Paper AS08

Invisible Influences: Estimands, Assumptions and Sensitivity in Inflammation & Immunologic Trial Interpretation

Author: Sukanya ES, Pfizer, Chennai, India

Abstract

Clinical trials in inflammation and immunology face unique challenges such as fluctuating disease activity, delayed treatment onset, immunogenicity, and high rates of rescue medication use that complicate endpoint interpretation. This paper explains how the ICH E9(R1) estimand framework defines precise treatment effects by specifying population, variable, intercurrent-event strategies (e.g., treatment-policy vs. hypothetical for rescue use), and summary measures. We highlight sensitivity techniques tailored to immunology trials, including reference-based imputation (e.g., jump-to-reference for early discontinuation), delta-adjusted models to simulate immunogenicity impact, and tipping-point analyses to assess robustness under missing-not-at-random assumptions. Special attention is given to identifying estimands that reflect real-world effectiveness versus idealized efficacy in diseases like atopic dermatitis, sJIA (Systemic Juvenile Idiopathic Arthritis) & Alopecia Areata. By integrating estimands with targeted sensitivity analyses, we expose how hidden assumptions can distort conclusions and how thoughtful design can mitigate them. Real-world examples illustrate how this dual framework enhances transparency, regulatory alignment, and interpretability in immune-mediated trials.

Introduction

Inflammation and immunologic indications pose distinctive statistical challenges: heterogeneous trajectories, flare–remission patterns, variable onset of action, immunogenicity that may attenuate response, and common use of rescue therapies or treatment switching. Before the advent of the estimand concept, analyses often blended scientific questions with ad hoc handling of intercurrent events and missing data. ICH E9(R1) establishes a structured path from trial objective to analysis by defining the treatment effect of interest via four attributes: population, variable, intercurrent-event strategy, and summary measure.

Major Inflammation & Immunology Trials in the Industry

The I&I space is one of the fastest-growing therapeutic areas in pharma, driven by autoimmune diseases, chronic inflammatory disorders, and advanced immune-modulating technologies. Some of the diseases are:

- Rheumatoid Arthritis (RA), Psoriatic Arthritis, Ankylosing Spondylitis
- Systemic Lupus Erythematosus
- Atopic Dermatitis (Eczema) & Allergic Disorders
- Psoriasis & Dermatologic Immune Disorders

Key Challenges in I&I Trials Before the Estimand Framework

1. **Unclear Treatment Effect**
   Trial objectives were vague (e.g., “Does the drug work?”) without specifying conditions like rescue medication use, population definition, or handling of missing data. This led to inconsistent interpretations of results.

2. **No Standard for Intercurrent Events**
   Events such as rescue medication, treatment discontinuation, or switching therapies were handled inconsistently. Analysts used ad hoc methods, creating uncertainty about whether results reflected true drug efficacy or real-world effectiveness.

3. **Inconsistent Missing Data Handling**
   High dropout rates were common, but methods like LOCF or multiple imputation were applied without clear rationale. Different assumptions across analyses reduced transparency and comparability.

4. **Opaque Sensitivity Analyses**
   Sensitivity analyses were performed mainly to satisfy regulators, with little explanation of assumptions or objectives. Post hoc methods further disconnected analyses from the main scientific question.

5. **Poor Communication of Results**
   Reports stated outcomes without clarifying assumptions (e.g., rescue medication allowed or not). This caused misunderstandings among clinicians, statisticians, and regulators, reducing confidence in conclusions.

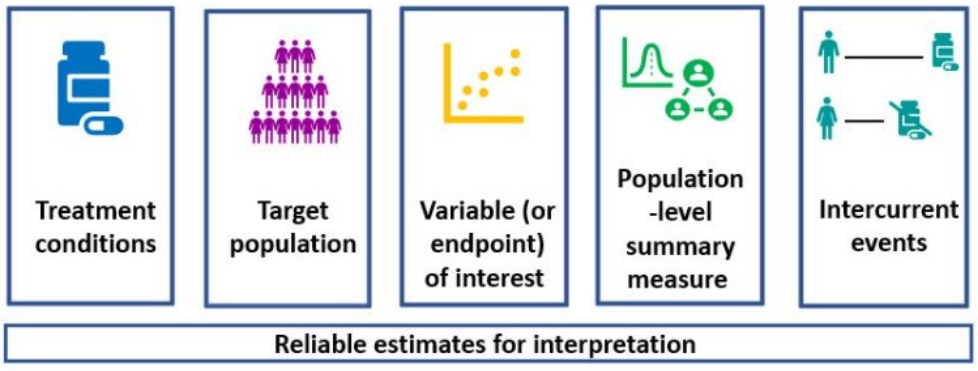
THE ESTIMAND FRAMEWORK (ICH E9 R1)

An Estimand framework is a structured approach that aligns trial objectives, trial design, analysis, and interpretation by precisely defining what treatment effect a study aims to estimate. The estimand framework ensures alignment across **planning, design, conduct, analysis, and interpretation** of clinical trials. Before the estimand framework was introduced I&I trials often had ambiguity about:

- How intercurrent events were handled?
- What treatment effect was truly being estimated?
- How the analysis aligned with clinical objectives?

ICH E9(R1) defines an **estimand** as:

A precise description of the treatment effect reflecting the clinical question posed by the trial objective.



An estimand has **five attributes**:

Estimand Attribute	Explanation
1. Population	The patient population targeted by the trial objective.
2. Treatment condition(s)	The intervention(s) and comparator(s).
3. Variable (endpoint)	The clinical outcome being measured.
4. Intercurrent events (ICEs)	Events after treatment initiation affecting interpretation or existence of outcome.
5. Population-level summary measure	How the treatment effect will be summarized (e.g., mean difference, hazard ratio).

Intercurrent Events (ICEs) in I&I Trials: sJIA, Atopic Dermatitis, Alopecia Areata

Intercurrent Events (ICEs) are events occurring after treatment initiation that affect the interpretation or existence of trial outcomes (ICH E9(R1)). They are extremely common in I&I diseases because these conditions fluctuate, involve rescue therapies, and often use background treatments. Below is a structured, disease-specific overview.

ICEs in sJIA (Systemic Juvenile Idiopathic Arthritis):

Intercurrent Event (ICE)	Definition / Context	Estimand Strategy	Regulatory-Style Interpretation
sJIA Disease Flare	Recurrence of fever + worsening in core JIA variables (e.g., joint count, physician/parent assessment, inflammatory markers).	Composite	Flare is often the primary endpoint in withdrawal designs; flare occurrence = treatment failure.
Discontinuation due to Sustained Inactive Disease	Patients stop treatment after achieving prolonged inactive disease (e.g., ≥24 weeks).	While-on-treatment	Not counted as treatment failure; data included until discontinuation only.
Discontinuation for Other Reasons	Stopping treatment for reasons unrelated to disease response such as adverse events, lack of efficacy, or patient/parent decision.	Composite	Typically treated as flare/treatment failure.
Use of Rescue Medication	Need for additional therapy to regain disease control such as corticosteroids, dose escalation, or extra immunomodulators.	Composite	Usually defines treatment failure or flare; may require censoring in some estimands.
Missing Visits / Operational Disruptions	Inability to collect scheduled assessments (e.g., missed visits, site closures).	Hypothetical	Data imputed as if the disruption had not occurred.
Transition to Extension or Post-Trial Phase	Moving from randomized to open-label or long-term follow-up.	While-on-treatment	Only randomized-phase data contribute; post-extension data excluded.

ICEs in Atopic dermatitis studies:

Intercurrent Event (ICE)	Definition / Context	Estimand Strategy	Regulatory-Style Interpretation
Use of Prohibited or Non-Allowed Medications	Use of disallowed topical or systemic therapies that may alter efficacy	Treatment-policy or Composite	If allowed under treatment-policy, data remain; if composite, use triggers

Intercurrent Event (ICE)	Definition / Context	Estimand Strategy	Regulatory-Style Interpretation
	assessments.		treatment failure.
Rescue Medication Use	Need for additional medication to control symptoms (e.g., corticosteroids, calcineurin inhibitors).	Composite	Considered treatment failure or marks endpoint event depending on protocol.
Treatment Discontinuation Due to Lack of Efficacy	Patient stops randomized treatment because AD is not improving.	Composite	Typically treated as non-response/treatment failure per regulatory expectations.
Treatment Discontinuation Due to Adverse Events	Patient discontinues due to safety/tolerability issues.	Principal Stratum or Composite	Can be evaluated under principal-stratum estimand or counted as treatment failure depending on study objective.
Use of Forbidden Concomitant Skin Care Products	Use of non-approved moisturizers or cosmetic products that may influence severity scores.	Treatment-policy or Hypothetical	Treatment-policy keeps data; hypothetical imputes assuming product was not used.
Non-Adherence / Missed Doses	Skipped or inconsistent treatment administration.	Treatment-policy	Real-world adherence is incorporated as part of the effect of treatment.
Missing Diary or PRO Data	Missing daily symptom scores (e.g., itch NRS, sleep disturbance).	Hypothetical	Missing values imputed under assumptions as if the data were collected.
Operational Disruptions (e.g., site closures)	Missed visits or assessments due to external operational issues.	Hypothetical	Data imputed assuming disruption did not occur.
Switch to Open-Label Extension	Patient enters extension phase before final endpoint assessment.	While-on-treatment	Only randomized-phase data contribute; post-extension data excluded.

ICEs in Alopecia Areata trials:

Intercurrent Event (ICE)	Definition / Context	Estimand Strategy	Regulatory-Style Interpretation
Use of Rescue Therapy	Introduction of additional treatment intended to promote hair regrowth (e.g., topical corticosteroids, minoxidil, intralesional steroids).	Composite	Rescue therapy typically indicates inadequate response and is treated as treatment failure or marks an endpoint event.
Use of Prohibited Medications	Use of non-allowed systemic immunomodulators, JAK inhibitors, biologics, or other agents that may influence hair growth.	Treatment-policy or Composite	Under treatment-policy, data remain; under composite, violation indicates failure.
Treatment Discontinuation Due to Lack of Efficacy	Participant stops assigned treatment because meaningful regrowth is not achieved.	Composite	Considered non-response/treatment failure per regulatory expectations.
Treatment Discontinuation Due to Adverse Events	Discontinuation due to tolerability/safety concerns.	Principal-stratum or Composite	Can define a sub-population (principal stratum) or can count as failure; depends on estimand.
Cosmetic Hair Procedures	Haircut, shaving, artificial hair fibers, dye, or camouflage that may interfere with scalp assessment (SALT).	Hypothetical or Treatment-policy	Hypothetical imputes data assuming no cosmetic interference; treatment-policy retains observed data.

Intercurrent Event (ICE)	Definition / Context	Estimand Strategy	Regulatory-Style Interpretation
Missed Assessments / Missing SALT or PRO Scores	Missing visits or missing Severity of Alopecia Tool (SALT) scores / patient-reported outcomes.	Hypothetical	Values imputed assuming data were collected per protocol.
Operational Disruptions	Missed visits due to non-medical logistical reasons (e.g., site closures).	Hypothetical	Imputed as if the disruption had not occurred; does not count as failure.
Early Switch to Open-Label or Extension Phase	Patients move to an extension arm before completion of randomized evaluation.	While-on-treatment	Only randomized-period data included in primary analyses; post-switch excluded.
Non-Adherence	Participants do not use investigational products as instructed (missed doses, inconsistent use).	Treatment-policy	Real-world adherence becomes part of the treatment effect; data remain in analysis.

Sensitivity Analyses Tailored to Immunology Trials

Sensitivity analyses probe robustness of conclusions to deviations from primary assumptions (e.g., missing at random). They should be aligned with the primary estimand and pre-specified in the SAP. Three families are particularly relevant in immunology:

Reference-Based Imputation (RBI)- Making Post-Discontinuation Behaviour Realistic:

RBI handles post-discontinuation missing data by referencing the control arm trajectory. In immunology studies, patients who discontinue early whether due to **lack of efficacy**, **adverse events**, or the need for **rescue medication** rarely continue behaving like their original assigned arm. Traditional MAR-based methods (MMRM/MI) unrealistically assume that missing data “look like” those who stayed on treatment.

RBI allows us to explicitly model **what happens after discontinuation**, by imputing missing values using observed data from a **reference group** (commonly the control arm).

In Jump-to-Reference (J2R), outcomes after discontinuation for the active arm are imputed to follow the control arm distribution, yielding conservative estimates for de facto effectiveness. Variants include Copy-Reference & Copy-Increment-from-Reference.

Some of the RBI types implemented in I&I trials:

RBI Type	Description	Typical ICE(s) Imputed in I&I Trials
Jump-to-Reference (J2R)	After ICE, trajectory follows control arm immediately	Treatment discontinuation; rescue; loss-of-efficacy withdrawal
Copy-Reference (CR)	All future values replaced by reference arm values	Rescue initiation; treatment switching; protocol-mandated escape
Copy-Increment-from-Reference (CIR)	Pre-ICE preserved; post-ICE increments from control	Rescue; drug interruptions; dose modifications
CR + MAR Hybrid	Pre-ICE MAR; post-ICE CR	Mixed ICEs such as discontinuation + rescue
Jump-to-Mean (J2M)	Jump to mean of reference distribution	Administrative ICEs; COVID-related missingness

RBI Type	Description	Typical ICE(s) Imputed in I&I Trials
Jump-to-Last-Mean (J2LM)	Jump to last mean of reference timepoint	Minor deviations; late rescue
Extended J2R (eJ2R)	Partial rebound toward reference	Biologic washout; long gaps in treatment
CR + Delta	CR plus negative delta shift	ADA-driven dropout; severe loss-of-efficacy events
J2R + Delta	J2R followed by delta penalty	Strong MNAR assumptions; non-adherent withdrawal
CIR + Delta	CIR with additional delta worsening	Rescue + worsening expectations; ADA+ groups

Delta-Adjusted Models

Delta-adjustment applies a prespecified shift to imputed values to reflect plausible MNAR mechanisms such as immunogenicity or adherence issues. For example, anti-drug antibodies might reduce expected benefit; a negative delta can be applied to post-dropout imputations for antibody-positive patients to assess impact on primary conclusions.

Types of Delta Adjustment Models:

Delta Type	Description	Typical ICE(s) Imputed in I&I Trials
Constant Delta	Same Δ for all post-ICE values	Treatment discontinuation; mild AE dropout
Tipping-Point Δ	Range of Δ values tested	All ICEs causing missing data (discontinuation; rescue)
Group-Specific Δ	Δ varies by subgroup	ADA+; rescue; non-adherent; lack of efficacy (LOE)
Time-Varying Δ	Δ increases over time	Biologic discontinuation; long drug-interruption gaps
Pattern-Mixture Δ	Δ applied by ICE pattern	AE dropout; LOE dropout; rescue; protocol deviation
Reference-Based + Δ	Δ layered on J2R/CR/CIR	Rescue initiation; discontinuation; treatment switching
Bayesian Δ	Δ has prior distribution	ADA-driven dropout; adherence-related discontinuation
Effect-Size Δ	Δ shifts treatment effect	Meta-analytic sensitivity analyses

Tipping-Point Analyses

Tipping-point analyses vary assumptions systematically (e.g., a grid of delta values) until the treatment effect loses statistical significance or clinical relevance. This maps the robustness region and provides an interpretable threshold for MNAR departures that overturn conclusions.

Disease-Focused Estimands and Sensitivity Setups

Atopic Dermatitis: Primary estimand may target EASI-75 at Week xx under a treatment-policy strategy permitting rescue topical corticosteroids; sensitivity analyses include J2R for early discontinuation and MNAR delta-adjustments for high pruritus-driven dropout.

SJIA: Hypothetical estimand for steroid-sparing ACR response without rescue, contrasted with treatment-policy estimand reflecting real-world practice; RBI/J2R for discontinuations and tipping-point around flare-associated missingness.

Alopecia Areata: Estimand targeting SALT ≤20 at Week xx with explicit handling of discontinuations; delta-adjustments for immunogenicity and adherence, plus tipping-point to quantify sensitivity to MNAR dropouts.

Why can endpoints alone mislead in I&I trials?

Endpoints (e.g., EASI-75 in atopic dermatitis, SALT≤20 in alopecia areata, flare/no-flare in sJIA) quantify outcomes only when they are observable. In real I&I studies, however, intercurrent events (ICEs) frequently intervene before the scheduled outcome assessment e.g., rescue medication, early discontinuation due to adverse events or lack of efficacy, flares prompting withdrawal, immunogenicity (ADA)-driven loss of effect, operational disruptions. If you analyse endpoints without specifying how ICEs are handled, you implicitly assume a data-generation story (often MAR - Missing At Random), which may be false and therefore biasing. The ICH E9(R1) addendum was written precisely to prevent such conceptual mismatches by requiring you to define the treatment effect in the presence of ICEs (the estimand), and to stress-test your inferences with sensitivity analyses that explore plausible MNAR departures from the primary assumptions.

For inflammatory/dermatologic trials with diary-heavy outcomes and frequent deviations, best-practice sensitivity methods include controlled multiple imputation (delta-adjusted and reference-based MI) and tipping-point analyses to quantify how much MNAR bias would be needed to overturn conclusions.

Conclusion

Inflammation and immunology trials sit at the intersection of biological complexity and methodological fragility. Traditional endpoint-focused interpretations often obscure the hidden assumptions that shape the evidence we generate assumptions about how intercurrent events unfold, why data go missing, and how participants behave once disease flares, rescue medication is triggered, or immunogenicity erodes treatment response. As this paper demonstrates, estimands make these invisible influences visible by explicitly defining the clinical question in the presence of real-world disruptions. They shift the conversation from *"**What was observed?**"* to ***"What exactly are we estimating, for whom, and under what assumptions?"***

But defining an estimand alone is not enough. Sensitivity analyses complete the picture, exposing how dependent conclusions are on unverifiable assumptions. Through reference-based imputation, delta-adjusted MNAR scenarios, and tipping-point assessments, we gain the ability to test whether our findings remain credible when confronted with the biological and behavioural realities of I&I conditions: flares, rapid loss of effect, ADA-driven discontinuation, symptom-dependent missingness, and protocol-directed rescue therapies. This dual framework clear estimands plus targeted, disease-appropriate sensitivity analyses transforms trial interpretation from a fragile, assumption-laden exercise into a transparent process where robustness can be demonstrated rather than inferred.

Ultimately, the integration of estimands and sensitivity methods does more than strengthen statistical validity. It enhances scientific honesty, aligns interpretations with regulatory expectations, and deepens clinical relevance by ensuring that results reflect the therapeutic journey of real patients, not idealized ones. In a field where disease trajectories are volatile and treatment pathways nonlinear, such clarity is not optional, it is essential. By illuminating the assumptions that shape our conclusions, we foster a more reliable, reproducible, and trustworthy evidence base for decision-making in inflammation and immunologic therapeutics.

References

ICH E9(R1) Addendum: Estimands and Sensitivity Analysis in Clinical Trials (FDA Guidance, 2021).

ICH E9(R1) Training Materials (ICH, 2021).

PHUSE Paper Guidelines (US Connect 2022/2023).

Bartlett JW. Reference-based multiple imputation – variance considerations (arXiv, 2021).

White IR et al. A causal modelling framework for reference-based imputation and tipping-point analysis (J Biopharm Stat, 2019).

Microsoft. (2026). Microsoft Copilot (Version 19.2601.43101.0) [Large language model] https://copilot.microsoft.com

Acknowledgments

Thanks to colleagues and reviewers within the Pfizer & PHUSE community for methodological discussions on estimands and sensitivity analyses.

Contact Information

Contact the author at:
Author Name: E S Sukanya
Company: Pfizer
Email: Sukanya.ErramachettySubramanyam@pfizer.com