AKI case study and causal roadmap tutorial

Table of contents

[Step 1a – Causal Question and Causal Estimand 2](#_Toc210301161)

[Introduction 2](#_Toc210301162)

[Causal Estimand 4](#_Toc210301163)

[Alternative Estimands 7](#_Toc210301164)

[Note on the Implicit Estimand in the Completed SOF vs Non-SOF AKI Analysis 12](#_Toc210301165)

[3. Alternative Estimand Strategies and Their Causal Questions 14](#_Toc210301166)

[Estimands and Thier Trade-offs for SOF vs Non-SOF Case Study 24](#_Toc210301167)

[Bias, interpretability, generalizability summary 26](#_Toc210301168)

[Common Summary Measures and Pitfalls for Time-to-Event Outcomes 27](#_Toc210301169)

[Choosing a Summary Measure for Time-to-Event Data 29](#_Toc210301170)

[General Advice on Estimand Selection in RWE Time-to-Event Safety Analyses 32](#_Toc210301171)

[Step 1b — Causal Model 36](#_Toc210301172)

[Study-type & background knowledge 36](#_Toc210301173)

[Specifying the Causal Model (DAG) 36](#_Toc210301174)

[Step 2 – Observed Data & Statistical Model 40](#_Toc210301175)

[Objective 40](#_Toc210301176)

[Observed‑data structure 40](#_Toc210301177)

[Key relationships & gaps 41](#_Toc210301178)

[Missingness & measurement error 41](#_Toc210301179)

[Why Step 2 matters 42](#_Toc210301180)

[Step 3: Identifiability 47](#_Toc210301181)

[Introduction to *Identification*: Linking the HCV→AKI Causal Question to the Observed Data 47](#_Toc210301182)

[Identification assumptions 47](#_Toc210301183)

[Why these assumptions are plausible (or not) in claims data 48](#_Toc210301184)

[7 Take‑home message for pharmacoepidemiologists 49](#_Toc210301185)

# Step 1a – Causal Question and Causal Estimand

## Introduction

In this chapter, we will formulate the scientific question of interest as a causal question, introduce estimands, choose an estimand for the analysis of the SOF vs non-SOF AKI data, and then discuss the implications of this decision and alternative estimands

### Formulation of the scientific question

In the causal inference roadmap, clearly defining the scientific question and target estimands is critical for valid analysis and interpretability. The first step in the causal roadmap involves translating a clinical or regulatory question into a clearly defined causal question. A well-formulated causal question clearly identifies the target population, interventions (or exposures), and outcomes of interest, helping define the target estimand.

**Causal question**

Among U.S. adults (≥ 18 y) with chronic hepatitis C who newly initiate direct-acting antiviral therapy, what would the 90-day cumulative risk of first acute kidney injury be if every patient remained on their initial regimen for the full 90 days, comparing sofosbuvir-containing regimens with non-sofosbuvir regimens?

The question is articulated in Roadmap language: population, intervention, comparator, endpoint, time horizon. This aslo aligns with the **ICH E9(R1)** estimand framework, which emphasizes the importance of clearly defining the treatment condition of interest, population, endpoint, intercurrent events, and the population-level summary measure.

### Estimand Introduction – ICH E9(R1) Estimand Framework in RWE Analysis

What is an estimand? An estimand is a precise description of the treatment effect or quantity we aim to estimate, aligning the study objective with how data are collected and analyzed. This is distinct from the statistical estimator or estimate:

* **Estimand**: The target causal quantity we want (e.g. risk difference in 1-year incidence between treated vs. untreated).
* **Estimator**: The statistical method or formula we use to compute an estimate of the estimand from data (e.g. a regression model, TMLE algorithm, etc.).
* **Estimate**: The result we get from the estimator applied to our data (e.g. an estimated 1-year risk difference of -3%, meaning 3% fewer injuries with treatment).

By first nailing down the estimand, we ensure the estimator and resulting estimate are meaningful for the policy or clinical question.

The **ICH E9(R1) Addendum (2019)** introduced the estimand framework to ensure clarity and consistency from study design through analysis and interpretation. The document was drafted and agreed by the ICH E9(R1) Expert Working Group (EWG)— team of statisticians drawn from the regulatory authorities of every ICH Regulatory Member. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is a non-governmental consortium whose purpose is to create harmonised scientific and technical guidelines. In essence, an estimand defines “the target of estimation to address the scientific question of interest” for a study. By adopting the estimand mindset, RWE analyses can be more regulatory-aligned, as they explicitly state the causal question and handling of complexities in a way that regulators increasingly expect.

Key attributes of an estimand:

1. **Population** – the patients of interest
2. **Treatment** – the therapy or exposure condition and comparator
3. **Endpoint** – the outcome variable
4. **Intercurrent Events** – post-treatment events that can affect interpretation or existence of the outcome
5. **Summary measure** – how the treatment contrast is quantified (e.g., risk difference, hazard ratio)

For time-to-event outcomes (like time to acute kidney injury), this framework is especially important because multiple intercurrent events can occur during follow-up (patients might discontinue treatment, switch therapies, die, etc.), complicating how we interpret the outcome. ICH E9(R1) emphasizes that an estimand must address how each intercurrent event is handled as part of the clinical question; we cannot fully define “what effect we’re estimating” without specifying what we do if something like treatment discontinuation, addition of a rescue medication, or death happens during the study. This clarity is especially crucial in RWE studies, where such events are common and not under investigator control.

By pre-specifying the estimand, we ensure the study design and analysis align with the real objective, avoiding common pitfalls. Historically, analyses sometimes defaulted to “intent-to-treat” principles or ad-hoc censoring without explicitly stating the question being answered. The estimand framework forces us to be explicit: Are we estimating the effect of initiating treatment no matter what happens after, or the effect while actually on the treatment, or under some hypothetical scenario, etc.? This upfront clarity improves communication with regulators and stakeholders, since everyone knows which treatment effect we mean. In fact as you will see in this example analysis, the estimand drives all subsequent study steps – design, conduct, analysis, and interpretation – ensuring they target the same goal.

## Causal Estimand

### Target population and eligibility

* Adults ≥ 18 y with ≥ 12 months continuous enrollment and at least one HCV diagnosis.
* No prior exposure to any DAA or documented AKI in the 12-month baseline.
* Eligible for either regimen at the index dispense (pragmatic equipoise).

The simulated dataset (generate\_hcv\_data() in DGP.R) mirrors age, CKD, cirrhosis and HIV prevalence observed in the HealthVerity HCV cohort.

### Interventions (treatment strategies)

| Treatment regime | Description (target intervention) |
| --- | --- |
| **A = 1** | Initiate any **sofosbuvir-containing** DAA regimen on Day 0 and **continue the same regimen without switching or discontinuation for 90 days**. |
| **A = 0** | Initiate a **non-sofosbuvir** DAA regimen on Day 0 and **continue the same regimen without switching or discontinuation for 90 days**. |

Note: In the real-world claims data, patients may switch or discontinue therapy. We will handle such deviations analytically in step 5; patients are censored at first switch/discontinuation and inverse-probability-of-censoring weights (IPCW) are applied—to recover the above hypothetical sustained-exposure estimand.

### Outcome & follow-up

* Outcome First AKI within 90 d (simulation event-time generator).
* Time 0: Start at dispense (t = 0); stop at 90 d, AKI, death, disenrollment, or regimen switch—according to chosen estimand.

### Intercurrent-event handling

Here is a table demonstrating how different estimands handle different potential intercurrent events. Below, we will describe each of these estimands in details, but for now, focus on our chosen estimand, which quantifies the effect of SOF in a hypothetical would where no patients switched treatment, as well as the intention to treat (ITT) approach, where the effect of prescribing the drug is quantified.

| Intercurrent event | Intention-to-treat (treatment-policy) | While-on-Tx HR | **Hypothetical no-switch** | Principal stratum |
| --- | --- | --- | --- | --- |
| **Regimen switch** | Ignore; follow patient regardless of changes | Censor 1 day after first switch | Intervention prevents switching via modeled hazard | Individuals who *would not* switch by week 8 belong to the stratum |
| **Death** | Treated as independent censoring† | Same (censor) | Same | Same |
| **Disenrollment** | Independent censoring (IPCW) | Same | Same | Same |

† Death is uncommon in the first 90 days of DAA therapy; a competing-risk sensitivity analysis will be reported in Step 6.

#### Primary Estimand: Intention To Treat (Treatment-policy risk difference)

We propose using a while-on-treatment estimand (as-treated approach) as the primary estimand for the SOF vs non-SOF AKI analysis, which is the same approach as in the completed analysis, but we lay it out explicitly here.

**Definition:**  
[ \_{} ;=; !{Y^{(1)}} ;-; !{Y^{(0)}}. ]

This means our estimand will target the causal effect of continuous treatment with a SOF-containing regimen versus a non-SOF regimen on the incidence of AKI, while patients remain on their initial treatment.

In practical terms, the population is all patients initiating treatment for which either SOF or an alternative could be used; the treatment condition is “initiate and continue SOF-based therapy” vs “initiate and continue non-SOF therapy”; the outcome is time to AKI; intercurrent events: treatment discontinuation or switching will be handled with a treatment-period (while-on-treatment) strategy (we’ll censor follow-up at the time of regimen discontinuation/switch in each group), and perhaps define death as a competing risk or censoring event (since death can preclude observing AKI – one could treat death as censoring if unrelated to treatment or make a composite if considering “AKI or death” as a broader safety outcome, but for simplicity, let’s say we censor at death as well). The summary measure could be a hazard ratio or risk difference in AKI at a certain time point during therapy (e.g., by 12 weeks) -later we will advocate for the use of risk differences or ratios over hazard ratios for a more causal interpretation. This estimand aligns with asking: “Comparing SOF to no-SOF, what is the difference in probability of AKI during the treatment period?”

The table below summarizes the proposed estimand:

| Attribute | Recommended specification for the **primary treatment-policy estimand** |
| --- | --- |
| **Population** | Adults (≥ 18 y) with chronic HCV who are eligible—at first contemporary treatment initiation—to receive **either** a SOF-containing or a non-SOF DAA regimen. |
| **Treatment strategies** | *Initiate* a SOF-containing regimen **vs** *initiate* a non-SOF regimen on day 0; **follow patients regardless of any subsequent discontinuation, augmentation, or switch** (treatment-policy strategy). |
| **Intercurrent-event strategies** | • **Regimen switch / discontinuation** → *ignored* (patient remains in original arm).• **Death** → treated as independent censoring at date of death.• **Lost to follow-up / disenrollment** → treated as censoring; handled with inverse-probability-of-censoring weights. |
| **Endpoint** | Cumulative incidence of the **first AKI** (KDIGO-compatible ICD-10 algorithm) within 90 days of treatment initiation; algorithm PPV documented in appendix. |
| **Summary measure** | **Primary:** 90-day absolute risk difference (SOF – non-SOF).**Key secondary:** 90-day risk ratio and ΔRMST0–90. |
| **Sensitivity / supplementary estimands** | (i) While-on-treatment cause-specific HR (censor at switch).(ii) Hypothetical “no-switch” ΔRisk via switch-hazard TMLE.(iii) Principal-stratum ΔRisk among patients who would adhere ≥ 8 weeks. |
| **Rationale** | Treatment-policy estimand aligns with ICH E9(R1) for public-health decision-making: it answers *“What is the excess 90-day AKI risk attributable to prescribing SOF?”* Assumptions needed (baseline exchangeability, positivity, independent censoring) are diagnosable; estimator (TMLE) is double-robust and ML-enabled, facilitating regulatory acceptance. |

**Interpretation**  
Absolute change in 90-day AKI risk attributable to prescribing a SOF-containing regimen, regardless of subsequent switching or discontinuation.

**Regulatory relevance**  
Aligns with ICH E9(R1) “treatment-policy” strategy; informs population-level risk–benefit and product labeling decisions.

## Alternative Estimands

The **current Roadmap** designates as **primary** a *hypothetical “no-switch”* estimand:  
> *90-day cumulative risk of first AKI if every patient had remained on their initial regimen for the full 90 days, comparing SOF-containing versus non-SOF regimens.*

This contrasts with the **completed analysis**, which reported a **naïve as-treated, censor-at-switch** estimand (★).  
Below we list that original contrast alongside other common options and outline when each is most informative.

| Purpose | Recommended estimand | What it answers | Key identification concerns |
| --- | --- | --- | --- |
| **Primary safety question for this Roadmap** | **Hypothetical “no-switch” (IPCW-adjusted)** | Intrinsic nephro-toxicity if patients stay on the starting regimen | Requires correct model for switching hazard; time-varying positivity |
| Historical analysis (completed study) | ★ **Naïve as-treated, censor 30 d after switch** | Risk while *observed* on drug; simple but susceptible to informative censoring | Switching related to early creatinine rise can bias estimates |
| Regulatory/labeling decision (“start or not?”) | **Treatment-policy ITT** | Real-world effect of initiating SOF vs non-SOF regardless of later changes | Cross-over dilutes contrast but avoids selection bias |
| Mechanistic / per-protocol efficacy | **G-formula or MSM per-protocol** | Risk if everyone *remained* adherent, adjusting for time-varying confounders | Requires rich covariate history and correct longitudinal models |
| Pragmatic composite | **AKI *or* regimen change** | Captures clinical action signal (switching counts as an event) | Blurs biological vs behavioural pathways; easier for decision-making |

### Choosing among estimands

* **Regulatory safety signal** – *Hypothetical “no-switch”* is preferred when the agency needs an unbiased estimate of inherent renal risk **while patients remain on therapy**.
* **Clinical decision (“start vs do not start”)** – ITT is most relevant because it mirrors real-world prescribing where switches and discontinuations occur.
* **Mechanistic/toxicology** – Per-protocol or principal-stratum contrasts isolate biologic effect but rely on stronger assumptions.
* **Historical comparability** – Retain the ★ as-treated estimate as a secondary analysis to benchmark against the original report.

Subsequent Roadmap steps (identification, estimation, sensitivity) must align with the chosen estimand, and are written to target the **primary hypothetical no-switch estimand** and present the ITT and naïve as-treated results as prespecified sensitivity analyses.

#### Table of Alternative Estimands

| **Label** | **Summary measure / 90-d target** | **Handling of switching** | **Handling of death / other competing risk** | **Causal question addressed** | **Identification assumptions / caveats** | **Regulatory / scientific role** |
| --- | --- | --- | --- | --- | --- | --- |
| **★ As-treated (censor at switch)** | RD or RR for AKI ≤ 90 d among continuous users | Censor 30 d after first switch | Censor at death / loss / DB close | Effect of *staying* on the initial regimen through Day 90 | Censoring independent of outcome given baseline *W*, treatment *A* (plus time-varying covariates if used) | Counselling patients during therapy; quality-of-care benchmarking |
| **ITT (treatment-policy)** | ΔRisk\_0-90, RR | Ignore switch (follow regardless) | Censor at death / loss | Effect of *initiating* SOF vs non-SOF, irrespective of later changes | No informative censoring; treatment cross-over dilutes contrast | Public-health impact; drug-label language (chosen primary) |
| **While-on-Tx HR** | Cox HR (as-treated clock-reset) | Censor 1 d after first switch | Censor at death / loss | Instantaneous effect while continuously on the original regimen | Non-informative censoring; hazards proportional | Clinician decision-support during treatment |
| **ΔRMST** | RMST difference 0-90 d | Same as ITT (ignore switch) | Censor at death / loss | Mean delay/advance in AKI onset over 90 d | Same as ITT plus correct RMST model if cov-adj | Patient-centred interpretation when PH assumption fails |
| **Per-protocol (model switch)** | RD / RR (g-formula / MSM) 0-90 d | Switch treated as time-varying exposure and explicitly modelled | Censor at death / loss | Risk if everyone *remained* on assigned therapy | Correct specification of time-varying confounders & switching model | Mechanistic or biologic efficacy estimate |
| **Hypothetical “no-switch” (IPSW)** | ΔRisk\_0-90, RR | Censor at switch **and** weight by IPSW | Censor at death / loss | Risk had all switching been *prevented* | Correct model for switching hazard; positivity | Mechanistic safety assessment; sensitivity analysis |
| **Composite (AKI *or* switch)** | RD / RR for first AKI **or** regimen change ≤ 90 d | Count switch as an event | Death censored | Effect on “AKI-related adverse outcome” combining toxicity and clinical response | Composite mixes pathways; assumes same utility weight | Pragmatic decision-making when switch signals renal toxicity |
| **Controlled Direct Effect** | RD / RR 0-90 d | Censor at switch | Treat death as competing risk to be **eliminated** counter-factually | Kidney toxicity if death could not occur | Requires counter-factual elimination of death; strong, often implausible assumptions | Exploratory mechanistic inquiry; rarely primary |
| **Principal Stratum (always-adherers ≥ 8 wk)** | RD / RR 0-90 d within latent stratum | Restrict to patients who *would* stay ≥ 8 wk on original regimen | Censor at death / loss | Biological efficacy among guaranteed adherers | Stratification on unobservable behaviour; strong monotonicity / exclusion assumptions | Hypothesis-generating; subgroup exploration |

## Note on the Implicit Estimand in the Completed SOF vs Non-SOF AKI Analysis

**Current analysis approach:** In the original cohort analysis, patients were classified at baseline into “SOF” vs “non-SOF” groups, and then followed for the outcome (AKI) followed over a period of time until the earliest of: 1) first AKI diagnosis, 2) 30 days after the last supply of the index regimen or first switch to an alternative DAA, 3) death, loss of insurance eligibility, or database close. A Cox model on the matched cohort therefore targeted a while-on-treatment (as-treated) hazard ratio for AKI during exposure (plus the 30-day wash-out).

The critical question is: how did the analysis handle patients who discontinued their treatment early or switched to a different regimen? This determines the implicit estimand. The analysis used a “while on treatment” (as-treated) estimand – meaning it focused on AKI occurrences during the period patients stayed on their initial therapy. In practical terms, if a patient stopped the initially prescribed treatment or switched to a different therapy, the analysis would censor their follow-up at that point (no longer counting their later time or events) so that only time actually exposed to SOF or to the comparison regimen contributed to the risk calculations. This corresponds to an estimand where the treatment effect is defined up until discontinuation of the originally assigned treatment. Intercurrent events like stopping or switching therapy are handled by essentially excluding (censoring) data after those events, rather than following the patient regardless. The underlying scientific question for this estimand is: “What is the effect of continuous treatment with SOF vs non-SOF on the hazard of AKI, while patients adhere to the initial treatment?”

**Assumptions and implications:** This implicit as-treated/while-on-treatment estimand assumes that once a patient ceases the initial treatment, any further risk of AKI is no longer attributable to that treatment (hence not counted). It treats treatment changes as censoring events, which conceptually asks: what would the incidence of AKI be if patients had remained on their originally assigned treatment (up until the time of AKI or treatment cessation)? An important assumption here is that censoring due to treatment discontinuation or switch is not introducing bias – i.e. that patients who stop treatment are not systematically different (in unmeasured ways) in their risk of AKI than those who continue. In reality, this can be a strong assumption: for example, if some patients discontinued SOF because their renal function was worsening (an early sign of AKI risk), censoring them at discontinuation might miss AKI events related to the drug, underestimating the true harm. An alternative approach is to explicitly model the probability of censoring due to switch and re-weight the analysis through an inverse probability of censoring weighed analysis. Despite this risk, the as-treated approach aligns with how drug safety is often assessed (focusing on events during exposure). It answers a clinically relevant question: “What is the kidney safety profile of the drug during the time patients are actually taking it?” Alternatively, had the analysis ignored treatment changes and simply followed everyone from initial treatment assignment to outcome (no matter if they switched or stopped therapy), it would correspond to a treatment-policy estimand (an ITT-like approach). A treatment-policy estimand treats intercurrent events as irrelevant – each patient is analyzed in their original group regardless of subsequent adherence or changes.

If the SOF vs non-SOF analysis was done this way, it would implicitly be answering a different question: “What is the effect of starting SOF vs not, on risk of AKI, irrespective of whether patients stay on that initial treatment or not?” This would include AKI events even after therapy changes. Such an approach might be more pragmatic (reflecting real-world usage patterns), but in a short-term safety context like AKI during HCV therapy, it could dilute the observable effect of the drug. (For instance, if many SOF patients switched off the drug early, an ITT-style analysis might show little difference in AKI rates because those who would have been harmed stopped taking it – essentially comparing groups that become similar in exposure over time.) Hernán and Scharfstein caution against blindly using an ITT estimand when it doesn’t match the clinical question – e.g. an analysis that effectively compares “treat until completion” in one group vs “start treatment but possibly stop if toxicity arises” in the other could be asking an irrelevant or misleading question (https://www.researchgate.net/publication/333705608\_A\_constructive\_critique\_of\_the\_draft\_ICH\_E9\_Addendum#:~:text=ICE%29%20and%20treatment%20%28e).

In our case, a treatment-policy estimand would be akin to comparing “initiating SOF (even if one has to stop due to AKI risk) vs not initiating SOF”. That’s arguably not the question clinicians or regulators care about if the goal is to understand the drug’s inherent safety – they would be more interested in “what happens under the treatment while it’s given.” In summary, the retrospective SOF vs non-SOF AKI analysis most likely (even if not stated outright) adopted an as-treated (while-on-treatment) estimand, censoring at treatment discontinuation or switch.

This approach interprets the AKI outcome as a treatment-emergent adverse event – focusing on the period of exposure to each treatment. It inherently assumes that once a patient leaves the original treatment, their subsequent risk is outside the scope of our comparison. This estimand is appropriate for capturing on-treatment causal effects, but we must be mindful of its limitations: if patients who discontinue are different, or if the drug causes some latent injury that only manifests after stopping, the as-treated estimand could bias or undercount the true effect. These considerations motivate examining alternative estimand strategies.

## 3. Alternative Estimand Strategies and Their Causal Questions

Different estimand strategies address intercurrent events in different ways, each corresponding to a slightly different causal question. The ICH E9(R1) Addendum outlines several strategies (treatment-policy, hypothetical, while-on-treatment, principal stratum, etc.) (https://pmc.ncbi.nlm.nih.gov/articles/PMC8112325/#:~:text=Estimand%20strategy%20Measurement%20of%20interest,32).

Here are four key strategies applied to the context of post-market drug safety and our SOF vs non-SOF case study, and clarify what question each estimand is answering:

### Treatment-Policy (Intention-to-Treat-like)

Under a treatment-policy strategy, intercurrent events are ignored in the sense that patients are analyzed according to the initial treatment assignment regardless of what happens afterward. This is analogous to the classical ITT principle in trials. The outcome of interest is measured no matter if the patient discontinued the drug, switched to another therapy, or took additional treatments. The causal question answered is: “What is the effect of starting treatment A versus treatment B on the outcome, considering the policy of initially assigning that treatment and following patients thereafter, no matter if they continue it or not?” For our case: a treatment-policy estimand would compare all patients who started SOF to all who started non-SOF, in terms of AKI risk, regardless of treatment changes. This could be relevant for a public health or intent-to-treat perspective, e.g. a regulator might ask: “If we introduce SOF broadly, what is the overall impact on AKI rates in the population, accounting for the fact that some patients may stop or switch treatments?” The advantage of this strategy is that it uses all data and isn’t biased by who adheres to treatment – it mimics a pragmatic scenario. However, the trade-off is interpretability for safety: if many patients discontinue SOF due to early renal problems, a treatment-policy estimand will still count their later AKI (or lack thereof) in the SOF group, potentially underestimating the drug’s true nephrotoxic potential. It effectively answers a more diluted question, combining both the drug’s effect and the impact of stopping it. For drug safety questions, regulators and clinicians often find a pure ITT approach less informative, because it doesn’t isolate the period of actual drug exposure (as Hernán and colleagues note, an estimand that compares “never treat” vs “treat even in the face of contraindications/toxicity” is not clinically relevant.

### While-On-Treatment (As-Treated)

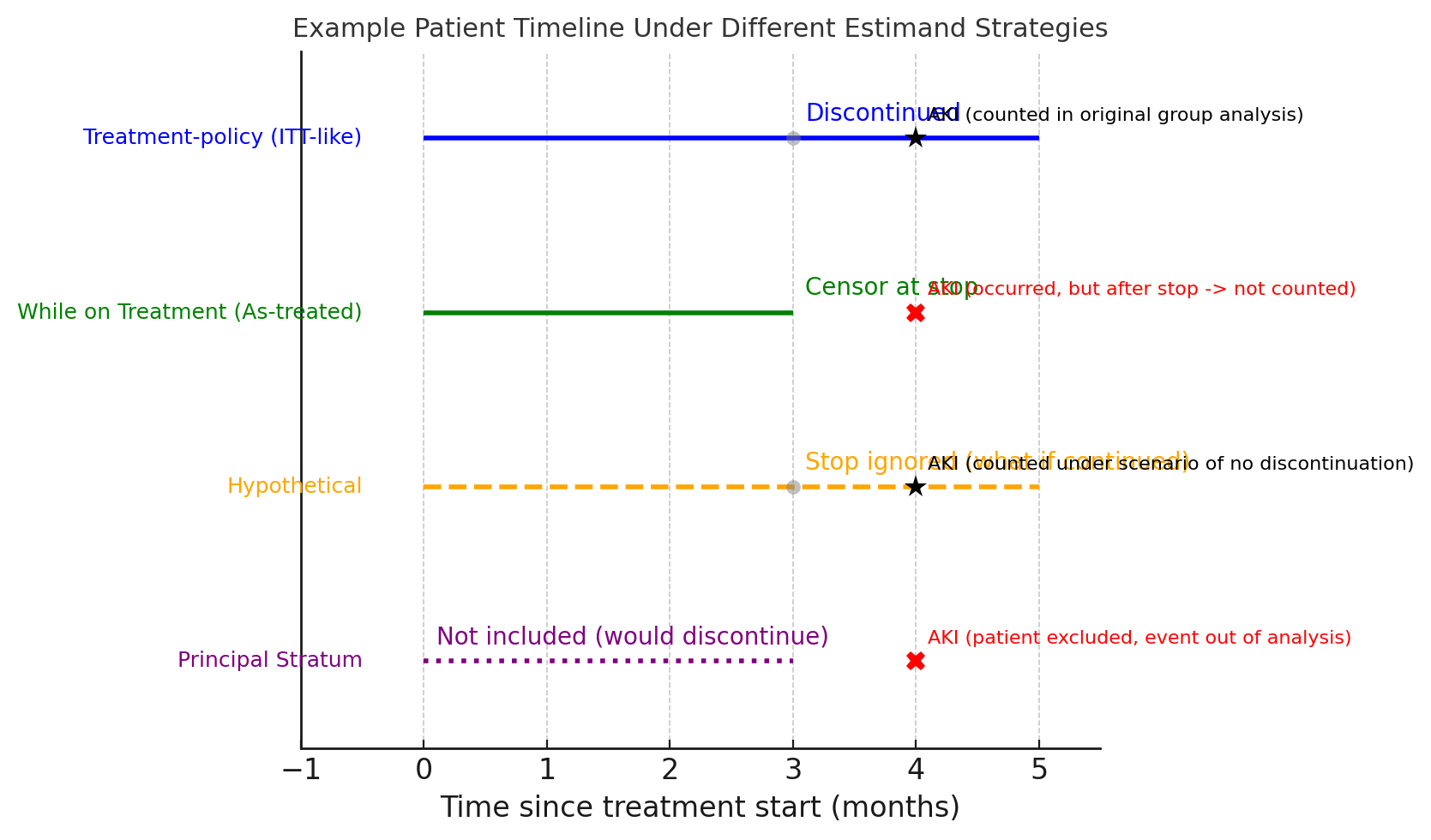
The while-on-treatment strategy (also called “as-treated” or “per-protocol” in some contexts) focuses on outcomes up until the time an intercurrent event occurs. Essentially, we follow patients only while they remain on the originally assigned treatment; if they discontinue or switch, we stop counting their outcomes thereafter (often treating them as censored at that point). The causal question here is: *“What is the effect of treatment A vs B on the outcome during the time patients actually stay on their intended treatment?”* This is the strategy we deduced the current SOF analysis used. In a safety context, this estimand is often most aligned with identifying the drug’s direct effects. For SOF vs non-SOF, it asks: *“Comparing patients continuously treated with SOF to those continuously treated with an alternative, what is the difference in AKI risk while on therapy?”* Any AKI that occurs after a patient stops or changes treatment is not attributed to the original treatment under this strategy. The benefit is clear interpretability for causation – it captures the period when the drug is actually present in the body and can cause harm. It aligns with the idea of treatment-emergent adverse events, which is often how clinical trials report safety (events are attributed to the treatment if they happened on or shortly after it). The strategy is particularly relevant to post-market pharmacovigilance if we want to know the risk during exposure – for example, it would feed into labeling like “AKI occurred in X% of patients during treatment”. The drawback is potential bias if the act of discontinuation is related to the outcome risk. In observational studies, one must consider that censoring at discontinuation assumes “non-informative” censoring (all prognostic factors leading to stopping are accounted for). If, say, patients with rising creatinine (a precursor to AKI) are more likely to stop SOF, a naive while-on-treatment analysis could censor them right before an AKI would have been observed, thereby underestimating AKI incidence on SOF. Despite this, analytic techniques (like modeling or inverse probability weighting) can mitigate some of this bias, and the while-on-treatment estimand remains very relevant for causal inference about the drug’s effect. It addresses the question most pertinent to clinicians: “What is the risk to my patient’s kidneys if they continue this drug versus if they were on an alternative?”

### Hypothetical Strategy

A hypothetical estimand poses a “what-if” question by assuming a certain intercurrent event did not occur, and then evaluating the outcome under that scenario. In other words, we imagine a hypothetical world in which, for example, all patients remained on their originally assigned treatment (no one discontinued or switched), and ask what the treatment effect on AKI would be in that world. This requires modeling or extrapolation because the data for patients after they discontinue are essentially missing in that hypothetical scenario. In our SOF example, a hypothetical estimand might be: “What would the incidence of AKI be comparing SOF vs non-SOF if no patient in either group ever discontinued or switched treatments (i.e., if everyone completed the full intended course of therapy)?” This is conceptually similar to the while-on-treatment approach, with the difference that it explicitly frames it as a counterfactual scenario and often involves imputing or modeling outcomes for those who did discontinue. The hypothetical strategy is useful if we want to remove the effect of certain real-world behaviors to isolate the treatment’s effect under optimal conditions or specific conditions. In post-market safety, one might use a hypothetical estimand if interested in, say, the safety profile had patients been able to tolerate the drug continuously. It is somewhat less common in pure safety analyses, but could be relevant for understanding intrinsic drug effects. The challenge with the hypothetical approach is implementation – because some outcomes are not observed (due to the event we assume away), one must rely on assumptions or statistical models (e.g., modeling the AKI risk as if treatment had continued for those who stopped). If those assumptions are wrong, it can introduce bias. Conceptually, though, it asks a clear causal question and is aligned with counterfactual reasoning. Regulatory guidance acknowledges hypothetical estimands as valid in certain cases (for example, in trials, “what would outcome be if patients did not take rescue medication” is a common hypothetical scenario. In RWE, we would use this strategy sparingly and carefully, usually to supplement an analysis – for instance, to estimate the full-course effect of SOF vs non-SOF if early discontinuation is frequent.

### Principal Stratum Strategy

The principal stratum strategy focuses on a subset of patients defined by some post-treatment behavior or event (often a subset in which an intercurrent event does or does not occur). The estimand is then the treatment effect within that subgroup – the subgroup is defined in a way that is not affected by treatment assignment (hence “principal stratum” in causal inference terms). A classic example: “the effect of treatment among those patients who would adhere to treatment in both the treatment and control conditions.” In our context, one could think of a principal stratum like “patients who would complete a full course of therapy regardless of whether they were on SOF or non-SOF.” That stratum would exclude patients prone to discontinuation. The causal question becomes: “Among patients who (hypothetically) would stay on their assigned treatment no matter which treatment they got, what is the effect of SOF vs non-SOF on AKI?” This is appealing because it compares outcomes in a group of patients unaffected by adherence issues – essentially a per-protocol effect without the usual adherence bias, if it could be identified. Another principal stratum of interest could be “the effect in patients who would not experience acute liver failure (another competing event) under either treatment” – any intercurrent event can define a stratum. The issue is that principal strata are defined counterfactually and can’t be directly observed (we don’t know which patients would adhere under both scenarios, for example). Estimating effects in principal strata often requires strong assumptions or specialized methods (e.g. monotonicity assumptions, sensitivity analyses). It’s rarely used as the primary estimand in observational studies because of these challenges. However, it’s conceptually relevant: for instance, regulators might be interested in the “intrinsic efficacy or safety in patients who can tolerate the drug”. In our SOF example, a principal stratum estimand might help answer: “If we consider only those patients who are able to complete therapy, does SOF have a different impact on AKI than comparator?” One could attempt to approximate this by analyzing a subset of patients who actually completed treatment in both groups (acknowledging that introduces selection bias). Trade-offs: principal stratum estimands improve interpretability for a specific scenario (e.g., ideal adherers), but sacrifice generalizability (the effect may not apply to all patients) and are hard to estimate without bias. They are more often seen in clinical trial contexts or sensitivity analyses. For routine pharmacovigilance, one would typically not choose principal stratum as the primary approach, but it’s good to be aware of it as a conceptual tool – it reminds us that treatment effects can differ in sub-populations defined by post-treatment events (for example, “compliers” vs “non-compliers”).



**Figure 1:** Illustration of how a hypothetical patient’s AKI event would be handled under different estimand strategies. In this scenario, the patient starts treatment at time 0, discontinues therapy at month 3 (gray circle), and then experiences an AKI at month 4 (star symbol). Under a treatment-policy estimand (blue line), we ignore the discontinuation and count the AKI as an outcome for the original treatment group (the patient is analyzed as if they were still on treatment). Under a while-on-treatment estimand (green line), we would censor the patient at the moment of discontinuation (vertical tick mark at month 3); any AKI occurring after stopping (the red “X” at month 4) is not counted for the original treatment. The hypothetical estimand (orange dashed line) imagines the patient had not discontinued – effectively, it considers the AKI at month 4 as if the patient were still on treatment (treating the gray stop marker as ignored). The principal stratum estimand focusing on completers (purple dotted line) would exclude this patient entirely, since they did discontinue; thus, their AKI outcome is outside the analysis for that stratum. This figure highlights how each strategy defines the “effect of treatment” slightly differently: treatment-policy uses all outcomes regardless of adherence, while-on-treatment uses only outcomes during actual treatment, hypothetical projects outcomes as if adherence were perfect, and principal stratum restricts to those without the intercurrent event. Each of these strategies can be appropriate depending on the objective of the study. In a post-market safety setting, we typically ask: “Does the drug cause the adverse outcome when used in practice?” If we are aiming to isolate the drug’s causal effect, a while-on-treatment or certain hypothetical estimand is often most suitable. If instead we care about the public health impact of using the drug (where adherence issues are part of the picture), a treatment-policy estimand might be informative. It’s important to formulate the estimand that best matches the causal question we want answered, as we will do next for the SOF vs non-SOF analysis.

### Unified Estimand Table

This table summarizes these four key estimands as well as some alternatives/extensions for time-to-event settings.

|  | Estimand Label | Population (P) | Intervention / Comparator (A) | Outcome / Endpoint (Y) | Handling of Intercurrent Events | **Time-horizon** | Summary Measure | Identification Assumptions | Primary Estimation Strategy | Rationale / Use-case | **Regulatory perspective** | **Used in AKI Study?** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | **Primary “Per-Protocol with Censoring at Switch”** | Adults with chronic HCV initiating first SOF vs non-SOF DAA, U.S. claims 2016–2023 | Start SOF-containing DAA vs active non-SOF comparator (new-user) | First AKI event within treatment + 30-day risk window | Switch → censor; other stops → censor; disenrol/death → censor | 0–180 d | RD & RR; cause-specific HR | Consistency; baseline exchangeability; non-informative censoring | 1:1 PS-matching ➜ Cox; TMLE for RD/RR | Regulatory signal-evaluation; simple to explain | **Mechanistic (biologic toxicity while on drug)** | **✓** |
| 2 | Treatment-Policy (Intention-to-Treat) | Same as #1 | Initiation strategy only; ignore later switching | AKI regardless of regimen changes | Treatment changes ignored; death/disenrol → censor | 0–180 d | RD & RR | Standard identifiability; independent admin censoring | KM / TMLE-IPW | Reflects real-world use | **Pragmatic primary** |  |
| 3 | Hypothetical “No-Switch” World | Same as #1 | Counterfactual risk if everyone remained on initial regimen | AKI up to 180 d | Model switching hazard; set hazard = 0 in g-formula | 0–180 d | Counterfactual RD | Positivity for remaining on drug; correct model for hazards | Continuous-time TMLE with zeroed switch hazard | Removes bias from informative switching | Mechanistic |  |
| 4 | Per-Protocol Weighted (IPCW for Switch) | Same as #1 | Remain on initial regimen; switch → IPCW | AKI as above | Switch → weight; death/disenrol → censor | 0–180 d | Weighted cumulative incidence; RD | Positivity & correct switch model; independent censoring | Stabilised IPCW ➜ KM / TMLE | Efficient alternative to simple censoring | Mechanistic |  |
| 5 | Composite Endpoint (AKI ∨ Switch ∨ Discontinue) | Same as #1 | Same as #1 | First of AKI, treatment switch, or discontinuation | Switch/stop **count as events** | 0–180 d | CIF diff; cause-specific HR | Competing-risk assumptions; independent censoring | Fine–Gray / CIF-TMLE | Captures any renal-related safety signal | Signal-detection composite |  |
| 6 | AKI with Death as Competing Risk | Same as #1 | Same as #1 | AKI; death treated as competing risk | Death → competing; switch → censor/weight | 0–180 d | Sub-distribution HR; CIF diff | Independent censoring/death | Fine–Gray / Aalen–Johansen TMLE | Policy-relevant when mortality >0 | Pragmatic supplementary |  |
| 7 | Long-Term (365-day) Treatment-Policy | Same as #1 but 1-year follow-up | Same as #2 | AKI within 365 d | Ignore treatment changes | 0–365 d | 1-yr RD & RR | As in #2 | TMLE RD/RR | Long-run public-health view | Pragmatic long-horizon |  |
| 8 | On-Treatment (While-On-Regimen) | Same as #1 | Effect during exposed time only; stop follow at drug stop | AKI during exposed time | Off-drug time → censor; switch → censor | 0–∞ (exposed time) | Incidence-rate diff; IRR | Independent censoring | Poisson / TMLE offset | Pure biologic exposure effect | Mechanistic |  |
| 9 | **Restricted Mean Survival Time (RMST)** | Same as #1 | As Treatment-Policy | AKI | Ignore treatment changes | 0–90 d | Δ RMST(0–90 d) | Standard identifiability | TMLE-RMST / g-comp | Robust to non-PH; intuitive “delay in AKI” | Pragmatic & communicable |  |
| 10 | Principal-Stratum (Would-Adhere ≥ 8 wk) | Same as #1 | Contrast within latent subgroup that would adhere 8 weeks to either regimen | AKI within 180 d | Restrict to principal adherer stratum | 0–180 d | RD in principal stratum | Monotonicity; strong principal-stratum ID assumptions | Bounding / sensitivity; TMLE-principal | Identify highest-risk subgroup | Exploratory biologic |  |

\*More detailed justification for chosen estimand in the appendix

## Estimands and Thier Trade-offs for SOF vs Non-SOF Case Study

Having recommended the while-on-treatment estimand, it’s important to acknowledge other reasonable estimand choices and discuss their pros and cons in this context. Below is a table comparing different estimand options, then a more detailed description of key estimands for time-to-event outcomes.

### Intention-to-treat as an alternative:

A treatment-policy estimand (ITT-like) could be justified if our objective was more about the overall impact of initiating SOF on patient outcomes, rather than the direct causal effect of the drug. For example, a public health authority might be interested in “if 1,000 patients are started on SOF vs 1,000 on older regimens, how many AKI cases will ultimately occur in each group?” This includes the fact that some patients may stop the drug – which in practice could limit the harm (because those who experience issues stop early). A treatment-policy analysis would capture that dynamic. The advantage is it reflects “real-world use” including adherence patterns. It also avoids the need to model or adjust for post-baseline variables – you simply follow everyone.

In a randomized trial, ITT is best for efficacy to avoid bias; for safety in observational data, however, the situation is trickier. The drawback of treatment-policy here is potential dilution of the drug’s effect. If SOF truly causes AKI, many patients might discontinue at first sign of kidney issues – those patients might avoid full-blown AKI (which is good for them, but the analysis would then count them as not having AKI while on SOF, even though SOF precipitated the problem that made them stop). Thus, the ITT estimand might conclude “no big difference in AKI rates,” whereas in truth SOF had nephrotoxic potential but it was mitigated by clinicians stopping the drug. From a regulator’s perspective, that ITT result is less useful because it doesn’t reveal the drug’s inherent risk. That said, a treatment-policy estimand could be an equally valid secondary analysis – it answers a complementary question: “what is the risk difference in a world where patients and doctors behave normally (stopping when needed)?” This might be relevant for risk-benefit assessment on a population level. For instance, if ITT shows only a tiny increase in AKI (because many at-risk patients stopped early), regulators might consider that in context of drug benefits. But they would still want to know the on-treatment risk to properly caution and manage patients. In summary, treatment-policy is not our choice for primary estimand due to interpretability concerns for safety, but it is not “wrong” – it’s just answering a different question. It has lower internal bias (no selection due to censoring) but mixes in the effects of patient management.

**Hypothetical estimand trade-offs:** The hypothetical strategy (imagine no one discontinued) in practice often ends up looking similar to the while-on-treatment analysis, except achieved via modeling rather than actual censoring. If our data shows, say, 90% of patients completed therapy, a while-on-treatment estimand already is very close to “everyone continued” scenario. If discontinuation rates are higher, one might consider a hypothetical estimand to estimate “if 100% continued.” The benefit of the hypothetical approach is that it directly addresses the question of full adherence without excluding those patients from the analysis (unlike censoring). Modern methods (like multiple imputation for treatment continuation) could be used to estimate what their outcomes might have been. The cost, however, is the strong modeling assumptions required. If we model kidney outcomes beyond discontinuation, we must correctly account for why they discontinued and how that relates to AKI risk. Mis-specification can lead to bias as well. In regulatory settings, hypothetical estimands have been used for efficacy (e.g. “if no rescue medication, what would the outcome have been” in glucose trials). For safety, it’s less common to explicitly do a hypothetical because the while-on-treatment analysis already addresses a similar question in a more straightforward way (by using observed data up to discontinuation). Nevertheless, one could consider a hypothetical estimand as a sensitivity analysis: for example, “assuming patients who stopped SOF had the same risk profile as if they stayed on it, how many AKIs would have occurred?” If that yields similar results to the as-treated analysis, it increases confidence; if it differs, it suggests the censoring might have missed something. In the SOF vs non-SOF study, a hypothetical estimand isn’t necessary given we can handle things via censoring, but it’s conceptually equivalent to ensuring our inference is about the treatment effect under continuous use.

### Principal Stratum considerations:

As noted, a principal stratum estimand like “effect among those who would adhere to treatment in both groups” is hard to identify but could be of scientific interest. It essentially removes the noise of non-adherence from both sides. In the real world example, adherence is high, so the principal stratum will mirror the

The trade-off is that it pertains to a subset of patients (those who can tolerate and stick with therapy). If, for instance, younger patients with no comorbidities are the ones who would never stop either treatment, the principal stratum effect applies mostly to that kind of patient. It might be different from the effect in older patients who tend to discontinue more. Thus, principal stratum estimands sacrifice generalizability – you answer a very specific question about a hypothetical subgroup. They also require either simplifying assumptions or sophisticated causal inference techniques (like instrumental variable approaches or sensitivity analyses) to estimate. In an observational SOF vs non-SOF study, one way to approximate this is to restrict the analysis to patients who actually completed therapy in both groups (observational analog of per-protocol analysis). That is essentially conditioning on adherence, which is biased (adherers may differ from non-adherers). So, one might use inverse probability weighting to create a pseudo-population that represents the principal stratum of completers. This is advanced and would likely be beyond what an RWE team needs for a routine analysis unless there’s a strong reason to focus on that question. Generally, we’d mention principal stratum only if, say, the sponsor or regulator asked: “What is the effect in patients who can actually complete the therapy course? Is it different?” If that’s a concern, one could do a secondary analysis on completers, acknowledging its limitations.

## Bias, interpretability, generalizability summary

Each estimand involves trade-offs:

* Bias: Treatment-policy avoids bias from informative censoring but can “dilute” the effect; as-treated isolates the effect but can introduce bias if not handled properly (need to account for why censoring happens). Hypothetical relies on model assumptions (risk of model bias). Principal stratum avoids some biases by narrowing focus but introduces others (selection bias, unless perfectly adjusted).
* Interpretability: As-treated and hypothetical directly address causal effects of the drug (easy for clinicians to interpret “while on drug vs while off drug”). Treatment-policy is a mix of drug effect and adherence behavior – interpretability is a bit more complicated (“effect of starting treatment strategy”). Principal stratum is clear in meaning (“effect in this subgroup”) but that subgroup is not directly observable – a bit abstract.
* Generalizability: Treatment-policy might generalize best to broad practice (since it includes everything that happens in practice). As-treated/hypothetical generalize to situations of good adherence or controlled treatment use – which is often what we want for understanding the drug, but if adherence patterns differ across populations, the effect could differ. Principal stratum by definition is not aiming to generalize to the whole population, just a part.

In the case of SOF vs non-SOF, we believe the bias introduced by as-treated is manageable (we can adjust for baseline differences and we will demonstrate techniques to handle differential censoring using targeted maximum likelihood estimation) and is worth the gain in causal interpretability. Treatment-policy would be more likely to understate a true causal harm if one exists (a form of bias in estimating the causal effect of continuous treatment, though it’s unbiased for the “policy” effect). Importantly, if the estimand is properly aligned to the question, then “bias” must be defined relative to that question. So an ITT analysis isn’t “biased” if the estimand of interest was truly the policy effect; it just might be answering a less relevant question for safety. Thus, a key part of estimand selection is picking the question such that the resulting estimate is meaningful and actionable. For completeness, we might pre-specify that we will also estimate a secondary treatment-policy estimand for AKI, to see the difference. If, for example, the as-treated analysis shows a significantly higher AKI risk on SOF, but the ITT analysis shows little difference, that tells a story: the drug causes problems, but patients and doctors are mitigating it by discontinuing (so ultimate outcomes converge). Both pieces of information can be valuable. We would explain that difference in regulatory discussions – it actually illustrates how clinical management can reduce harm. Conversely, if both ITT and as-treated show a similar elevated risk, it means even including the “real-world adherence” factor, SOF still carries the same risk, underscoring a robust safety signal.

## Common Summary Measures and Pitfalls for Time-to-Event Outcomes

In time-to-event (survival) analysis, a common practice is to report a hazard ratio from a Cox proportional hazards (PH) model. While hazard ratios are popular, we need to understand their limitations for causal interpretation. Let’s unpack what estimands are possible in time-to-event studies and why the hazard ratio may fall short.

### What Estimand Are We Getting from a Cox Model?

The Cox PH model gives a hazard ratio – roughly, the ratio of instantaneous event rates between two groups (exposed vs. unexposed) at any given time. If you include covariates, the Cox model yields an adjusted hazard ratio (often interpreted as the effect of treatment *holding confounders constant*). Many epidemiologic studies report only this HR as the measure of effect.

However, the hazard ratio is **not a direct probability, risk, or survival difference**. Importantly, a hazard ratio from an observational Cox model is a conditional measure (conditional on covariates and on surviving up to a given time). It does not directly answer questions like “how many more/fewer patients have the event by 12 months if treated vs. untreated?”. In fact, the hazard ratio “does not correspond to a clearly defined causal effect” on its own. It’s a rate ratio averaged over follow-up and, unless hazards are proportional and no other biases, it isn’t straightforward to interpret causally.

### Pitfalls of the Hazard Ratio:

* **Assumes Proportional Hazards**: Cox models assume the hazard ratio is constant over time. In reality, treatment effects may start strong and wane, or vice-versa. If the HR is not constant, the single number reported is some complex average of time-varying effects. For example, a treatment might increase early risk but improve long-term outcomes, yielding an average HR ~1.0 – masking important time patterns. Reporting only an “average” HR can be misleading if effects change over follow-up.
* **Built-in Selection Bias (Survivor Bias)**: The hazard at time t is among those who have not yet had the event by t. If treatment affects who remains event-free, the treated and untreated groups at later times become inherently different (“depletion of susceptibles”). Miguel Hernán pointed out that period-specific hazard ratios have a “built-in selection bias”. For instance, if susceptible individuals in the treatment arm experience the event early, the remaining treated patients are a healthier subset, which can make the hazard ratio appear to favor treatment later regardless of true long-term effect. This is one reason a treatment with no real long-term benefit could show an HR < 1 in later years purely by selection of who’s left.
* **Lack of Collapsibility**: Hazard ratios (like odds ratios) are non-collapsible, meaning the adjusted HR is not equal to any simple ratio of marginal (population-level) risks. Even if there is no confounding, conditioning on covariates can change the numerical value of an HR. This makes it hard to interpret the adjusted HR as a population effect. In contrast, measures like risk differences are collapsible (they can be aggregated without distortion).
* **Clinical Interpretation**: Physicians and policymakers often find absolute probabilities more intuitive (e.g. “5% of patients had kidney injury with drug vs 8% without”). A hazard ratio of 0.7 does imply a relative reduction, but it’s not obvious how that translates to absolute risk reduction without additional calculations. As the CONSORT guidelines note, reporting both absolute and relative measures is ideal, “as neither alone gives a complete picture”. An HR alone doesn’t tell you baseline risk or NNT (number needed to treat).

Bottom line: A hazard ratio from a Cox model is a useful associative measure, but endowing it with a causal interpretation is tricky. Hernán’s article “The Hazards of Hazard Ratios” cautioned that treating HR as *the* causal effect measure is risky. It may obscure time-varying effects and introduce bias due to the very way it’s defined over time at risk. Indeed, one review bluntly states that an HR from a Cox model “may be estimated… but does not correspond to a clearly defined causal effect”.

### Illustrative Example: Hazard Ratio vs. Absolute Risk

Consider a hypothetical (but inspired by real data) example: In a cohort of patients with chronic HCV infection, suppose we compare those treated with a new antiviral vs. those untreated, and we observe a hazard ratio of 0.70 for developing chronic kidney disease (CKD). In fact, a published real-world study found about a 30% hazard reduction in CKD risk with HCV treatment (HR 0.70, 95% CI 0.55–0.88). This suggests the treatment is beneficial. But what does HR = 0.70 mean in tangible terms?

* If the 5-year cumulative incidence of CKD in untreated patients is, say, 5%, an HR of 0.70 might correspond to roughly a 3.5% 5-year incidence in treated patients. That’s an absolute risk reduction of ~1.5 percentage points (5% vs 3.5%).
* If untreated risk is higher, e.g. 15% over 5 years, HR 0.70 might correspond to ~10.5% in treated – an absolute reduction of 4.5%.

The hazard ratio alone doesn’t tell us these absolute risks. In the HCV example, the authors reported incidence rates: untreated patients had about 10.8 CKD cases per 1000 person-years versus 6.7 per 1000 PY in effectively treated patients. Over a few years of follow-up, that implies only a few percent of patients developed CKD in either group. So the HR=0.70, while showing a relative benefit, translates to a modest absolute risk difference (a few fewer cases per 100 patients treated). This absolute effect size might matter for cost-benefit or clinical decisions, but it’s not apparent from the HR alone.

Especially in RWE settings with relatively low event rates, an impressive HR can correspond to a small absolute risk reduction. To fully answer our causal question (“should we treat HCV to prevent CKD?”), we likely want to know the absolute benefit (e.g. percentage of patients spared CKD over X years by treating). That is why careful summary measure selection is needed – perhaps we want our summary measure to be the risk difference at 5 years, rather than a hazard ratio.

## Choosing a Summary Measure for Time-to-Event Data

Given the pitfalls above, how should we define our estimand for time-to-event outcomes? The estimand should align with a meaningful causal contrast. Common choices include measures of survival probability or cumulative incidence at a certain time, or contrasts of survival distributions. Here are some estimand options:

* **Absolute Risk (Cumulative Incidence) at time** : e.g. “Probability of being event-free through 12 months.” From this we can derive risk difference or risk ratio between groups at time .
* **Survival Curve Difference**: Comparing the entire survival curves over time between treatment and control (e.g. showing adjusted Kaplan-Meier curves). This can be summarized at specific time points (risk at 1 year, 2 years, etc.) or by an area/difference measure.
* **Restricted Mean Survival Time (RMST)**: The average time without the event up to a milestone time (the area under the survival curve up to ). The difference in RMST between groups is an estimand (how much longer, on average, patients survive without the event with treatment vs. control, within a fixed horizon). This is an alternative summary that is often more interpretable when hazards are non-proportional.
* **Hazard-based estimands**: If truly the hazard function itself is of interest, one could define summary measures like a time-specific hazard ratio at a certain time, or the average hazard ratio over follow-up. However, as discussed, these are harder to interpret causally, so they are less commonly chosen as target summary measures in a causal analysis (they might be more a by-product of a model).

The key is to pick an summary measure that directly answers the causal question. Often for decision-making, risks and risk differences are very useful summary measures:

* **Risk difference at time**  (also known as absolute risk reduction): Tells how much the treatment changes the probability of the outcome by time . It’s easy to interpret (e.g. “treatment reduces 1-year event risk by 5 percentage points”) and can be translated to Number Needed to Treat (NNT = 1/(risk difference)).
* **Risk ratio at time**  (or relative risk): Tells how many times more or less likely the outcome is by time in one group vs. the other (e.g. “0.5 times as likely at 1 year” meaning a 50% relative reduction). Some prefer relative measures for their stability across populations.
* **RMST difference**: Tells the average gain or loss in event-free time within a certain period due to the treatment (e.g. “on average, patients lived 2 months longer without kidney failure over a 3-year period with treatment”). This can be very intuitive in some contexts (like quality of life or survival time gained).
* **Hazard ratio**: as discussed, it’s a relative measure of hazard rates. It’s commonly reported but should be linked to a causal estimand if used. For example, one might define the estimand as something like “the hazard ratio if the treatment were applied to everyone vs to no one, under PH assumption”. However, because of its issues, many causal analyses de-emphasize the HR as the primary estimand. Instead, they might report HR as a secondary analysis, acknowledging its limitations.

Below is a summary table of different summary measures for time-to-event outcomes and when you might use them:

Common summary measures for time-to-event outcomes and their interpretations. Each summary measure answers a slightly different question. Generally, for causal inference in RWE, absolute risk measures (risk differences) are recommended to convey real-world impact, often alongside a relative measure.

| Summary measure | Definition | Use Case & Interpretation |
| --- | --- | --- |
| Risk (Survival) at time | Probability of having (or not having) the event by a specific time under a given treatment strategy. Often expressed as cumulative incidence. | Useful for clear time-bound outcomes (e.g. 1-year event risk). Directly interpretable. Can compare risk in Treatment vs. Control to get risk difference or ratio. |
| Risk Difference at | Difference in cumulative incidence by time between two strategies (e.g. treated minus untreated risk). | Best for communicating absolute effect. Answers “How many fewer (or more) events by if everyone treated vs. no one treated?”. Policy-friendly (can compute NNT). |
| Risk Ratio at | Ratio of cumulative incidences by time between two strategies. | A relative measure at a concrete time point. Interpretation: “Patients treated have 0.xx times the risk by 12 months compared to untreated.” Useful for epidemiologic comparison; still fairly intuitive if communicated as “% reduction”. |
| Hazard Ratio (over follow-up) | Ratio of hazard rates (instantaneous event risk) between groups, typically assumed constant in Cox PH model. It’s a conditional relative measure, averaged over the follow-up period. | Common in literature, but caution: does not directly translate to absolute risk. Use when PH assumption is reasonable and when a relative rate measure is needed. Always consider also presenting absolute measures. Not a pure causal estimand unless proportional hazards and no confounding (in RCT). |
| Restricted Mean Survival Time (RMST) Difference | Difference in the area under the survival curve up to time between two groups. Equivalently, the difference in average event-free time by . | Good when timing of events matters or when hazards are non-proportional. E.g., “Over 5 years, treatment A gives 3 months more event-free survival on average than treatment B.” Clinically intuitive in many settings (time gained). |
| Median Survival Time Difference (if applicable) | Difference in median time to event between groups (or ratio of medians). | Sometimes used in oncology (e.g. median survival). Requires enough events to estimate median. Interpretation: how much longer median survival is with treatment. Could be considered if PH fails and median is of interest. |

Choosing the summary measure depends on the question: For example, if stakeholders care about “how many events are prevented by treatment within 1 year,” the summary measure should be a 1-year risk difference. If they care about long-term prognosis, perhaps a 5-year survival probability or RMST difference is appropriate. The summary measure should be decided first; the statistical approach comes next to estimate that summary measure.

## General Advice on Estimand Selection in RWE Time-to-Event Safety Analyses

The AKI safety study primarily adopted an **as-treated (per protocol), censor-at-switch estimand (★)**. This choice reflects a regulatory and clinical preference for quantifying the direct, drug-induced renal toxicity experienced while actually receiving the assigned therapy. **Why choose the as-treated estimand?**

* Aligns closely with standard regulatory practice for safety signals.
* Clearly interpretable in terms of drug-exposure-related harm.
* Assumptions required (non-informative censoring) considered acceptable in this short-term context. Alternative estimands are planned as secondary or sensitivity analyses to assess robustness and potential biases arising from informative censoring or switching.

To conclude, here are some generalizable tips for selecting estimands in real-world evidence studies of post-market safety (time-to-event outcomes):

* Start with a well-defined research question: Clearly articulate what decision or effect you are interested in. Is it the effect of initiating a therapy vs not (or vs another therapy) on an outcome, regardless of what happens later? Or the effect of actually being exposed to the therapy on that outcome? Or perhaps the effect in a particular patient subset? Writing this out in plain language helps identify the estimand. For example: “Do patients have a higher risk of AKI while on Drug A compared to Drug B?” points to a while-on-treatment estimand, whereas “If we prescribe Drug A instead of Drug B, will fewer patients develop AKI within 1 year?” points to a more treatment-policy estimand.
* Emulate a “target trial” to define estimand attributes: As a framework, imagine you were designing the ideal randomized trial to answer your question.
* Define who you would include (population), what the treatment and comparison arms would be (treatment strategies), how you’d handle changes in treatment in the protocol, when follow-up would start and end, and what outcome you’d measure. This exercise naturally defines the estimand. In the trial protocol you’d specify if patients are allowed to switch or if they’d be taken off study drug upon certain events – those translate to estimand strategies. By doing this, you make sure your observational study’s estimand is concrete. For instance, target trial emulation for a safety study might decide that if a patient in the trial has a toxicity, they discontinue per protocol – aligning with a while-on-treatment estimand (because after discontinuation, their outcome isn’t counted towards primary endpoint). You then ensure your observational analysis mirrors that (censor at discontinuation in analysis). Hernán et al. note that making the target trial explicit is a good practice and a “reasonably well-defined trial” should be emulated as closely as possible – if you can’t even define a meaningful target trial, the RWE study question might be too vague.
* Identify all relevant intercurrent events: In time-to-event analyses, common intercurrent events include treatment discontinuation, switching to a new treatment, addition of concomitant treatments, and death. Also consider events like hospitalization that might interrupt treatment, etc., if relevant. For each type of intercurrent event, consciously decide on a strategy before you see the data. Ask how each event impacts the interpretation of the outcome. For example, “If the patient switches to another drug, do I still attribute subsequent events to the original drug (treatment-policy), or do I stop follow-up at that point (while-on-treatment), or do I consider a hypothetical scenario where they hadn’t switched?” There isn’t one correct answer for all studies – it depends on the question. But you need to pre-specify it. This prevents bias and p-hacking, and it makes the study reproducible and transparent.
* Align the estimand with the study’s purpose (stakeholder needs): Different stakeholders may care about different effects. Clinicians might want to know the per-protocol effect (what happens if my patient actually takes this drug as intended). Regulators might want the policy effect (what happens in aggregate if this drug is on the market and used with typical adherence). Patients might want to know their personal risk if they adhere to treatment. Usually for safety, the clinical and regulatory interest is in the actual causal effect of the drug, so an estimand that gets closer to that (like while-on-treatment) is favored. But if your study is meant to inform, say, formulary decisions or cost-effectiveness, a more pragmatic estimand might be used. The key is: think from the end-user’s perspective – what question do they need answered? Then choose the estimand that answers it. This prevents situations where you present an analysis that, while statistically correct, doesn’t address the real concern of decision-makers.
* Beware of implicitly unrealistic estimands: As Hernán & Scharfstein warned, don’t inadvertently answer a question that nobody is asking or that is not actionable. For example, an estimand that effectively compares “patients who stay on Drug A no matter what” to “patients who can switch away from Drug B if issues arise” would be a mismatched comparison. Ensure symmetry in comparisons and clinical relevance. If certain intercurrent events are very likely (e.g., many will switch), a pure hypothetical of “nobody switches” might be too far from reality to be useful (unless you’re specifically interested in that scenario). Balance realism with the desire to isolate effects.
* Plan for analysis methods that match the estimand: Once you choose an estimand, make sure your statistical approach follows suit. For a while-on-treatment estimand, you will likely be censoring data – so use methods to handle that (Kaplan-Meier, Cox models with appropriate censoring, maybe inverse probability of censoring weights if censoring is informative). If we have nonproportional hazards (NPH) and censoring, note that the estimand that is estimated by the hazard ratio from Cox regression depends on the censoring distribution, so we present targeted maximum likelihood estimators in step 5 as an alternative. For any estimand, Restricted Mean Survival Time (RMST) may be a useful additional summary measure, which is especially beneficial under non-proportional hazards conditions due to interpretability as an average delay in event occurrence.
* For a treatment-policy estimand, ensure you’re not censoring at treatment changes (but you might need methods to handle treatment switching as a form of confounding if patients switched due to risk – e.g., rank-preserving structural models or treat switch as time-varying exposure in a sensitivity analysis). If using a hypothetical estimand, define the modeling approach to impute or project outcomes (like “we will use a joint model to extrapolate kidney function trajectory as if treatment continued”). Essentially, the estimand tells you what data to use or not use, and the analysis must implement that faithfully.
* Also, consider missing data separate from estimand – per ICH E9, missing data (loss to follow-up, etc.) is not an intercurrent event but a challenge to address analytically. So plan imputation or sensitivity for missing data after defining the estimand.
* Conduct sensitivity analyses for alternative estimands- we highlight this in Step 6. It’s often informative to do your primary analysis under one estimand and a secondary under another to see how conclusions differ. Especially in RWE, where assumptions are strong, demonstrating consistency across estimand strategies can bolster confidence. For example, if you choose while-on-treatment as primary, you might also report an ITT-like analysis. If results converge (e.g., both show elevated risk), great – robust finding. If they diverge, that tells a story (as discussed, perhaps indicating the effect is manifested only while on drug). Similarly, one could try a principal stratum approximation (like per-protocol analysis in those who adhered, acknowledging biases) to see if the effect is larger or smaller in that subset. Consistency across these gives insight into how intercurrent events influence the observed effect. When presenting to regulators, acknowledging these alternate analyses shows you have thoroughly examined the question from multiple angles and understand the impact of your assumptions
* Document and justify your choices: In any protocol or report, explicitly state the estimand (with its components) and why it was chosen. For instance, “We chose a treatment-period estimand because we are interested in the on-treatment causal effect of Drug X on Event Y, and this aligns with how the adverse event would be attributed to the drug clinically.” Also explain how each type of intercurrent event is handled and why that is appropriate. This level of detail will make it easier for others (e.g., a regulatory reviewer) to follow your logic and agree that the analysis is answering the right question. It also makes the study reproducible, as another researcher could apply the same estimand definition to a different dataset and expect to address the same question.
* Involve clinical experts in estimand discussions: Estimand selection is inherently multidisciplinary. Engage clinicians or pharmacovigilance experts who understand the disease, treatment, and real-world patient behavior. They can provide insight on which events are important and how best to define the question. For example, a clinician might say “If a patient’s creatinine starts rising, I’d stop the drug immediately” – which supports a while-on-treatment approach or a composite endpoint including “rise in creatinine leading to stop.” Such input ensures the estimand is grounded in real clinical pathways.
* Keep the estimand simple and focused: Especially for an audience not deeply familiar with the estimand framework, it’s important to communicate the estimand in a clear, concise way. Avoid overly technical language when explaining it to stakeholders – you can use plain language alongside the formal definition. For instance, you might say: “Our estimand is essentially looking at the risk difference in AKI between the two treatments during the time patients are actually on those treatments.” Once that concept is understood, you can layer in the finer points of censoring rules, etc. The goal is everyone (analysts, clinicians, decision-makers) has a shared understanding of what is being estimated. By following these principles, an RWE team can confidently select and implement an estimand that makes their analysis robust, transparent, and fit for purpose. In the case of our SOF vs non-SOF AKI analysis, this approach led us to focus on the treatment-period effect, providing a clear answer to the causal question of interest. Adopting the estimand framework in observational research ultimately improves the quality of evidence we generate, ensuring it answers the questions that matter in a rigorous way.

### Summarizing Step 1

This step highlights the importance of clearly defining the scientific question, causal model (DAG), and causal estimands early in the analysis roadmap. Each decision made here shapes the analytic approach, informs interpretation, and provides transparency regarding the assumptions and their potential impact on results.

#### References

(need to format and fill out from offline doc)

Dang et al. (2023). JCTS. Rufibach (2018). Pharmaceutical Statistics. ICH E9 (https://pmc.ncbi.nlm.nih.gov/articles/PMC8112325/#:~:text=Estimand%20strategy%20Measurement%20of%20interest,32).

# Step 1b — Causal Model

## Study-type & background knowledge

Step 1b in the causal roadmap explicitly defines the statistical model and discusses assumptions necessary for causal identification. Clear articulation of these assumptions is crucial for valid causal inference.

### Target-trial emulation schematic

We emulate a two-arm pragmatic trial that randomises at t = 0 (first DAA dispense) to

A = 1 Sofosbuvir-containing regimen

A = 0 Non-sofosbuvir regimen

and follows individuals for 90 days to the first occurrence of acute kidney injury (AKI) or censoring. The analytic data used here are fully simulated by generate\_hcv\_data() in the DGP.R file, with parameters calibrated to the real HealthVerity HCV cohort. No protected health information is present.

*Roadmap link:* this schematic translates Step 1a’s causal question into an explicit target-trial design

## Specifying the Causal Model (DAG)

Causal inference typically involves representing relationships between variables visually using directed acyclic graphs (DAGs). A DAG helps clarify the assumptions necessary for valid causal inference, particularly:

* **Exchangeability (no unmeasured confounding)**
* **Positivity (treatment assignment possible across covariates)**
* **Consistency (treatment definitions correspond to reality)**

Consider the following simplified DAG illustrating relationships in our AKI study:

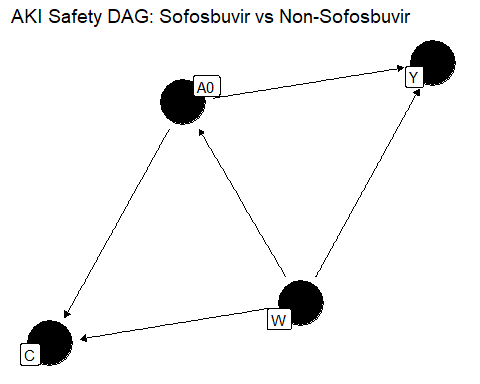
Warning: package 'ggplot2' was built under R version 4.4.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.2 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

Warning: package 'dagitty' was built under R version 4.4.3

Warning: package 'ggdag' was built under R version 4.4.3

Attaching package: 'ggdag'  
  
The following object is masked from 'package:stats':  
  
 filter



**Explanation of DAG components:**

* **W (Baseline Covariates):** Factors like age, sex, baseline kidney function, diabetes, liver cirrhosis, healthcare utilization. These influence both treatment selection (A) and AKI outcome (Y).
* **A0 (Baseline Treatment):** Initial choice between SOF-containing or non-SOF-containing DAAs.
* **Y (Outcome):** AKI event within 90 days after treatment initiation.
* **C (Censoring events):** Events that lead to incomplete follow-up, such as death, regimen switching, or loss of insurance. The DAG explicitly assumes no direct arrows from unknown or unmeasured variables to A and Y (conditional on W), reflecting the key assumption of conditional exchangeability.

The DAG encodes our current subject-matter understanding: baseline factors influence both treatment choice and AKI risk; treatment may induce early kidney events and may also trigger regimen switching or dropout (informative censoring).

### Intercurrent events & time-varying mechanisms

| Intercurrent event | Representation | Roadmap implication |
| --- | --- | --- |
| **Regimen switch** (non-SOF → SOF or vice-versa) | First crossover time enters C(t) | Competes with AKI; requires clear *estimand* choice (treatment-policy vs hypothetical no-switch). |
| **Death** | Cause-specific hazard in C(t) | Competing risk; can be handled via composite outcome or Fine–Gray estimand. |
| **Loss of follow-up / disenrollment** | Administrative censoring | Treated as random given measured W; verify positivity and apply IPC weighting if needed. |

Because the current dataset is *simulated* with baseline treatment assignment only (no programmed switching), time-varying confounding is absent **by design**. This simplification lets us focus on demonstrating the Roadmap mechanics without additional longitudinal complexity. Note the additional simulated case study at the end of this tutorial demonstrating longitudinal TMLE to untangle longitudinal confounding in HIV treatment adherence.

### Identification conditions

1. **Exchangeability:** (Y^{a} !!!A W).
2. **Positivity:** (0 < P(A = a W = w) < 1) for all (w) in support (W).
3. **Consistency:** Observed (Y) equals (Y^{a}) for the treatment actually received.
4. **Correct model for censoring:** (C) independent of counterfactual outcomes conditional on (W, A).

These minimal assumptions will be revisited formally in Step 3 (Identifiability Assessment).

#### Link to simulation design

| Feature in the simulated DGP | Rationale for inclusion | Estimands that depend on it |
| --- | --- | --- |
| **Baseline covariates (W) generated under exchangeability** | Guarantees the causal assumptions for ΔRisk and ΔRMST hold by design, allowing validation of estimator bias. | All estimands |
| **No programmed regimen switching** | Ensures ITT and “no-switch” estimands coincide in simulation so any divergence in real data highlights switching bias. | Diagnostic reference for ITT, hypothetical no-switch, while-on-Tx |
| **Event-time generator calibrated to 2 % 90-day AKI risk** | Provides realistic incidence for variance demonstrations and power calculations. | All estimands (affects standard errors and coverage) |

##### Note on fitness-for-purpose audit

A “fitness-for-purpose” data audit is normally completed at this stage to verify that variable definitions, time-stamps, and measurement reliability align with the causal model. Because we are using fully simulated data whose generating mechanism is known and transparent, such an audit is not applicable. When we later port the workflow to the real HealthVerity cohort, this audit will be mandatory.

# Step 2 – Observed Data & Statistical Model

**Roadmap reminder.** Step 2 documents *exactly* what we observe (either from a simulated data‑generating mechanism or a real‑world database) and shows how those measurements map onto the causal model selected in Step 1. Getting this right is critical: every identification and estimation claim downstream relies on the tuple we define here.

## Objective

*Provide a transparent, reproducible description of the observed‑data vector and the non‑parametric statistical model that contains its distribution.*

## Observed‑data structure

We will work with two nested datasets derived from the simulation code in generate\_hcv\_data() (see *DGP.R* code below for details):

| Dataset | Purpose | Censoring rule |
| --- | --- | --- |
| **Primary (no‑switch)** | Targets the hypothetical *no‑switch* estimand. | Censor at earliest ofevent (AKI), censor\_admin, censor\_switch. |
| **Sensitivity (ITT)** | Targets the treatment‑policy estimand. | Censor at event or censor\_admin; *ignore* censor\_switch. |

Each individual record is

O = (W,A,T,Delta),

with

* $W $ Baseline covariate block generated in Sections 1–2 of the DGP.
* $A $ Binary treatment (treatment) simulated in Section 4.
* $T $ follow\_time = min(event time, admin censor, switch + 30 d).
* $ Delta $ Indicator that event\_time ≤ follow\_time (1 = AKI, 0 = right‑censor).

No parametric form is imposed; the statistical model is the usual fully non‑parametric class $ mathcal M={P(O)} $.

### Visual map

(age, sex, ckd, …) ─┐  
 │ ┌────────────┐ AKI  
Baseline covariates │──▶│ Treatment │──┐ (event, Δ=1)  
 │ │ A (0/1) │ │  
 └──▶└────────────┘ │  
 │ censor at  
 ▼ 30 d post‑switch  
 ┌──────────────────────────┐  
 │ Switch indicator + time │  
 └──────────────────────────┘

The arrows represent the *observed* temporal ordering, **not** causal assumptions (those live in Step 1b).

## Key relationships & gaps

| Relationship in causal model | How observed? | Potential gap |
| --- | --- | --- |
| Baseline confounding ( , ) | 33 binary + 3 continuous covariates. | eGFR (lab) unavailable in claims → unmeasured residual renal risk. |
| Treatment → Switch → Censoring | tx\_days, switch, censor\_switch. | Reasons for switching *unobserved* → informative censoring risk. |
| Treatment → AKI (outcome) | Exact event time in sim; ICD codes in RWD. | Misclassification in RWD; simulation assumes perfect coding. |
| Death as competing risk | Not simulated; present in RWD. | Requires sensitivity check when analysing real cohort. |

**Take‑home:** Exchangeability may fail if baseline eGFR is a strong confounder; independent censoring may fail if declining renal function triggers switching. Both issues inform the identification step.

## Missingness & measurement error

*Simulation*

* optional MCAR missingness via add\_missing (5 % in region, 10 % in ckd).
* optional imputation via **missForest**.

*Real‑world cohort*

* Gaps in enrolment, race, lab results; plan multiple‑imputation or incorporate missing indicators as part of $W $.
* Validate ICD‑based AKI against lab‑based definition in a subset.

## Why Step 2 matters

* **Identifiability preview.** By documenting which arrows in the causal graph are broken by missing data or coarse measurement, we can assess whether exchangeability and positivity are plausible (Step 3).
* **Estimator choice.** Time‑varying censoring indicators imply IPCW or longitudinal TMLE rather than a simple Kaplan–Meier.
* **Reproducibility.** Future readers can regenerate the exact simulated datasets by running generate\_hcv\_data(seed = …) and confirm that the variables align with the narrative.

| id | treatment | follow\_time | event | ckd | cirrhosis |
| --- | --- | --- | --- | --- | --- |
| 1 | 0 | 8.312041 | 0 | 0 | 0 |
| 2 | 0 | 203.737839 | 0 | 0 | 0 |
| 3 | 1 | 37.598959 | 0 | 0 | 0 |
| 4 | 0 | 300.613925 | 0 | 0 | 0 |
| 5 | 1 | 388.130767 | 0 | 0 | 0 |
| 6 | 1 | 86.817428 | 0 | 0 | 0 |

*Table*: First six rows of select variables of the simulated analysis dataset

## ---------------------------------------------------------------  
## Simulate an HCV cohort for SOF vs non-SOF renal-safety analyses  
## ---------------------------------------------------------------  
generate\_hcv\_data <- function(  
 N = 125000, # cohort size ≈ large claims database  
 p\_sof = 0.36, # marginal probability of starting SOF  
 h0 = 3.5e-4, # baseline daily AKI hazard (non-SOF)  
 HR\_early = 1.50, # SOF vs non-SOF HR during first τ days  
 HR\_late = 0.70, # SOF vs non-SOF HR after τ days  
 tau = 90, # change-point (days) for piece-wise HR  
 max\_follow = 180, # admin cut-off (days) – short DAA course  
 risk\_window = 30, # 30-day wash-out after stopping / switch  
 np\_hazard = FALSE, # TRUE → use piece-wise HR\_early / HR\_late  
 dep\_censor = FALSE, # TRUE → informative admin censoring  
 censor\_base = 1/100, # base admin-censoring rate (≈ median 69 d)  
 complexity = FALSE, # TRUE → add non-linear terms + interactions  
 treat\_override = c("simulate","all\_treated","all\_control"),  
 add\_missing = FALSE, # add a little MCAR missingness  
 impute = FALSE, # run missForest on missing vars  
 seed = NULL # for reproducibility  
){  
  
 ## -- housekeeping ----------------------------------------------------  
 if (!is.null(seed)) set.seed(seed)  
 treat\_override <- match.arg(treat\_override)  
 if (impute) requireNamespace("missForest")  
 requireNamespace("tidyverse")  
  
 ## 1 DEMOGRAPHY ------------------------------------------------------  
 # Simulate age, sex and enrolment window to mimic U.S. claims data  
 raw <- tibble(  
 id = seq\_len(N),  
 age = pmax(rnorm(N, 48, 13), 18), # truncated at 18 y  
 sex\_male = rbinom(N, 1, 0.58),  
 race = sample(c("white","black","hispanic","asian","other"),  
 N, TRUE, prob = c(.48,.14,.06,.02,.30)),  
 region = sample(c("NE","MW","S","W"),  
 N, TRUE, prob = c(.20,.18,.37,.25)),  
 enroll\_days = rpois(N, 420) # pre-index continuous enrolment  
 )  
  
 ## 2 CLINICAL HISTORY & MEDICATIONS ---------------------------------  
 # Create binary indicators for comorbidities and nephro-toxic meds  
 add\_bin <- function(p) rbinom(N, 1, p)  
 raw <- raw %>%  
 mutate(  
 # Kidney-specific  
 ckd = add\_bin(.08),  
 prior\_aki = add\_bin(.05),  
  
 # Liver + systemic disease  
 heart\_failure = add\_bin(.07),  
 sepsis = add\_bin(.03),  
 dehydration = add\_bin(.06),  
 obstruction = add\_bin(.04),  
 cirrhosis = add\_bin(.18),  
 portal\_htn = add\_bin(.04),  
 esld = add\_bin(.02),  
 hiv = add\_bin(.04),  
  
 # Metabolic and vascular  
 diabetes = add\_bin(.20),  
 hypertension = add\_bin(.45),  
 bmi = rnorm(N, 28, 5),  
 overweight\_obese = add\_bin(.20),  
 smoking = add\_bin(.40),  
 alcohol = add\_bin(.18),  
 substance\_abuse = add\_bin(.25),  
 cancer = add\_bin(.08),  
 chemo = add\_bin(.01),  
  
 # Out-patient drugs that may modify AKI risk  
 nsaid = add\_bin(.25),  
 acearb = add\_bin(.30),  
 diuretic = add\_bin(.22),  
 aminoglycoside = add\_bin(.05),  
 contrast = add\_bin(.08),  
 statin = add\_bin(.15),  
 aspirin = add\_bin(.10),  
 beta\_blocker = add\_bin(.14),  
 ccb = add\_bin(.16),  
 art = add\_bin(.05), # antiretroviral therapy  
  
 # Prior DAA exposure → exclusion criteria  
 prior\_sof = add\_bin(.05),  
 prior\_nonsof = add\_bin(.05)  
 )  
  
 ## 3 BASELINE EXCLUSIONS --------------------------------------------  
 cohort <- raw %>%  
 filter(enroll\_days >= 365, # ≥1 y continuous enrolment pre-index  
 age >= 18,  
 prior\_aki == 0, # no history of AKI  
 !(prior\_sof == 1 | prior\_nonsof == 1)) # new users only  
  
 ## 4 TREATMENT ASSIGNMENT -------------------------------------------  
 # Simulate (or override) initial SOF vs non-SOF regimen  
 if (treat\_override == "simulate") {  
  
 # Linear predictor capturing confounding between W and A  
 lp0 <- with(cohort,  
 0.015\*age + 0.30\*cirrhosis + 0.25\*ckd + 0.15\*hiv + 0.10\*diabetes -  
 0.10\*cancer + rnorm(nrow(cohort), 0, 0.6)  
 )  
  
 if (complexity) {  
 # Add non-linearities & interactions for SuperLearner challenge  
 lp0 <- lp0 +  
 0.02\*(cohort$bmi^2)/100 -  
 0.3\*sin(0.1\*cohort$bmi) +  
 0.5\*(cohort$age/50)^3 +  
 1.5\*cohort$ckd\*cohort$cancer +  
 0.8\*cohort$hiv\*log1p(cohort$age)  
 }  
  
 # Calibrate intercept so marginal P(SOF)=p\_sof  
 alpha0 <- qlogis(p\_sof) - mean(lp0)  
 p\_trt <- plogis(alpha0 + lp0) |> pmin(0.95) |> pmax(0.05)  
  
 cohort$treatment <- rbinom(nrow(cohort), 1, p\_trt)  
 } else {  
 # Force everyone into one arm (useful for counter-factual sims)  
 cohort$treatment <- ifelse(treat\_override == "all\_treated", 1L, 0L)  
 }  
  
 ## 5 BASELINE HAZARD FOR AKI ----------------------------------------  
 # Build subject-specific baseline hazard multipliers  
 if (!complexity) {  
 lp\_out <- with(cohort,  
 -2.8 + 0.03\*age + 0.7\*ckd + 0.5\*cirrhosis +  
 0.3\*heart\_failure + 0.25\*nsaid + 0.20\*contrast  
 )  
 } else {  
 lp\_out <- with(cohort,  
 -2.8 +  
 0.03\*age + 0.0005\*age^2 +  
 0.7\*ckd + 0.5\*cirrhosis +  
 0.02\*(bmi^2)/100 -  
 0.3\*sin(0.1\*bmi) +  
 0.4\*heart\_failure\*acearb + # drug–disease interaction  
 0.6\*nsaid\*treatment + # effect-modifier for treatment  
 0.3\*contrast\*log1p(age)  
 )  
 }  
 base\_rate <- h0 \* exp(lp\_out) # personalised baseline hazard  
  
 ## 6 AKI EVENT TIMES -------------------------------------------------  
 # Option A: single-rate exponential with HR\_early  
 if (!np\_hazard) {  
 rate <- base\_rate \* ifelse(cohort$treatment == 1, HR\_early, 1)  
 cohort$event\_time <- rexp(nrow(cohort), rate = rate)  
  
 # Option B: piece-wise exponential with HR\_early then HR\_late  
 } else {  
 # helper: random piece-wise exponential sampler  
 rpexp\_piece <- function(n, r1, r2, tau){  
 u <- runif(n); p1 <- 1 - exp(-r1\*tau); t <- numeric(n)  
 e <- u <= p1  
 t[e] <- -log(1 - u[e]) / r1[e] # early events  
 t[!e] <- tau - log((1 - u[!e])/(1 - p1[!e]))/r2[!e] # late events  
 t  
 }  
 r1 <- base\_rate \* ifelse(cohort$treatment == 1, HR\_early, 1)  
 r2 <- base\_rate \* ifelse(cohort$treatment == 1, HR\_late, 1)  
 cohort$event\_time <- rpexp\_piece(nrow(cohort), r1, r2, tau)  
 }  
  
 ## 7 ADMINISTRATIVE & SWITCH CENSORING ------------------------------  
 # Administrative censoring: independent or dependent on covariates  
 if (!dep\_censor) {  
 censor\_admin <- rexp(nrow(cohort), rate = censor\_base)  
 } else {  
 c\_rate <- censor\_base \* exp(0.4\*lp\_out + 0.3\*cohort$treatment)  
 censor\_admin <- rexp(nrow(cohort), rate = c\_rate)  
 }  
 cohort$censor\_admin <- pmin(censor\_admin, max\_follow)  
  
 # Simulate treatment duration & switching (≈ 3% switch rate)  
 cohort$tx\_days <- ifelse(cohort$treatment == 1,  
 rpois(N, 84), # SOF courses often 12 wk  
 rpois(N, 70)) # others slightly shorter  
 cohort$switch <- rbinom(nrow(cohort), 1, 0.03)  
 cohort$censor\_switch <- ifelse(cohort$switch == 1,  
 cohort$tx\_days + risk\_window,  
 max\_follow)  
  
 # Determine observed follow-up time & event indicator  
 cohort$follow\_time <- pmin(cohort$event\_time,  
 cohort$censor\_admin,  
 cohort$censor\_switch)  
 cohort$event <- as.integer(cohort$event\_time <= cohort$follow\_time)  
  
 ## 8 FINAL ANALYSIS DATASET -----------------------------------------  
 ana <- cohort %>%  
 select(-enroll\_days, -prior\_aki, -prior\_sof, -prior\_nonsof,  
 -tx\_days, -event\_time, -censor\_admin, -censor\_switch)  
  
 ## 9 OPTIONAL MISSINGNESS + IMPUTATION ------------------------------  
 if (add\_missing) {  
 # MCAR missingness on two illustrative variables  
 ana$region[sample(nrow(ana), 0.05\*nrow(ana))] <- NA  
 ana$ckd[sample(nrow(ana), 0.10\*nrow(ana))] <- NA  
  
 if (impute) {  
 imp\_vars <- c("age","race","region",  
 "ckd","cirrhosis","hiv",  
 "diabetes","hypertension","bmi")  
 imp\_in <- ana %>%  
 select(all\_of(imp\_vars)) %>%  
 mutate(across(c(race,region), as.factor))  
 ana[, imp\_vars] <- missForest::missForest(as.data.frame(imp\_in),  
 verbose = FALSE)$ximp  
 }  
 }  
  
 return(ana)  
}

# Step 3: Identifiability

## Introduction to *Identification*: Linking the HCV→AKI Causal Question to the Observed Data

Rigorous causal inference from observational data requires translating the scientific questions posed in **Step 1** into statistical parameters that are identifiable from the observed data.

Before we can run any models we must answer two questions:

1. **What exact counterfactual quantity are we after?** (The *causal estimand* defined in step 1)
2. **Can that quantity be expressed with the data we observe?** (= *identifiability*)

This **identification step** formally links the counterfactual (causal) world to the factual world we can measure. Doing so relies on three well-known assumptions—**Consistency, Conditional Exchangeability, and Positivity**—described below in the context of our revised primary and sensitivity estimands.

*Reminder:* Causal estimands expressed in mathematical notation

Primary – *Hypothetical no‑switch* risk at 90 days *“What would the 90‑day cumulative risk of first AKI be if every patient stayed on their initial regimen for the full 90 days?”*

Sensitivity – Treatment‑policy (ITT) risk at 90 days *“What is the 90‑day AKI risk difference if clinicians simply start SOF vs non‑SOF, allowing any subsequent changes?”*

## Identification assumptions

To express those counterfactual risks with our data we need four assumptions.

### Consistency + SUTVA

The outcome we observe equals the potential outcome under the regimen actually received, and one patient’s treatment choice does not influence another patient’s AKI risk.

### Conditional exchangeability (no unmeasured confounding)

All factors that affect both initial regimen choice and AKI risk are captured in .

### Positivity

*Treatment positivity* — every covariate pattern with positive probability has some chance of receiving either regimen:

*Time‑varying positivity* (primary estimand) — no regimen‑and‑history pattern forces an immediate switch/discontinuation over 0–90 d.

Positivity assumptions are both theoretical and practical- i.e, if some patient characteristics mean they will never be prescribed SOF due to contraindications, that is a theoretical violation. If we observe some patients (as defined by combinations of ) never or rarely recieve SOF, that is a practical violation. We will check positivity using propensity score diagnostics in Step 4.

### 4.4 Independent censoring (primary estimand)

After adjusting for observed history the chance of remaining uncensored is unrelated to the counterfactual outcome:

We relax this by modelling the switching/death hazard and using inverse‑probability‑of‑censoring weights (IPCW) inside the TMLE.

## Why these assumptions are plausible (or not) in claims data

* **Consistency / SUTVA:** Each patient’s AKI diagnosis is determined by their own regimen and physiology; spill‑over is unlikely.
* **Exchangeability:** We include rich baseline medical history from the claims data (CKD stage, eGFR, liver disease, etc.). Residual confounding remains possible if, for instance, prescriber preference correlates with unmeasured kidney function.
* **Positivity:** Both SOF and non‑SOF regimens are widely used across all patient profiles in commercial and Medicare claims, but we will check overlap via propensity‑score diagnostics.
* **Independent censoring:** Switching is often triggered by rising creatinine, so naïve censoring would be informative. IPCW mitigates this by modelling the switching hazard as a function of observed labs and comorbidities.

## 7 Take‑home message for pharmacoepidemiologists

The Identification step distils *clinical intent* into a pair of mathematically precise contrasts and spells out exactly what we must believe about the data for those contrasts to be recoverable. Once these assumptions are on the page, colleagues and reviewers can scrutinise them, and the subsequent estimation step (Step 3) flows in a straight line from the formulas above.

*Next up:* we implement doubly‑robust estimation of these risks with continuous‑time TMLE (concrete) and evaluate how sensitive the conclusions are to the assumptions listed here.