Target Trial Emulation: A Framework for Causal RWE

By InuitionLabs.ai • 10/17/2025 • 40 min read

target trial emulation | real-world evidence | rwe | causal inference | observational studies | real-world data | epidemiology | regulatory affairs

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Executive Summary

This report examines how *real-world evidence* (RWE) can be used to support causal conclusions in healthcare decision-making, by adopting the **target trial emulation** framework to avoid specialized jargon. Real-world data (RWD) – such as electronic health records, insurance claims, and patient registries – provide vast information about patients' treatments and outcomes outside clinical trials (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). Regulatory agencies like the U.S. Food and Drug Administration (FDA) and Europe's EMA have increasingly recognized RWE as valuable for evaluating medical products; indeed, the U.S. 21st Century Cures Act (2016) mandated guidance on RWE, defining it as "clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD" (pmc.ncbi.nlm.nih.gov). In practice, RWE has already been used for drug approvals, label changes, and safety assessments (www.fda.gov). However, using RWD for causal inference is challenging due to potential biases and confounding factors that are absent by design in randomized trials.

The **target trial emulation** approach offers a structured way to draw causal insights from observational data by explicitly designing and conducting an observational study as if it were a randomized trial (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). The key idea is to specify the protocol of the ideal randomized trial – including eligibility criteria, treatment strategies, start of follow-up, and outcomes – and then apply those elements to real-world data (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). In this way, analysts intentionally "mimic" random assignment by carefully controlling who enters each group and adjusting for differences between them. By adhering closely to the trial blueprint, target trial emulation minimizes common errors in observational research (such as immortal-time bias and confounding) and makes the analysis transparent and reproducible (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).

This report provides a **comprehensive overview** of RWE and target trial emulation for decision-makers. We first define RWD and RWE, discuss their growing role in regulation and practice, and outline the challenges to making causal claims from observational data. We then explain the target trial emulation framework step by step—describing how to translate each component of a randomized trial into the observational context (see **Table 1**). We highlight analytical techniques (e.g. propensity scoring, inverse probability weighting, and sensitivity analyses) that help control biases. Throughout, we cite peer-reviewed studies and industry examples. For instance, a J. Natl. Cancer Institute study emulated a breast cancer trial and produced hazard ratios nearly identical to the actual randomized trial but with tighter confidence intervals due to larger sample size (pmc.ncbi.nlm.nih.gov). Another example is a Danish registry study that emulated a head-to-head trial of two diabetes drugs, finding no significant difference in cardiovascular outcomes (www.ahajournals.org). On the other hand, systematic reviews have found that even well-done observational comparisons may contradict RCT findings in about 18% of cases, underlining why rigorous design like target trial emulation is essential (pmc.ncbi.nlm.nih.gov).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

We also present **case studies and comparisons**: e.g. how statin use and dementia risk was studied with a target trial – finding modest long-term benefit from sustained statin use (pmc.ncbi.nlm.nih.gov) – as well as broader meta-analyses comparing many RCTs and observational results (pmc.ncbi.nlm.nih.gov). Regulatory contexts are discussed: the FDA and EMA have issued guidance on RWD use, and bodies like NICE in the UK explicitly endorse trial-like methodologies in their RWE framework (pmc.ncbi.nlm.nih.gov) (www.evidencebaseonline.com). At the same time, industry uptake has been slower than expected (pmc.ncbi.nlm.nih.gov) (www.evidencebaseonline.com).

Finally, we consider implications and future directions. Target trial emulation helps decision-makers gain *causal insights* from existing data while avoiding misleading deductions. It demands careful planning, high-quality data, and transparency about assumptions. Emerging practices (like the "Causal Roadmap" framework (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov)) formalize these steps. Advances in data science (common data models, machine learning) and regulatory standards are making RWE more reliable. However, challenges remain: unmeasured confounding cannot be fully ruled out, and many RWD sources still suffer from incompleteness or biases (www.mdpi.com) (pmc.ncbi.nlm.nih.gov). In sum, RWE – when properly analyzed through target trial emulation – holds great promise for informed decision-making, both now and in the future.

Introduction and Background

Real-world data (RWD) refer to information collected outside the setting of classical controlled trials. Sources include electronic health records (EHRs), health insurance claims, patient registries, and even data from digital devices. Such data capture "what happens in routine practice" – how patients are treated and what outcomes they experience in the actual healthcare system. When properly analyzed, RWD can generate **real-world evidence (RWE)** about the effects and safety of medical interventions. In recent years, regulatory agencies and healthcare stakeholders have embraced RWD/RWE as increasingly important inputs into decision-making. For instance, the U.S. 21st Century Cures Act (2016) explicitly prompts the FDA to provide guidance on using RWE. In fact, the Act defines RWE as *"clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of real-world data"* (pmc.ncbi.nlm.nih.gov). Globally, many agencies (EMA, FDA, as well as payer HTA bodies like NICE in the UK) are developing frameworks to integrate RWE into evidence review and policy decisions (pmc.ncbi.nlm.nih.gov) (www.mdpi.com).

Examples demonstrate the scope of RWE's use. Since 2011, the FDA's Center for Drug Evaluation and Research (CDER) and Center for Biologics (CBER) have documented dozens of regulatory decisions that involved RWE. </current_article_content>These include new drug approvals supported in part by retrospective cohort studies, label expansions relying on observational analyses, and safety assessments from post-market data (www.fda.gov). Table 1 of the FDA's report lists cases like using medical record cohorts to confirm treatment in frostbite, or leveraging national death records to measure 28-day mortality in a clinical trial (www.fda.gov).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Likewise, an EMA review notes that RWE can complement clinical trial results in both pre- and post-approval contexts (www.ema.europa.eu). In sum, RWE is moving from peripheral to core importance in medical product evaluation (www.linkedin.com).

However, while RWD are abundant and diverse, they pose **significant challenges for causal inference** (determining cause-effect relationships). Unlike randomized controlled trials (RCTs), real-world data are observational by nature: treatments are not randomly assigned, and data capture is not standardized. Observational analyses can suffer from biases and confounding. For example, if healthier patients preferentially receive a new drug while sicker patients do not, a simple comparison of outcomes will be misleading – any difference might reflect baseline health rather than treatment benefit. There are many such pitfalls (see Section 2).

The concept of causality itself requires care. In an RCT, the random assignment ensures that, on average, treatment groups are similar except for the treatment, which justifies attributing differences in outcomes to the treatment. In observational data, groups typically differ in systematic ways that can confound results. As Hernán and Robins have emphasized, drawing causal inference requires imagining the *"target trial"* you wish you had, and then emulating it carefully (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). This approach has become mainstream in epidemiology and outcomes research. Indeed, the idea of emulating randomized experiments using observational data dates back decades (to Cox 1958 and Rubin 1974) (pmc.ncbi.nlm.nih.gov), and was formalized explicitly in recent years by Hernán & Robins (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).

In practice, rigorous design and analysis are needed to yield reliable RWE. A key insight from recent reviews is that *study design is as important as data quality* (www.mdpi.com). Some analyses of observational data have been dismissed because design flaws, not data insufficiency, drove the bias. In one perspective, Zou and Berger note that many discrepancies between RCT results and RWD analyses occurred precisely because those studies *did not embed a target trial design* (www.mdpi.com). With proper design – explicitly mirroring a hypothetical RCT – analysts can reduce these discrepancies.

This report aims to explain these concepts in accessible terms, focusing on the target trial emulation framework for decision-makers. We will cover:

- **The nature of RWD/RWE:** definitions, sources, regulatory context, and data quality issues.
- **Challenges of observational data:** confounding, biases, missing data, and their effects on causal claims.
- **Target Trial Emulation:** rationale, core principles, and step-by-step procedure. We will use simple language – for example, describing how "making groups comparable" serves the role of randomization.
- **Implementation and methods:** how to actually adjust for differences (propensity scores, weighting, matching) and handle time-related issues (like immortal time bias).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

- **Evidence and case studies:** concrete examples where target trial emulation was applied, including comparisons to RCTs and lessons learned.
- **Regulatory and policy aspects:** how agencies view trial emulation; successes and limitations in actual submissions.
- **Future trends:** upcoming methodological advances, data quality frameworks, integration with AI, and open challenges.

Throughout, we emphasize an **evidence-based approach**: all claims will be supported by citations to peer-reviewed studies, regulatory documents, or authoritative reviews. The goal is to provide decision-makers with a thorough but understandable picture of RWE causality and target trial methods, so they can appreciate both the opportunities and the caveats in using such evidence for healthcare and policy decisions.

Real-World Data (RWD) and Evidence (RWE)

Sources and Definitions

Real-World Data (RWD) are data pertaining to patient health or healthcare delivery that are collected outside of traditional clinical trials. Common sources include:

- **Electronic Health Records (EHRs):** Digital medical records from hospitals or clinics, capturing diagnoses, treatments, lab results, etc.
- **Claims and billing databases:** Records of all billed procedures, prescriptions, and diagnoses from insurers.
- **Patient Registries:** Organized collections of detailed data for patients with particular conditions (e.g. cancer registries, rare disease registries).
- **Wearables and apps:** Data from devices like fitness trackers, smartphone apps, or patient surveys collecting daily life and physiologic data.

These sources have complementary strengths. For example, claims data cover large populations longitudinally and accurately record dates of healthcare events, but often lack detailed clinical measures. EHRs contain rich clinical details (vitals, imaging, labs) but may have missingness and inconsistent coding. Registries often have high-quality fields for specific diseases but may lack a general population comparator.

When RWD are analyzed to answer a specific question, the result is *Real-World Evidence (RWE)*. For instance, one could use RWD to estimate the comparative effectiveness of two diabetes drugs on heart attack rates. If done correctly, the analysis yields RWE on the drugs' real-world

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

benefits and risks. It is important to recognize that **RWE is only as reliable as the data and methods used**. Poorly collected data or naïve comparisons can lead to spurious conclusions. This has led to emphasis on "*fit-for-purpose*" data – i.e., RWD that are sufficiently complete and accurate to answer the question at hand (www.mdpi.com) (www.mdpi.com).

Regulators and professional bodies have also codified what aspects of RWD matter. Recent data quality frameworks (DQFs) outline dimensions like data completeness, consistency, and representativeness. For example, the EMA's Data Quality Framework stresses transparency (knowing how data were collected), reliability (accuracy of coding), and coherence (ability to link data over time) (www.mdpi.com). Ultimately, effective RWE requires both high-quality data *and* rigorous study design (www.mdpi.com).

Growing Role in Decision-Making

The use of RWE has grown substantially. In extensions of drug labels, in comparative effectiveness research, and in safety monitoring, RWD have become central. During the COVID-19 pandemic, for instance, RWD from health records and registries was used to assess vaccine effectiveness in near real-time (pmc.ncbi.nlm.nih.gov). Observational analyses provided rapid evidence on booster need, on vaccine protection in special populations, and on comparative performance of different vaccines (pmc.ncbi.nlm.nih.gov).

In terms of policy, both FDA and EMA (and other agencies) now actively seek RWD inputs. The FDA's RWE Program and EMA's Big Data initiatives illustrate that. Surveys show that regulatory and payer decisions increasingly cite RWE (pmc.ncbi.nlm.nih.gov) (www.mdpi.com). For example, the FDA lists numerous approvals (drug/device) from 2011-2024 where RWE contributed. These include using registry data as external controls in single-arm trials, or leveraging EHR data to rule out safety signals (www.fda.gov).

Despite this acceptance, stakeholders remain cautious. A 2024 landscape analysis noted that nearly all (95%) of reviewed RWE studies had at least one avoidable methodological issue (pmc.ncbi.nlm.nih.gov). This grim statistic underscores that lots of published RWE is flawed by design or analysis, not by data per se. Therefore, current efforts emphasize *"doing it right"*: adopting standardized methods (like target trial emulation) and rigorous reporting. The U.S. FDA and the UK's NICE have issued methods documents focusing on comparative effect studies (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). In fact, NICE's Real-World Evidence framework (2022) explicitly lists target trial emulation as a key tool for high-quality analyses (pmc.ncbi.nlm.nih.gov), although surveys show industry uptake of TTE methods is still catching up (pmc.ncbi.nlm.nih.gov) (www.evidencebaseonline.com).

Data Quality Considerations

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Before delving into methodology, it is useful to acknowledge practical issues with RWD. RWD are collected for clinical or administrative purposes, not research, so they may have missing entries, miscoded values, or variable follow-up. For example, a patient might have multiple EHR entries with slightly different diagnoses codes, or a claims database might misclassify the reason for hospital admission. There can also be systematic omissions: some mild cases may never be recorded, so disease incidence can be underestimated. Data may come from multiple sites with different coding systems.

Regulatory guidances and academic reviews stress evaluating RWD quality upfront (www.mdpi.com). This can involve technical checks (are dates plausible? Are key values missing often?) and content review (does the data capture the clinical detail needed?). The MDPI perspective [20] outlines specific criteria for screening RWD sources. Some frameworks (like the EU's EHD Working Party) enumerate dimensions such as accuracy, completeness, timeliness, and linkability (www.mdpi.com).

While data quality is indispensable, experts note it is only one pillar; *study design* is equally crucial (www.mdpi.com). Even with perfect data, a poor analysis plan can give wrong answers. On the other hand, rigorous design (following the structure of a plausible trial) can salvage imperfect data. As the authors of [48] (Zou & Berger) summarize: *"Beyond data quality, it has become apparent that study design is equally, if not more, important to the creation of credible RWE."* They found that many RCT-vs-RWE discrepancies were due to flawed observational designs rather than just data issues (www.mdpi.com).

In practice, high-quality RWE generation thus entails (a) using RWD that meet purpose-specific criteria for completeness and validity, and (b) following methodological best practices (such as target trial protocols) to control biases. The rest of this report will focus on item (b), which is the emphasis of the target trial emulation approach.

Causality and Observational Data

The Causal Inference Challenge

In research, *causality* means understanding what would happen under different interventions or exposures. Rather than just noting that "treatment X is associated with better outcomes," decision-makers want to know if X actually **causes** the change – would giving X to a patient change their outcome? In randomized trials, this question is answered by comparing randomly allocated groups. Randomization ensures that, apart from the treatment itself, the groups are statistically similar at baseline. So any difference in outcomes can (on average) be attributed to the treatment. In formal terms, randomization guarantees *exchangeability*: at time of assignment, the treatment groups have comparable distributions of all factors (known and unknown) that affect outcomes (pmc.ncbi.nlm.nih.gov).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

In observational studies, however, treatments are chosen, not assigned. Thus groups may differ: for instance, sicker patients might be more likely to receive a new aggressive therapy, or doctors may prescribe differently based on age, co-morbidities, etc. This *confounding* by indication can lead to spurious associations. A classic example (often cited) is the apparent protective effect of hormone replacement against heart disease, an association later disproven when RCTs accounted for healthier characteristics of women who chose HRT. In RWD, similar issues abound: e.g., if healthier patients systematically receive Drug A, naive comparisons would make Drug A look more effective than it really is.

Other biases in RWD can further muddy causal claims. **Selection bias** can occur if certain patients are excluded or lost in a non-random way. **Immortal time bias** arises when patients must survive a certain period before being counted in a treatment group, artificially inflating survival in that group. **Measurement bias** occurs if outcomes are recorded differently across groups (e.g. follow-up is more complete for one group). **Time-dependent confounding** happens when intermediate health changes affect both treatment decisions and outcomes. Each of these can flip results or create false effects.

A helpful mental picture is the “causal gap” – what we want to compare (the counterfactual outcome difference) versus what we actually measure. If patient A took Drug X and survived, the question is: would A have survived had they taken Drug Y instead? Counterfactuals (the world where the same patient experiences the alternative) are invisible, so we compare different patients (or the same patient over time) under different treatments. Ensuring those patients are truly comparable is the core challenge.

Many statistical methods exist to adjust for confounding (e.g. regression, matching, stratification). However, any method relies on correct model specification and on having measured all confounders (“no unmeasured confounders” assumption). In RWD settings, often not all relevant factors are captured. Thus, analysts must be careful: transparency about assumptions is key.

Target Trial Emulation Concept

The target trial emulation framework provides a way to *structure* observational analyses so they parallel a randomized trial as closely as possible (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). The steps usually include:

1. **Specify the target trial protocol.** Imagine the ideal randomized trial that would directly answer your causal question. Explicitly write down all components: who would be eligible (inclusion/exclusion criteria), what treatments or exposures would be compared, when follow-up would start, what outcomes would be measured, and for how long. Also specify how patients would be assigned (randomly) and what causal effect you want (e.g. intention-to-treat risk difference, per-protocol hazard ratio, etc).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

2. **Emulate each element using observational data.** Use your RWD to *mimic* each component. For example, define eligible patients in the data who meet the trial criteria at some baseline time `t0`. Define treatment groups based on what therapy these patients actually received at `t0`. Define follow-up starting at `t0` (the "index date") and ending at outcome or censoring. Crucially, make sure the observational definitions mirror the trial's design as closely as possible (see **Table 1**).

3. **Adjust for confounding and bias.** Since we cannot randomize in RWD, we mimic randomization statistically. This typically involves adjusting for all measured baseline differences between groups. Propensity score methods (matching or weighting) or regression adjustment can balance observed covariates so that, conditional on those covariates, the treated and comparison groups are as similar as possible (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). In advanced forms, one may also address time-varying confounders. We explain these below.

4. **Analyze outcomes just as in a trial.** Using the emulated cohort, estimate the effect measure specified (e.g. risk ratio, hazard ratio). Interpret this as the causal effect your hypothetical trial would have measured. Conduct sensitivity analyses (e.g. test how robust results are to potential unmeasured confounding).

The key is that by **pre-specifying the trial structure**, analysts hard-wire transparency and guard against "researcher degrees of freedom". Every choice (inclusion rules, covariates for adjustment, analysis model) should be justified by the trial protocol, not by hunting data patterns. As Arnold et al. note, TTE improves *"clarity, transparency, and reliability"* of results (pmc.ncbi.nlm.nih.gov).

Table 1 below summarizes the major components of a target trial and how they map to an emulated observational study. Adhering to this table helps ensure that important design features (like start time and comparison group) are not overlooked.

Trial Component	Target Trial (Ideal RCT)	Emulated Observational Study
Eligibility criteria	Predefined inclusion/exclusion rules (e.g. age, diagnosis).	Apply the same criteria to select eligible subjects from RWD (index cohort).
Treatment strategies	Two (or more) arms randomly assigned (e.g. Drug A vs Drug B).	Define patient groups based on their actual initial treatment in data (new-user design).
Assignment (randomization)	Patients randomly allocated at baseline, ensuring balanced groups.	No randomization; instead, adjust for measured baseline differences via matching/weighting.
Time zero (start of follow-up)	The time of randomization or treatment initiation in trial.	Choose an index date (e.g. first prescription date) analogous to randomization time. Exclude immortal waiting periods.
Follow-up period	Fixed or specified follow-up time after randomization.	Follow each patient from index date until outcome, censoring, or end of study, mirroring trial length.
Outcomes	Standardized definitions and scheduled assessment of outcomes.	Define outcomes with the same clinical criteria (e.g. diagnosis codes) and timeframe in RWD.
Analysis approach (estimand)	e.g. Intent-to-treat effect or per-protocol effect in trial.	Estimate the same effect measure using the emulated cohorts (often intention-to-treat by including everyone from index). Include sensitivity analyses for adherence.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Table 1. Components of a Target Trial and their Observational Emulation. This schematic shows how each key trial element should be defined in the real-world data analysis. For example, "time zero" in the target trial becomes the patient's treatment initiation date in the data – this avoids immortal-time bias (patients must be alive to enter a group) by aligning follow-up properly (pmc.ncbi.nlm.nih.gov).

The target trial analogy also makes clear why certain pitfalls occur in naive analyses. For instance, **immortal time bias** happens if one arm's patients have to survive longer before entering the analysis cohort (e.g. if treatment starts after a waiting period, those who died early are incorrectly left out of the treatment group). By specifying time zero as treatment start, and excluding patients who experienced the outcome or were lost before that time, we avoid giving the treatment arm an unfair "immortal" period (pmc.ncbi.nlm.nih.gov).

Similarly, "confounding by indication" is addressed by ensuring that eligibility and assignment definitions capture all the clinical factors determining treatment. In an RCT, randomization automatically balances *even unknown* factors. In emulation, we instead **measure and adjust** for the known patient characteristics that influenced treatment choice (pmc.ncbi.nlm.nih.gov). While we cannot guarantee we have measured everything (a limitation discussed later), careful covariate selection can approximate the exchangeability that randomization provides.

Analytical Methods

Once the target trial is emulated in data, the main tasks are to adjust for differences between groups and to estimate the treatment effect. Common techniques include:

- **Propensity score matching or weighting:** Estimate each patient's probability of receiving the treatment given their baseline characteristics (the *propensity score*). Then, match treated patients to similar untreated patients, or weight patients so that the two groups have similar covariate distributions. This mimics a randomized comparison on measured factors. Numerous published examples use stabilized inverse probability weights to achieve balance (pmc.ncbi.nlm.nih.gov) (www.ahajournals.org).
- **Regression adjustment:** Include the measured covariates in an outcome regression model (e.g. Cox model for survival) to control for confounding. This is straightforward but relies on correct model specification.
- **Stratification or subclassification:** Divide data into strata of similar propensity score or covariate patterns, and compare outcomes within each stratum.
- **Advanced methods (g-formula, marginal structural models, g-estimation):** These handle time-varying treatments or confounders – e.g., if a patient's health status changes over time and affects both future treatment and outcome. Such methods are beyond basic scope, but they are part of the epidemiologist's toolkit.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Regardless of method, diagnostics are crucial. One should check that after adjustment, the baseline characteristics of the emulated groups are indeed similar ("balance check"). If not, residual confounding will bias results. Sensitivity analyses are often performed to gauge how much unmeasured confounding would be needed to change findings significantly. As one study explains, a **quantitative bias analysis** can estimate the impact of hypothetical unmeasured confounders (pmc.ncbi.nlm.nih.gov).

Table 2 below lists common biases in RWD and how target trial emulation addresses them:

Bias / Issue	How It Distorts Causal Claims	Mitigation in Target Trial Emulation
Confounding	Treatment groups differ in baseline risk (observed/unobserved factors).	Adjust for all measured patient characteristics (age, disease severity, etc.) via matching/weighting (achieve exchangeability) (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). Pre-specify these variables based on the protocol.
Immortal-time bias	Patients must survive an "immortal" period to receive treatment, making drug look better.	Align start of follow-up with time of eligibility/treatment (index date), excluding any pre-treatment immortal time (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).
Selection bias	Excluding or censoring patients non-randomly (e.g. EHR dropout).	Apply similar inclusion criteria to all arms; use censoring models/weights for loss to follow-up; consider *intention-to-treat* approach.
Measurement misclassification	Errors in recording exposures or outcomes (e.g. miscoding).	Use validated definitions (code lists) for diseases and treatments; consider requiring multiple records for confirmation; sensitivity analysis assuming plausible error rates.
Time-dependent confounding	A time-varying factor affected by prior treatment influences future treatment or outcome.	Use advanced causal methods (e.g. marginal structural models) when needed, so adjustment doesn't block causal pathway; align measurements at baseline where possible.
Channeling bias	Sicker patients channelled to one treatment due to prescribing habits.	Carefully characterize treatment indication in emulation protocol; adjust for indicators of disease severity.
Other biases (lead time, detection)	Differences in diagnostic intensity can skew observed outcomes.	Emulate trial's standardized follow-up and endpoint definitions; e.g. start follow-up only after diagnosis to avoid lead-time bias.

Table 2. Common biases in observational health studies and how target trial emulation mitigates them. For example, "confounding" is addressed by **adjusting** for baseline variables that predict both treatment and outcome (pmc.ncbi.nlm.nih.gov). "Immortal-time bias" is avoided by ensuring that no subject is counted until the treatment of interest could plausibly be received (pmc.ncbi.nlm.nih.gov). The key is that each bias is prevented by carefully specifying the cohort and analysis plan to mirror the hypothetical trial.

In practice, researchers create a **protocol** much like a trial registry entry. This protocol lists all inclusion criteria, which covariates will be adjusted for, the follow-up window, handling of missing data, and primary analysis methods (e.g. Cox regression with inverse probability weights). Following a pre-specified protocol reduces the risk of "data mining" or choosing methods based on the results. As the Causal Roadmap authors emphasize, prespecifying design and analysis plans and assessing assumptions transparently are crucial for credible RWE (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Case Studies and Examples

To illustrate these principles, we present several examples from the literature where target trial emulation has been applied. These demonstrate successes, limitations, and how real-world analyses compare to randomized trials.

Example: Metastatic Breast Cancer (MBC)

A 2023 study in the *Journal of the National Cancer Institute* emulated a randomized trial of first-line therapies for metastatic breast cancer (pmc.ncbi.nlm.nih.gov). The hypothetical trial (E2100) compared paclitaxel alone versus paclitaxel plus bevacizumab in HER2-negative MBC. The analysis used the French ESME-MBC real-world cohort (N=5538 patients).

Emulation steps: The investigators specified the target trial protocol (eligibility, treatments, outcomes) based on the E2100 trial. They excluded any patients who did not meet trial criteria at baseline. "Time zero" was defined as the date of first-line therapy initiation, aligning patient timelines. To adjust for confounding, they employed stabilized inverse-probability weighting and G-computation, using a rich list of baseline covariates (pmc.ncbi.nlm.nih.gov).

Results: After emulation, 3211 patients were eligible (larger than the original trial due to real-world size). The estimated overall survival in the emulated trial favored the combination therapy (paclitaxel+bevacizumab) over paclitaxel alone. The **hazard ratio** for death was **0.88**, similar to the actual E2100 randomized trial (HR=0.88), though with a non-significant *p*-value (P=0.16) (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). Because the real-world cohort was large, confidence intervals were narrower than in the RCT, implying more precise estimates. The study also performed bias analyses, concluding that residual unmeasured confounding was unlikely to overturn the finding (pmc.ncbi.nlm.nih.gov).

Interpretation: This case shows strong agreement between RWD emulation and the RCT result, suggesting the validity of both approaches. The observational emulation reproduced both a similar hazard ratio and the conclusion that overall survival advantage was modest. The targeted design (including removing immortal time and matching baseline) was credited for this concordance. Importantly, the real-world data allowed analysis of more patients, improving the precision: a textbook example of where RWD may complement trials (larger *N*), while emulation prevents biases.

Example: Statins and Dementia Risk

Hernán and colleagues (Caniglia et al., Neurology 2020) conducted a target trial emulation on statin initiation and 10-year dementia risk (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Protocol: The target trial was: among adults without dementia, compare those who start statins (at any point) versus those who do not, and follow for 10 years for new dementia. The investigators used a large US claims and EHR database. Baseline covariates (age, vascular risk factors, health behaviors) were measured at "time zero" (the statin start date for initiators; some index date for non-initiators). Propensity score weighting balanced groups on these factors.

Findings: They estimated both "intention-to-treat" (statin initiation vs no initiation) and "per-protocol" (continuous statin use) effects. The results showed that **sustained statin use** (not just any initiation) was associated with a modestly lower 10-year dementia risk (risk difference –2.2% for dementia or death) (pmc.ncbi.nlm.nih.gov). In contrast, simply initiating statins had little effect. The authors cautioned interpretation due to small numbers of initiators and possible unmeasured confounding, but suggested a potential benefit of long-term statin adherence on cognitive outcomes.

Significance: This emulation addressed a question difficult to answer in RCTs (would require decades of follow-up and large numbers). It shows how intention-to-treat vs per-protocol concepts translate to RWD. It also illustrates the need for caution: bias analysis revealed the results could be sensitive to unmeasured factors (pmc.ncbi.nlm.nih.gov). For decision-makers, the key takeaway is that well-designed RWD studies can suggest potential causal effects (here, statins possibly reducing dementia risk), but confidence depends on data quality and analyses.

Example: Comparative Effectiveness of Diabetes Drugs (Empagliflozin vs Dapagliflozin)

In *Circulation* 2024, a Danish population-based study emulated a head-to-head trial of two diabetes drugs called SGLT2 inhibitors (www.ahajournals.org) (www.ahajournals.org). The question: do patients starting empagliflozin versus dapagliflozin have different 6-year cardiovascular outcomes? (No direct RCT compares them.)

Design: The target trial was: among patients with treated type 2 diabetes, randomize to empagliflozin or dapagliflozin and follow for 6 years for a composite cardiovascular endpoint. The authors used nationwide health registries (linking prescriptions, hospital data, death records). They identified "new users" of either drug from 2014-2020, applied inclusion rules (age, diabetes diagnosis, no prior evidence of drug use), and set time zero at drug initiation. Baseline factors (57 covariates including age, sex, diabetes duration, prior events, kidney function) were used in inverse probability weighting to balance the arms (www.ahajournals.org).

Results: The 6-year cumulative risk of the composite outcome was nearly identical for both drugs. The **adjusted risk ratio** was about 1.02 (no significant difference) in the main analysis (www.ahajournals.org). This finding held in subgroups (e.g. with or without prior cardiovascular disease) and persisted in per-protocol analyses.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Interpretation: According to this emulation, empagliflozin and dapagliflozin are equivalently effective for cardiovascular outcomes. Since no large trial directly compared them, this RWD analysis provides timely evidence for clinicians deciding between these drugs. It also underscores that target trial emulation can be applied to *comparative effectiveness* questions, not just treatment-vs-no-treatment.

RCT vs Observation Concordance

Beyond single studies, systematic reviews have assessed how often observational (properly done) results agree with RCTs. Hong et al. (2021) reviewed 29 systematic reviews that compared drug effect estimates from observational studies and RCTs (pmc.ncbi.nlm.nih.gov). They found that in about 80% of comparisons there was no significant difference in estimated effect sizes between RCTs and observational analyses (pmc.ncbi.nlm.nih.gov). However, in roughly 18% of comparisons the results differed significantly, with some even pointing in opposite directions.

This mixed picture means: often, a well-conducted observational study (especially one mirroring an RCT) gives the same answer as the trial. But substantial discrepancies do occur. Disagreement can stem from differences in patient populations, outcome definitions, or unresolved biases in the observational data. For decision-makers, this implies that target trial emulation can often reproduce RCT findings (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov), but each case must be evaluated on its design merits. Large agreement percentages (80%) are encouraging, but the non-trivial discordance rate emphasizes why transparent methods and sensitivity checks are essential.

Broader Perspectives and Context

Regulatory and Policy Viewpoints

Regulators and health technology assessment (HTA) bodies are now grappling with how to use RWE. The FDA and EMA both have issued guidance documents encouraging the generation of fit-for-purpose RWE and acknowledging the role of methods like target trial emulation. For example, the FDA's recent reports highlight RWE's use in approvals and safety assessments (www.fda.gov). The FDA defines the causal contrasts to estimate (the "estimands") clearly, echoing target trial language, and urges demonstration of how biases are handled.

In Europe, the EMA's RWE initiative similarly aims to integrate observational data in regulatory decisions. Some country-level HTA agencies (e.g. NICE in England, IQWiG in Germany, HAS in France) have also published frameworks. Notably, NICE's 2022 RWE framework explicitly endorses emulating randomized trials when using observational studies (pmc.ncbi.nlm.nih.gov). That means if a company submits an RWE analysis to NICE, using a target trial protocol aligns

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

with NICE's methods expectations. As one analysis noted, "NICE's real-world evidence framework" considers trial emulation central (pmc.ncbi.nlm.nih.gov).

Nevertheless, actual adoption has lagged. A recent editorial by Castanon et al. observed that despite this endorsement, no submissions to NICE explicitly used a formal target trial emulation approach (pmc.ncbi.nlm.nih.gov). They suggest reasons might include lack of expertise or uncertainty in industry, as well as possibly insufficient understanding that such methods are supported. Similarly, industry blogs note that HTA agencies increasingly *appreciate* causal methods in RWD (to guide reimbursement decisions), but that many companies have not yet integrated TTE into their standard toolkit (www.evidencebaseonline.com). This represents a training and communication gap.

Importantly, regulators emphasize **transparency**. They encourage pre-registering RWE study protocols, akin to clinical trial registration, and publishing analysis plans in advance. Target trial emulation naturally leads to such prespecification: by writing a protocol that mirrors a trial, analysts commit up front to specific design and adjustments. This stance counters the old practice of purely data-driven analyses. As one source explains, target trial emulation *"explicitly ties the analysis to the trial it is emulating"*, making biases more visible (pmc.ncbi.nlm.nih.gov). Decision-makers favor this clarity.

Perspectives from Methodologists

Epidemiologists and biostatisticians have developed this field vigorously in recent years. Hernán and Robins wrote seminal papers in 2016 on using big observational data to emulate trials (pmc.ncbi.nlm.nih.gov). The broader causal inference community has endorsed the target trial as *"a powerful tool"* to improve observational research (academic.oup.com). Training programs and textbooks are now including this approach, aiming to elevate the standard of practice. Journals are increasingly publishing RWE studies that explicitly use trial emulation, and calling for detailed reporting of each trial element.

While enthusiasm is high, experts also caution that target trial emulation is not a magic bullet. It **reduces** bias but does not eliminate it if critical assumptions fail. In particular, unmeasured confounding remains the Achilles' heel. If an important predictor of treatment/outcome (say, a lab value or lifestyle factor) is not recorded, adjustment cannot account for it. Authors of the Roadmap framework stress that every RWD analysis should declare and evaluate the plausibility of its key assumptions (pmc.ncbi.nlm.nih.gov). Indeed, nearly all published RWE studies have had at least one identified methodological flaw (pmc.ncbi.nlm.nih.gov). But by highlighting assumptions and conducting sensitivity analyses, target trial frameworks enable at least a judgment of result reliability.

Another methodological issue is *generalizability*. RCTs typically have strict eligibility, which can make their results less applicable to the broader patient population. One advantage of RWD is that it reflects real clinical populations. By choosing broader inclusion, TTE studies can yield

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

evidence for groups under-represented in trials. On the flip side, care must be taken: if the emulated trial's eligibility is too broad, heterogeneity may complicate the interpretation of the causal contrast (are we averaging many different subpopulations?). In some cases, hierarchical or subgroup analyses may be needed.

Industry and Other Stakeholder Views

Beyond regulators and methodologists, industry executives, clinicians, and patient advocates care about RWE too. Payers often view RWE as a way to address uncertainty (especially in post-launch studies) or to evaluate cost-effectiveness. Patient groups may push for RWE when trial participation is infeasible. All these stakeholders generally want results they can trust. Thus, communicating findings in plain language is vital. Target trial emulation, by conceptually aligning with the familiar idea of a clinical trial, enhances credibility. Non-technical audiences can better accept an RWE conclusion if it is presented as "we designed this study like a trial, with clear start and end dates and defined groups".

There are also diverse opinions on the pace of RWE use. Some industry writers hail a "transformational" shift towards RWE and AI, expecting routine incorporation into drug development and approval decisions (www.tandfonline.com). Others caution that hype may oversell speed gains; cleaning and analyzing RWD still takes time and expertise. Nevertheless, cases like the FDA's acceptance of registry controls for rare disease approvals show that RWE can accelerate decisions when done well. The ultimate perspective is that RWE complements but does not replace RCTs. Where RCTs are impossible or too slow, RWE is invaluable. In other settings, RWE can confirm or refine RCT findings. Decision-makers should view them as two legs of evidence.

Implications and Future Directions

Looking forward, several trends will shape the use of RWE and target trial emulation:

- **Data improvements:** As healthcare systems adopt interoperable EHRs and data-sharing, RWD quality should improve. Wider use of common data models (like OMOP) will standardize variable definitions, making multi-center RWE studies more feasible. Moreover, new data types (genomics, wearables) will enrich RWD, but also pose new challenges for bias adjustment.

- **Methodological innovations:** Machine learning and AI can help in propensity modeling and finding latent confounders, but they require transparency about model behavior. Ongoing research on "double machine learning" and automating the Q-matrix in the Roadmap will make target trial analyses more robust and less labor-intensive. Additionally, methods for unmeasured confounding (e.g. instrumental variable approaches) may complement trial emulation when used carefully.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

- **Regulatory evolution:** FDA and EMA are expected to continue releasing guidance on RWE, likely including more on causal methods. The International Council for Harmonisation (ICH) is also working on templates for RWE submission. Regulatory bodies may start to expect trial emulation designs in more submission types (e.g. external control arms in single-arm trials). HTA agencies may eventually demand target trial-style analyses for efficacy claims based on RWD.
- **Education and standards:** For RWE to mature, the field needs standard reporting checklists (similar to CONSORT for RCTs). Some checklists are underway for RWE (e.g. RECORD, STaRT-RWE) that incorporate trial protocol elements. Broad adoption will help ensure consistency. Universities and training programs are already adding causal inference and RWE courses; we can expect more decision-makers to be conversant in basic concepts (just as epidemiology became part of medical education).
- **Ethical and societal issues:** A powerful RWE capability also raises questions. For example, if observational data can mimic trials, should some randomized trials be replaced by large-scale record studies? Ethical frameworks will need updating. Privacy concerns will also grow as more patient data is used; encryption and governance models must keep pace. Transparency to patients about how their data might inform evidence will be an important trust issue.
- **Real-world comparative trials:** Interestingly, the line between observational and randomized designs is blurring. *Pragmatic trials* randomize patients in routine care settings, often using EHR to capture outcomes. *Registry-based RCTs* enroll patients within a registry infrastructure. These hybrid designs can be seen as formal trial emulations with partial randomization. As more "real world trials" are conducted, the distinctions will merge. Decision-makers should recognize that trial emulation is part of a continuum from pure observational studies to actual RCTs.

In sum, the future of RWE looks promising but requires commitment to rigor. Decision-makers in healthcare, policy, and industry should encourage the use of structured, transparent frameworks like target trial emulation. This will enable them to lean on RWD when it is strong, without being misled when it is not. As the methodological literature puts it, we should aim to *"produce high-quality estimates of causal effects using RWD when possible, and to honestly evaluate whether the proposed methods are adequate for drawing causal inferences"* (pmc.ncbi.nlm.nih.gov). When done properly, RWE can significantly expand the evidence base for decisions, catching real-world subtleties that RCTs alone cannot.

Conclusion

Real-world evidence has moved from the fringes to the mainstream of medical and regulatory decision-making. It offers a way to learn from actual patient care at scale, filling gaps that traditional trials leave. However, observational data are fraught with complexities that can mislead naive analyses. The **target trial emulation** framework provides a disciplined solution: by explicitly designing an observational study as if it were a randomized trial, we can make much more reliable causal inferences. This involves specifying eligibility, treatment strategies, follow-up, and outcomes – all grounded in a hypothetical trial – and then executing that design in the data.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

Throughout this report, we have emphasized that target trial emulation is not just a statistical trick, but a philosophy of clarity and rigor (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). It forces analysts to confront biases head-on (for instance, by avoiding immortal time bias through proper cohort entry) and to adjust for confounding transparently. The many citations and case examples provided (from metastatic cancer to diabetes drugs) show that, when applied carefully, trial emulation can yield results consistent with gold-standard trials (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov) or can illuminate new truths when trials are not available (www.ahajournals.org) (www.ahajournals.org).

For decision-makers, the implications are clear. Do not dismiss RWE out of hand, but do demand strong study design. When evaluating observational findings, ask whether the study was explicitly structured like a trial. Check if it pre-specified comparisons, handled immortal time, and adjusted for key confounders. In regulatory reviews or reimbursement negotiations, insist on seeing the causal reasoning and sensitivity analyses.

Looking ahead, as data become richer and methods more advanced, RWE will only grow in importance. Yet the need for caution will remain: “95% of reviewed studies had a potential bias” (pmc.ncbi.nlm.nih.gov) reminds us that poor design can invalidate conclusions. Tools like the Causal Roadmap, machine learning diagnostics, and standardized reporting promise to raise the bar. Importantly, collaboration across stakeholders – regulators, industry, academia, clinicians, and patients – will ensure that RWE is harnessed to its full potential.

In conclusion, target trial emulation offers a bridge between the controlled world of RCTs and the messy realm of real-life practice. When used appropriately, it empowers decision-makers with causal insights drawn from large-scale, real-world experience – but without obscuring reality with jargon. The underlying principle is simple: *make your observational study as much like the perfect trial as you can*, and then interpret the results with the same care and caveats as a trial. By following this approach, we can maximize the utility of RWD while maintaining scientific integrity (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). Future healthcare decisions, grounded in robust RWE, will be better informed and more responsive to the needs of patients in the real world.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies including Scilex Holding Company (SCLX) and leading CROs across North America.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at https://intuitionlabs.ai/contact for a consultation.

© 2025 IntuitionLabs.ai - North America's Leading AI Software Development Firm for Pharmaceutical & Biotech. All rights reserved.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by Adrien Laurent, a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.