

A Plain-Language Rewrite of ICH E9(R1)

Defining the effect you actually want to measure

1. The core problem

A trial exists to answer a question: *what does this treatment do?* But that is vaguer than it sounds. Before you analyze anything, you have to pin down exactly which effect, in which people, on which outcome, accounting for which real-world complications, summarized in which way. Skip this, and two statisticians can analyze the same dataset and report different “treatment effects,” each correct for a different unstated question.

The fix is to write down the effect you intend to estimate *before* the trial — precisely enough that the analysis method follows from it, not the other way around. Call this precisely defined effect the **Target Effect**.

2. The five parts of a Target Effect

A Target Effect is fully specified only when all five of these are nailed down:

1. **Treatments:** which treatment(s) and comparator(s), and the regimens around them.
2. **People:** the patient population the question is about.
3. **Outcome:** the specific measurement that captures the clinical question (the endpoint).
4. **Disruption handling:** how you account for events that occur after treatment starts and muddy the outcome (see §4). The most neglected part, and the heart of this document.
5. **Group summary:** how individual outcomes are combined into one number comparing treatments (difference in means, odds ratio, hazard ratio, etc.).

Change any one of the five and you have changed the question.

3. Disruptions vs. simply-missing data

A **Disruption** is an event that happens *after treatment starts* and either changes the meaning of the outcome or makes it hard to interpret. Examples are starting a rescue medication, stopping the study drug, switching arms, or dying before the outcome is assessed. A Disruption is *not* the same as missing data, and conflating the two is the most common error here:

- A **Disruption [Intercurrent Event or ICE]** is something that happened to the treatment or to the meaning of the outcome. The outcome you want is still conceptually defined — the only question is how the effect should treat the event. A *question-definition* problem.
- **Missing data** is the plain absence of a value you wanted, where that value stays perfectly well defined — e.g., a patient moves away. Nothing about the question changed; you just do not have the number. An *estimation* problem, handled with imputation and similar tools.

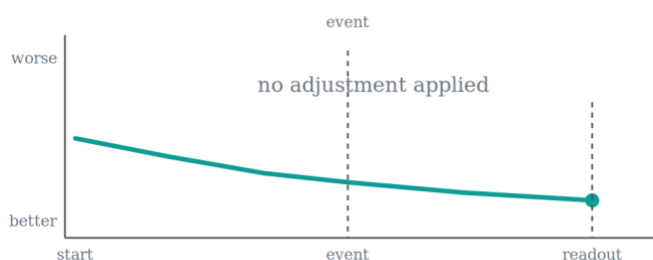
The two stack: a Disruption can later *produce* missing data. Handle them as separate steps — first decide how the Target Effect treats the Disruption, then deal with any data that ends up missing.

4. The five ways to handle a Disruption

Each strategy is a different answer to “given that this event can happen, what effect do we actually want?” None is universally right; the clinical question decides. The pictures below are the **archetypes** — reusable across different events, because the picture goes with the *strategy*, not the event.

(a) Count Regardless [aka Treatment Policy]

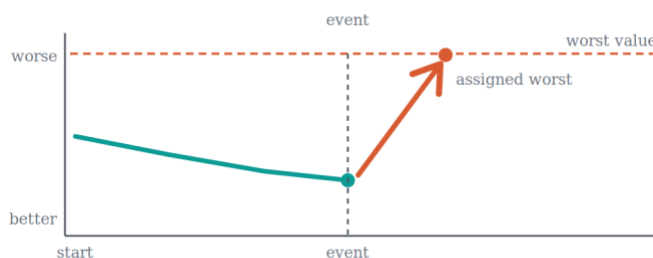
Use the outcome value no matter what — whether or not the Disruption happened. The event is treated as part of real life and ignored in defining the effect. Answers the pragmatic question: what happens when you *start* this treatment, knowing some patients will stop, switch, or supplement? Note that this strategy requires that you actually collect the outcome after the Disruption; therefore this strategy cannot be used when the Disruption makes the outcome impossible to obtain (e.g., death, for a non-survival outcome).



Archetype: used as observed. The line runs through the event untouched.

(b) Bake-In [aka Composite Variable]

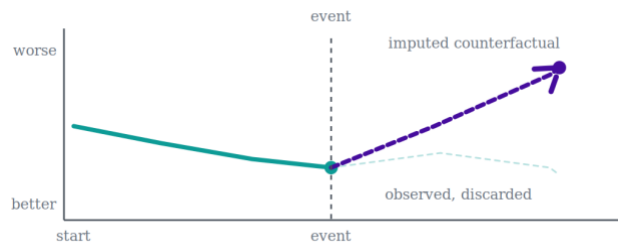
Make the Disruption part of the outcome itself. Rather than measuring the clinical value alone, you define a combined endpoint in which experiencing the event counts as a result — usually a failure. You are not estimating what the value would have been; you are declaring that the event itself is an outcome. The high value is a *definition*, not an estimate — which is why the archetype is a clean, hard arrow to a fixed value.



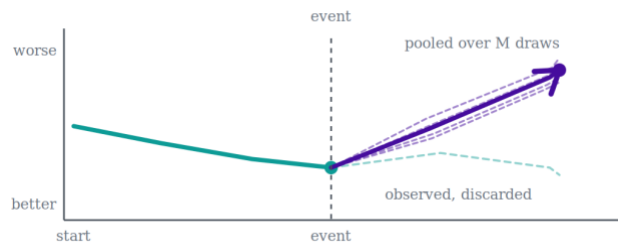
Archetype: definitional floor. The value is assigned to “worst” by fiat.

(c) What-If [aka Hypothetical]

Target the value the outcome *would* have taken in an imagined world where the Disruption never occurred. You specify the hypothetical condition (e.g., “rescue had not been available”), then estimate what would have happened. Unlike Bake-In, you are not scoring the event — you are reconstructing a counterfactual value, which carries uncertainty. A single deterministic reconstruction looks like the first picture below; multiple imputation looks like the second — the scatter of draws is the visual signature of “estimated, not declared.”



Archetype: single imputed path. Observed tail discarded, one counterfactual drawn. A single deterministic reconstruction here means a model-based conditional-mean estimate had rescue not been available, not a carry-forward (LOCF/BOCF/WOCF).

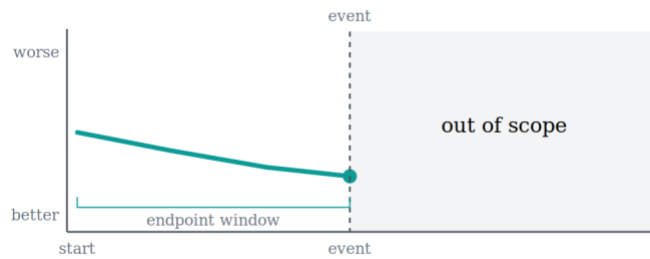


Archetype: multiple imputation. A fan of draws (uncertainty) with a pooled estimate.

Caveat for reference-based imputation: the spread of draws here depicts between-imputation variability as intuition only. Under jump-to-reference, the pooled variance can be artificially narrow relative to what the fan suggests — do not read quantitative uncertainty off the picture.

(d) Until-Disruption [aka While On Treatment]

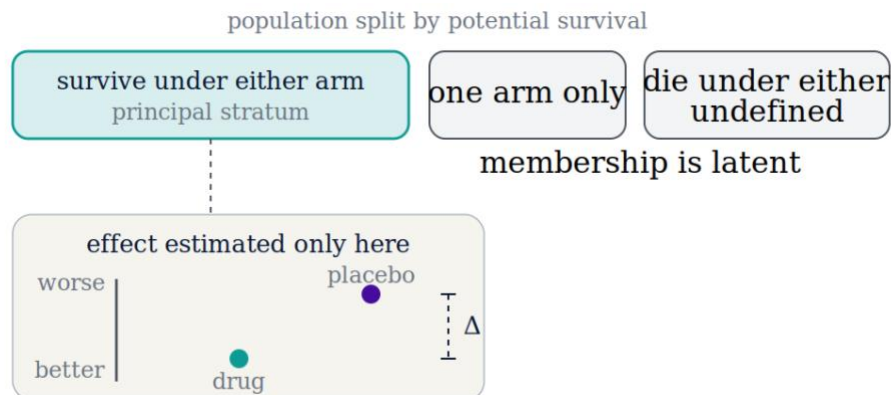
Define the outcome only over the period *before* the Disruption. The question is the effect while the patient is actually on treatment, accepting that different patients contribute different lengths of time. You do not recover or imagine the post-event period — it is simply outside the question that the estimand is answering. Natural for cumulative or longitudinal outcomes, not for a single fixed-timepoint snapshot.



Archetype: truncation. The endpoint window ends at the event; the rest is out of scope.

(e) Same-Type Subgroup [aka Principal Stratum]

Narrow the question to a specific *type* of patient: those who would (or would not) experience the Disruption regardless of which treatment they were assigned. Because that membership is a fixed trait independent of the assignment actually received, comparing treatments within that type is a fair, like-with-like comparison — unlike naively subsetting to whoever actually avoided the event, which breaks randomization. The catch: membership is latent (each patient is assigned once), so identification rests on assumptions plus sensitivity analysis. This is the one strategy that is *not* about a trajectory over time — so its picture deliberately breaks format.



Archetype: stratum subgroup. Carve out the latent type; compare only inside it.

5. Worked examples: same event, different strategy

The same Disruption appears under several strategies — only the column-2 choice changes. The *How it is implemented* column is tagged: **SINGLE** = one procedure; **PICK ONE** = true alternatives (the selector is noted); **ALL STEPS** = a sequence run in order. The final column shows a thumbnail of the row's event-specific figure (full size in §7). Study withdrawal is deliberately absent — it is missing data, not a Disruption.

Disruption	Strategy & why	How it is implemented	Figure
Rescue medication	Count Regardless real-world effect of starting the drug; rescue is part of the lived regimen.	SINGLE Follow on schedule; use the post-rescue value as observed.	
Rescue medication	Bake-In needing rescue is itself failure, so it goes inside the endpoint.	PICK ONE <i>by endpoint type</i> : - Continuous: assign worst observed change at first rescue (rank-based). - Binary: responder = met threshold AND no rescue.	
Rescue medication	What-If effect in a world where rescue was not available.	ALL STEPS - Censor from first rescue onward. - Impute counterfactual (MAR vs. reference-based). - Analyze the reconstructed no-rescue outcome.	
Rescue medication	Until-Disruption only the response over the window before rescue matters.	PICK ONE <i>by endpoint shape</i> : - Snapshot: symptom change from baseline at first rescue. - Cumulative: on-treatment average/AUC truncated at rescue.	
Discontinuation (AE)	Count Regardless effect of the regimen, accepting some stop for tolerability.	SINGLE Continue assessments after the drug stops; use final outcomes as observed.	

Disruption	Strategy & why	How it is implemented	Figure
Discontinuation (AE)	What-If effect had the patient been able to stay on drug.	ALL STEPS 1. Censor at discontinuation. 2. Impute an on-treatment-consistent trajectory under MAR. 3. Analyze.	
Discontinuation (AE)	Until-Disruption benefit only while patients tolerate and stay on drug.	PICK ONE by endpoint shape: • Snapshot: endpoint at time of discontinuation. • Cumulative: summarize over the on-treatment window.	
Death before assessment	Bake-In death is the worst result; bake it into the endpoint.	PICK ONE by endpoint type: • Continuous: rank values, place deaths below worst observed. • Categorical: composite where death = failure.	
Death before assessment	Same-Type Subgroup effect among patients who would survive under either arm — the only group for whom the outcome is defined.	ALL STEPS 1. Define the latent survivor stratum. 2. Adopt an identifying assumption (monotonicity vs. principal-score). 3. Estimate the within-stratum effect. 4. Sensitivity: analyze the untestable assumption.	
Death before assessment	Until-Disruption symptom benefit during the time alive and on drug.	SINGLE Define the endpoint over the period before death (e.g., mean score while alive).	
Death before assessment	Count Regardless — Not Applicable	N/A No value exists after death, so this strategy cannot actually be used here.	
Switch to other drug	Count Regardless real-world effect of the initial assignment, switching included.	SINGLE Keep ITT; use final outcomes regardless of switch.	

Disruption	Strategy & why	How it is implemented	Figure
Switch to other drug	What-If effect had switching not been possible.	ALL STEPS 1. Censor at switch. 2. Reconstruct the no-switch outcome (RPSFTM vs. IPCW). 3. Analyze.	

6. The workflow

The pieces are planned in order, before data arrive:

Trial objective → Target Effect (all five parts) → estimation method → pre-specified Stress Tests

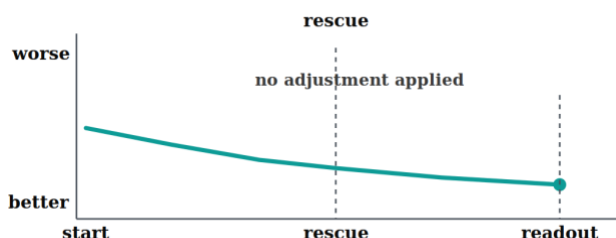
Perform Stress Tests and Sidea Analyses in addition to the primary analysis.

A **Stress Test** keeps the same Target Effect and asks whether the conclusion holds if the untestable assumptions are wrong in plausible ways. A **Side Analysis** is broader or exploratory — often a different Target Effect — and is not a robustness check. All of it belongs in the protocol and SAP up front. The point of the framework is that the question drives the analysis, not the reverse.

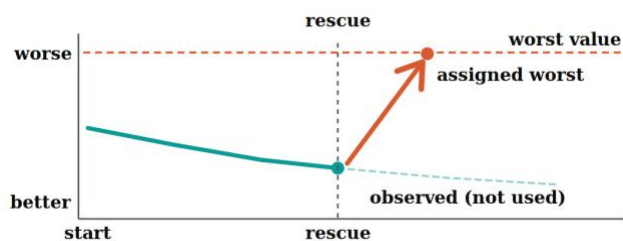
7. Figures By event

The table above uses one archetype picture per strategy. Here are the same pictures rendered for each specific event — note how the archetype stays constant while the event label and a few event-specific details change. That constancy is the point: the picture goes with the *strategy*, not the event.

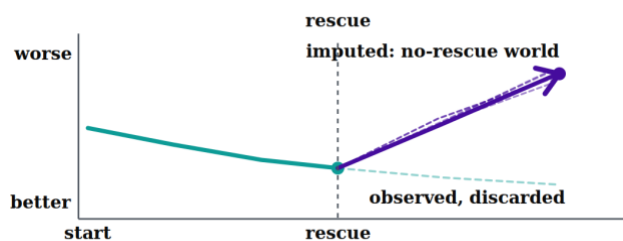
Rescue medication



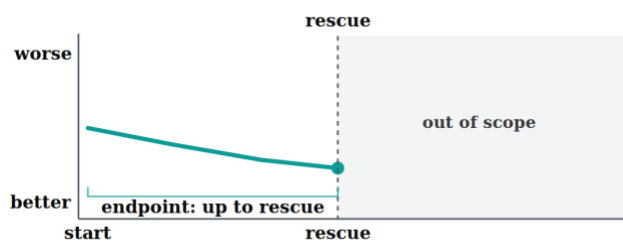
Count Regardless — the line runs through rescue untouched.



Bake-In — value assigned worst at rescue; the real later outcome is collected but not used.

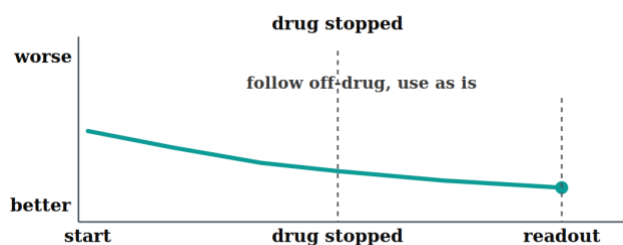


What-If — observed tail discarded; imputed toward a worse no-rescue world (fan = MI).

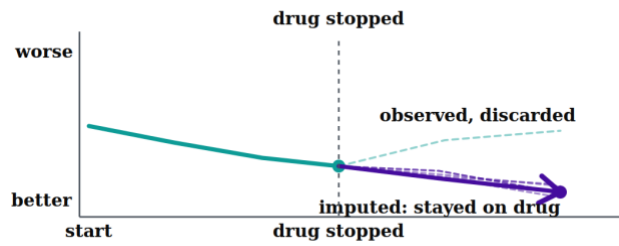


Until-Disruption — endpoint defined only up to rescue.

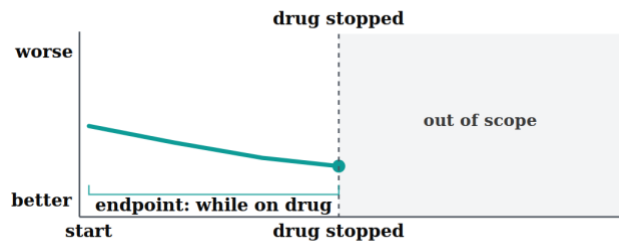
Treatment discontinuation (AE)



Count Regardless — follow off-drug, use values as observed.

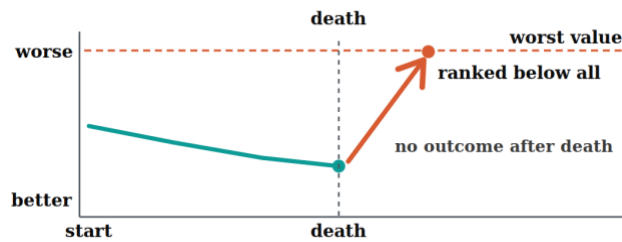


What-If — imputed as if the patient had stayed on drug (keeps improving); observed drift discarded.

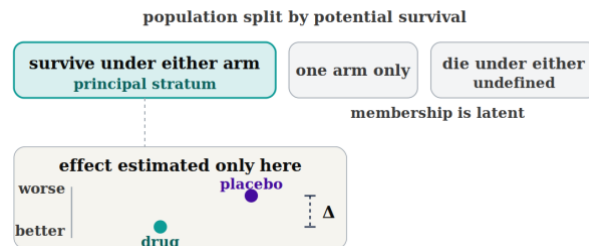


Until-Disruption — endpoint defined only while on drug.

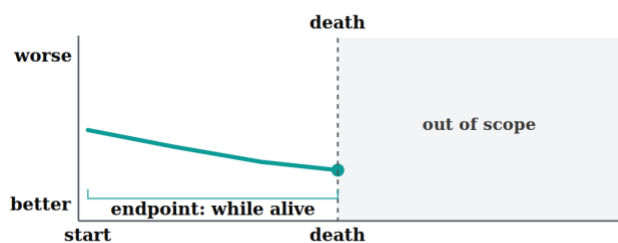
Death before assessment



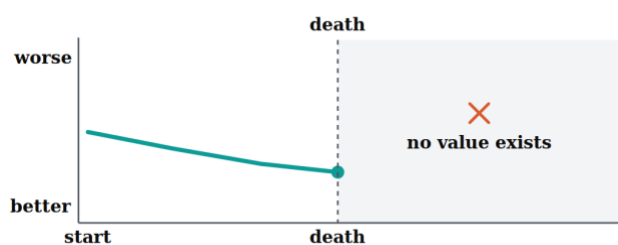
Bake-In — ranked below all observed; no tail, because no outcome exists after death.



Same-Type Subgroup — the template-breaker: compare only within the survive-under-either stratum.

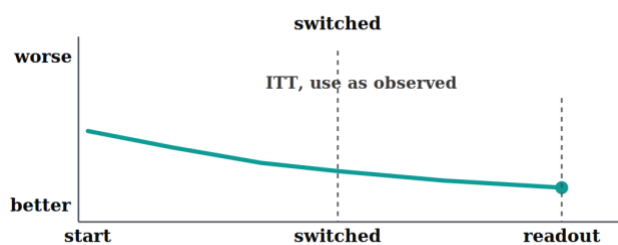


Until-Disruption — symptom score while alive on treatment; the rest is out of scope.

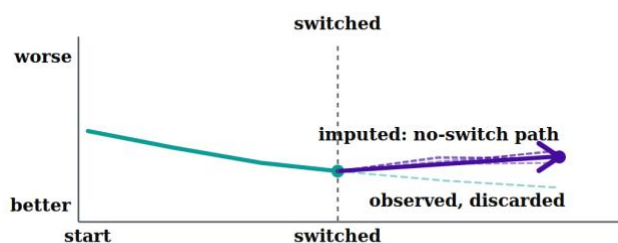


Count Regardless — not usable: no value exists after death to use as observed.

Switch to other drug



Count Regardless — keep ITT; use final outcomes regardless of switch.



What-If — imputed along the no-switch path; observed post-switch values discarded.

Legend: new term → ICH E9(R1) term

This document	ICH E9(R1)
Target Effect	Estimand
Disruption	Intercurrent event
Group summary	Population-level summary
Count Regardless	Treatment policy strategy
Bake-In	Composite variable strategy
What-If	Hypothetical strategy
Same-Type Subgroup	Principal stratum strategy
Until-Disruption	While on treatment strategy
Stress Test	Sensitivity analysis
Side Analysis	Supplementary analysis

Appendix

Below is an example definition of an Estimand, with comments and critiques.

Anatomy of a Real Estimand

● Critical error ● Reword ● Good — keep □ Alternative not chosen

