation and any related sensitivity analyses, including the assumption that any treatment discontinuation, change in dose or frequency, or treatment switching is clinically valid.

## Conclusion

The TBASEL study was an attempt at an ECA analysis in oncology to estimate PP effects. The choice of a clinically relevant treatment regimen, as well as retrospective data limitations limited our ability to operationalize PP effects. While the ITT estimand is of primary interest to HTA decision-makers, future research should explore how the PP effect could complement ITT analysis. The PP effect may offer valuable additional insights for clinicians by focusing on the ‘effect of taking’ a drug, rather than the ‘effect of prescribing’ it, as captured by ITT. Additionally, effective dissemination in both written and visual formats that are clear and comprehensible to nonexperts is crucial to instill trust and enhance associated discussions.

There is a need for clear guidance on the design and analysis of ECA studies using real-world data (RWD) in general so that they can provide sound evidence for decision-making. There will inevitably be differences across jurisdictions regarding the minimum required standards of real-world evidence to inform decision-making. With the increase in cross-collaborative networks and joint committees, such as EUnetHTA [29], as well as the regulation on HTA [30] that will replace it, this guidance should become more aligned in the future. Complementary to these initiatives, additional demonstration and benchmarking studies [31,32], and improvements to reporting [33] in RWD-based ECA analyses are warranted.

### Executive summary

- External comparator arms (ECAs) can contextualize single-arm oncology trials when randomized trials are infeasible, but they pose challenges for per-protocol (PP) effect estimation.
- Defining the estimand and protocol adherence requirements is critical, requiring multidisciplinary input and clear alignment with data availability.
- The TBASEL study attempted to estimate PP effects using the target trial emulation framework, benchmarking against the OAK trial in advanced non-small cell lung cancer.
- Missing time-varying data on progression, lab values and treatment discontinuation reasons in trials and real-world data limited PP estimation feasibility.
- Real-world data often lacks harmonized measures for post-baseline treatment and adherence needed for robust PP analyses.
- Emulating target trials with ECAs requires early planning to align eligibility criteria, protocol violations and data collection with PP estimation needs.
- Residual bias due to unmeasured confounding and incomplete data may necessitate quantitative bias analyses to contextualize results.
- While ITT remains the primary estimand for HTA decisions, complementary PP estimates can inform clinicians on the effect of taking treatment under real-world adherence.
- Clearer guidance, methodological demonstration and reporting improvements are needed to advance the use of PP estimation in ECA studies using real-world data.

## Acknowledgments

We thank Gerald Smith, Cytel, for his invaluable contributions to the design and implementation of the TBASEL emulation. We also thank Isaac Gravestock, F. Hoffmann-La Roche Ltd., for providing methodological input. Finally, we are grateful to Riley Geason, Cytel, for his exceptional editorial support.

### Financial disclosure

This study was funded by Roche.

### Competing interests disclosure

N Scheuer and T Sanglier report stock ownership in Roche. T Sanglier and G Machniki are employees of Roche. N Scheuer was an employee of Roche at the time of this study. The authors have no other competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript apart from those disclosed.

### Data sharing statement

The authors certify that this manuscript reports the secondary analysis of clinical trial data that have been shared with them, and that the use of this shared data is in accordance with the terms (if any) agreed upon their receipt. The source of this data is NCT02008227.

### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Goodman CS. *HTA 101: introduction to health technology assessment*. National Institutes of Health, MD, USA (2014).
2. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin. Trials* 9(1), 48–55 (2012).
- **Highlights the limitations of relying solely on the intention-to-treat (ITT) effect and introduces practical considerations for per-protocol (PP) analyses in comparative effectiveness research.**
3. Patel D, Grimson F, Mihaylova E *et al.* Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health* 24(8), 1118–1125 (2021).
4. Sola-Morales O, Curtis LH, Heidt J *et al.* Effectively leveraging RWD for external controls: a systematic literature review of regulatory and HTA decisions. *Clin. Pharmacol. Ther.* 114(2), 325–355 (2023).
5. Seeger JD, Davis KJ, Iannacone MR *et al.* Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. *Pharmacoepidemiol. Drug Saf.* 29(11), 1382–1392 (2020).
6. National Institute for Health and Care Excellence. Use of a real-world data external control arm. Available from: <https://www.nice.org.uk/corporate/ecd9/chapter/introduction-to-real-world-evidence-in-nice-decision-making>
7. Mishra-Kalyani PS, Amiri Kordestani L, Rivera DR *et al.* External control arms in oncology: current use and future directions. *Ann. Oncol.* 33(4), 376–383 (2022).
8. Jaksa A, Louder A, Maksymiuk C *et al.* A comparison of seven oncology external control arm case studies: critiques from regulatory and health technology assessment agencies. *Value Health* 25(12), 1967–1976 (2022).
9. Wang X, Dormont F, Lorenzato C, Latouche A, Hernandez R, Rouzier R. Current perspectives for external control arms in oncology clinical trials: analysis of EMA approvals 2016–2021. *J. Cancer Policy* 35, 100403 (2023).
10. Boyne DJ, Brenner DR, Gupta A *et al.* Head-to-head comparison of FOLFIRINOX versus gemcitabine plus nab-paclitaxel in advanced pancreatic cancer: a target trial emulation using real-world data. *Ann. Epidemiol.* 78, 28–34 (2023).
- **Illustrates target trial emulation using real-world data in oncology, relevant to methodological parallels with the TBASEL study.**
11. Deng Y, Polley EC, Wallach JD, Herrin J, Ross JS, McCoy RG. Comparative effectiveness of second line glucose lowering drug treatments using real world data: emulation of a target trial. *BMJ Med.* 2(1), e000419 (2023).
12. Dickerman BA, Garcia-Albeniz X, Logan RW, Denaxas S, Hernan MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat. Med.* 25(10), 1601–1606 (2019).
13. Murray EJ, Hernan MA. Improved adherence adjustment in the coronary drug project. *Trials* 19(1), 158 (2018).
14. Food and Drug Administration. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials (2021). Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical>
15. Hernan MA, Scharfstein D. Cautions as regulators move to end exclusive reliance on intention to treat. *Ann. Intern. Med.* 168(7), 515–516 (2018).
- **Discusses regulatory considerations in moving beyond ITT, underscoring policy relevance for adopting PP estimates in decision-making.**
16. Polito L, Liang Q, Pal N *et al.* Applying the estimand and target trial frameworks to external control analyses using observational data: a case study in the solid tumor setting. *Front. Pharmacol.* 15, 1223858 (2024).

17. Hernan MA, Robins JM. Per-protocol analyses of pragmatic trials. *N. Engl. J. Med.* 377(14), 1391–1398 (2017).
18. Hernan MA, Robins JM. *Causal Inference: What If*. Chapman Hall/CRC Press, FL, USA (2024).
19. Murray EJ, Caniglia EC, Petito LC. Causal survival analysis: a guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence. *Res. Methods Med. Health Sci.* 2(1), 39–49 (2020).
- **Provides a guide on estimating ITT and PP effects from trials with nonadherence, highly relevant for understanding methodological underpinnings.**
20. Rittmeyer A, Barlesi F, Waterkamp D *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a Phase III, open-label, multicentre randomised controlled trial. *Lancet* 389(10066), 255–265 (2017).
21. Birnbaum B, Nussbaum N, Seidl-Rathkopf K *et al.* Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. *arXiv* 2001.09765 (2020). Available from: <https://arxiv.org/abs/2001.09765>
22. Ma X, Li L, Madden S, Blythe JS, Baxi SS. Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR. *medRxiv* 2020.03.16.20037143 (2020). Available from: <https://www.medrxiv.org/content/10.1101/2020.03.16.20037143v3>
23. Zhang Q, Gossai A, Monroe S *et al.* Validation analysis of a composite real-world mortality endpoint for patients with cancer in the United States. *Health Serv. Res.* 56(6), 1281–1287 (2021).
24. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat. Sci.* 25(1), 1–21 (2010).
25. Wilkinson S, Gupta A, Scheuer N *et al.* Assessment of alectinib vs ceritinib in ALK-positive non-small cell lung cancer in Phase II trials and in real-world data. *JAMA Netw. Open* 4(10), e2126306 (2021).
26. National Institute for Health and Care Excellence. NICE real-world evidence framework (2022). Available from: <https://www.nice.org.uk/corporate/ecd9/chapter/overview>
27. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol. Drug Saf.* 15(5), 291–303 (2006).
28. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am. J. Epidemiol.* 177(4), 292–298 (2013).
29. Woodford Guegan E, Cook A. European network for health technology assessment joint action (EUnetHTA JA): a process evaluation performed by questionnaires and documentary analysis. *Health Technol. Assess.* 18(37), 1–296 (2014).
30. European Commission. Regulation on health technology assessment. Available from: [https://health.ec.europa.eu/health-technology-assessment/regulation-health-technology-assessment\\_en](https://health.ec.europa.eu/health-technology-assessment/regulation-health-technology-assessment_en)
31. Crown W, Dahabreh IJ, Li X, Toh S, Bierer B. Can observational analyses of routinely collected data emulate randomized trials? Design and feasibility of the observational patient evidence for regulatory approval science and understanding disease project. *Value Health* 26(2), 176–184 (2023).
32. Wang SV, Schneeweiss S, Franklin JM *et al.* Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 329(16), 1376–1385 (2023).
- **Demonstrates the feasibility and limitations of emulating randomized clinical trials using nonrandomized data analyses, offering benchmark insights for your study's design considerations.**
33. Hansford HJ, Cashin AG, Jones MD *et al.* Reporting of observational studies explicitly aiming to emulate randomized trials: a systematic review. *JAMA Netw. Open* 6(9), e2336023 (2023).