# Implementing Estimands & Target Trial Emulation (TTE) in Real-World Evidence: Case Studies & Perspectives Q&A
# 24 February 2026

**Answered Live (edited)**

**1. To Letizia: What is the difference between using IPCW and marginal structural models for adjusting treatment switching?**

**Letizia:** IPCW (inverse probability of censoring weighting) and marginal structural models (MSMs) are not competing choices, but rather two parts of the same causal toolkit.

We use IPCW to address informative censoring. In our study, we censor patients at the moment they switch treatments. To avoid selection bias, since those who switch treatments differ from those who don't, we assigned more weight to the remaining uncensored patients who share similar characteristics.

An MSM is the broader structural model used to estimate the causal effect across the entire population. When we use IPCW to adjust for that treatment switch, the final weighted Kaplan–Meier curve or the weighted hazard ratio is actually the result of fitting an MSM.

By using IPCW to handle the time-varying nature of treatment switching and its relationship with disease progression, we are implementing an MSM to estimate a hypothetical effect. IPCW is the engine that allows the MSM to provide a valid answer when standard methods fail due to time-dependent confounding.

**2. To Letizia: Does this method estimate a marginal or a conditional probability?**

**Letizia:** In our study, we targeted a marginal probability by applying a double-weighting approach to our data. While the weights themselves – the IPTW-ATT and IPCW – are derived from conditional models that account for baseline and time-varying covariates such as disease progression, the application of these weights creates a pseudo-population where the distribution of confounders is balanced, effectively emulating the "Population" attribute of our target trial. By analysing this pseudo-population, we move from individual conditional probabilities to a population-level summary measure, specifically a marginal hazard ratio. This allows us to estimate the average treatment effect on the treated (ATT), which captures the causal contrast of interest: the difference in overall survival had the population followed a hypothetical strategy where no subsequent therapy was received. Ultimately, our estimator uses conditional modelling as the engine to produce a population-level estimate that accurately reflects that marginal estimand.

**Not Answered Live**

**3. To Jufen: Did you get any comments from the agency regarding the definitions of PFS being different between SAT and RWD?**

**Jufen:** In the target RCT, PFS was defined as the time to first documented disease progression or death from any cause. Initiation of a new anticancer therapy was addressed using a hypothetical strategy: if a patient had not experienced progression or death prior to starting a new therapy, PFS was censored at the time of initiation.

However, in RWD, progression dates were unavailable for many patients. Therefore, the PFS comparison was conducted using a composite strategy, in which initiation of a new anticancer therapy was considered an event for patients in both RWD and ELARA trials. In other words, progression, death, or initiation of a new anticancer therapy were treated as PFS events. Patients without any of these events were censored at the date of last contact.

This represents a limitation of the emulation of the target RCT. However, because the same PFS definition was consistently applied to both the ELARA trial data and the external RWD – where initiation of a new anticancer therapy was considered a PFS event in both cohorts – and because this approach was pre-specified in the SAP prior to analysis, the health authority accepted it without further questions.

**4. To Jufen: Did you weight both baseline characteristics and the censor?**

**Jufen:** In this case study, the weighting-by-odds approach was applied to adjust for potential confounding in baseline characteristics. Because the target estimand was the average treatment effect on the treated (ATT), the ELARA cohort did not require weighting to construct the tisa-cel group – each patient in ELARA was assigned a weight of 1.

For the real-world (RW) cohort, each patient was assigned a weight of $\hat{e}_i/(1 - \hat{e}_i)$, where $\hat{e}_i$ represents the estimated probability of belonging to the ELARA cohort. This approach creates a weighted RW population with a covariate distribution comparable to that of the ELARA cohort.
We assessed baseline balance between the RWD and ELARA patients before and after weighting by examining standardised mean differences (SMDs). SMDs were evaluated for both pre-weighted and post-weighted data. The weighting-by-odds approach was considered adequate in reducing imbalances in measured baseline covariates, as indicated by absolute SMDs below 25%.

Censoring was not separately weighted in this analysis. After achieving covariate balance, causal effects for PFS and OS were estimated by fitting a weighted Cox regression. For details, please refer to: Hampson L.V., Chu J., et al. (2024). Combining the Target Trial and Estimand Frameworks to Define the Causal Estimand: An Application Using Real-World Data to Contextualize a Single-Arm Trial. Statistics in Biopharmaceutical Research, 16(1), 1–10. https://doi.org/10.1080/19466315.2023.2190931.

**5. To Jufen: A question about the last slide of the second presentation - Could you elaborate on the opinions of the HTA about the combination of the frameworks for the ELARA use case? Did they flag any limitations that are different to the EMA or the FDA? Were they interested in different research questions or analysis? Thank you!**

**Jufen:** The formal analysis report of the indirect comparison between the ELARA trial data and external RWD was submitted to the HTA to support the final pricing and reimbursement decision. The HTA interactions were managed by another group of real-world evidence experts, and I wasn't directly involved in those discussions, so I can't provide detailed insight into all aspects of their feedback.

However, to the best of my recollection, the HTA raised a considerable number of follow-up questions, particularly regarding the potential limitations of RWD and underlying analysis. They requested more detailed justification and clarification of these limitations. The HTA asked for additional sensitivity analyses, to further assess the robustness of the findings and to support their evaluation.
ed".

**6. What are some of the challenges in including observational data in clinical trials? How are regulators reacting to the use of RWD for decision-making?**

**Jufen:** We had multiple interactions with the EMA from the beginning of the study. We consulted EMA scientific advice before we conducted the analysis and then we included the analysis report in the EMA submission dossier. The contribution of RWE was recognised. However, unfortunately, the RWE data was not accepted for inclusion in the EU label, but it was included in the EPAR after approval. During this process, the target trial and estimand frameworks played an important role in facilitating transparent and constructed discussion, and helped us address the causal inference questions.

For the FDA, we discussed the RWD for the ELARA study at the pre-submission meeting, and the FDA indicated that the single-arm data alone is sufficient for the decision-making in this setting. We therefore decided not to include it in the submission dossier.

**Gerd**: Including RWD in decision-making has been successful. It requires early agency communication, transparency, suitable data quality (relevance and reliability) and rigid statistical methodology. Corresponding challenges are finding the right data source with sufficient data quality (data feasibility step), where critical covariates which are needed for causal inference to overcome the non-randomised setting are appropriately available.

Regulators react differently, with the FDA being at the forefront by actively issuing guidelines describing how RWD can be applied for FDA regulatory decision-making. If RWD is planned to support a submission, the agencies should be contacted for feedback, because the RCT design remains the gold standard for studies being conducted for regulatory decision-making. However, the FDA has approved at least 60 drugs without an RCT since 1999 (Hatswell A.J. et al. (2016). Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. BMJ Open 6(6):e011666). If the treatment effect is very large (e.g. as seen in a single-arm trial or in an external comparator study), an RCT may be too much of a good thing. One example is a new CAR-T treatment with a response rate of 90%. Chemotherapy has a response rate of 30%. In such a case, an EC study design using RWD for the comparator might be more appropriate. Typical indications where RWD is increasingly used are rare diseases and oncology. Adding RWD to a single-arm trial submission (e.g. by delivering EC study results) typically strengthens the submission.

# Implementing Estimands & Target Trial Emulation (TTE) in Real-World Evidence: Case Studies & Perspectives


## Abstract

Webinar 3 in the PHUSE series, At the Intersection of Estimands and Target Trial Emulation (TTE) for RWE, moves from theory to practice. Through two applied case studies and an expert perspective, presenters will show how the estimand framework clarifies the clinical question (treatment strategies, target population, outcomes, summaries, and handling of intercurrent events), while TTE provides the operational blueprint for eligibility, time-zero, exposure rules, follow-up and analysis.
We will walk through external comparator analyses using real-world data and address common challenges such as data completeness, index date alignment, baseline confounding and treatment switching. A closing commentary synthesises considerations when jointly applying estimands and TTE in external comparator studies.


## Speakers

**Gerd Rippin, *IQVIA***
Dr Gerd Rippin, BSc, MSc, PhD, is a Senior Director Statistical Services Real-World Solutions at IQVIA. He has worked in the industry since 2000 and at IQVIA since 2017. Gerd specialises in real-world methodology such as causal inference, time-to-event analysis and external comparator studies, and has published papers with EMA co-authors.

**Jufen Chu, *Daiichi Sankyo***
Jufen Chu, PhD, is a director in the Biostatistics Department at Daiichi Sankyo. Prior to joining Daiichi in June 2024, Jufen served as a lead statistician in the haematology CAR-T programme at Novartis. From 2016 to 2024, she contributed to multiple CAR-T programmes focused on non-Hodgkin's lymphoma and multiple myeloma at various stages of development. Jufen gained her PhD in Statistics from the University of Texas at Dallas in 2016.

**Letizia Polito, *Roche***
Letizia Polito, PhD, is a Principal Real-World Data Scientist at Roche, with nine years of pharmaceutical industry experience, including at GSK. She leads real-world evidence generation in haemophilia after extensive work in lung cancer. In her day-to-day role, Letizia focuses on tackling methodological hurdles as they emerge, applying advanced biostatistics and causal inference to design fit-for-purpose solutions.

# Implementing Estimands & Target Trial Emulation (TTE) in Real-World Evidence: Case Studies & Perspectives

## Moderators/Organisers

**Thanh G.N. Ton, *Roche***

Thanh G.N. Ton, MPH, PhD, is a Senior Director at Roche in the Real-World & Clinical Data Strategy group within the Product Development Data Sciences function. She is an epidemiologist, with over 25 years of experience designing and applying observational methods in disease areas including oncology, neurodegenerative conditions, cardiovascular disease and immunologic conditions. Thanh is passionate about achieving scientific rigour and committed to incorporating innovative methodologies into her research.

**Aaron Crowley, *Genesis Research Group***

Aaron Crowley, MA, is Director of Biostatistics at Genesis Research Group – an HEOR and RWE consultancy supporting pharmaceutical and biotech sponsors. He brings more than a decade of experience designing and analysing real-world studies and clinical trials, with methodological expertise in causal inference, externally controlled trials, and indirect treatment comparisons. Aaron's work routinely supports regulatory submissions, HTA dossiers, and market access strategies, spanning oncology, cardiovascular, immunology and rare diseases.

**Matt Baldwin, *Amgen***

Matt Baldwin leads the PHUSE Real World Evidence Working Group project on Estimands for RWD/RWE. He is a Biomedical Data Stewardship Senior Manager at Amgen, where he leads data standards efforts for ADaM datasets. Four years ago, after 10 years as a biostatistician, Matt made the switch to data standards – and he couldn't be happier. He has a passion for improving clinical data analysis processes with statisticians and programmers. His current interests include the implementation of estimands and aspects around real-world data. Matt started volunteering with DIA in 2015, but now exclusively participates in volunteer teams with CDISC and PHUSE. His PHUSE Working Group involvement mainly falls under Optimizing the Use of Data Standards and Real World Evidence.