⬇ Download PDF **GitHub repository**

Illustrative within-host modeling study / hypothesis generation — not validated experimental or clinical findings, not a cure. Developed with AI assistance (disclosed).

Sequencing antagonism between anti-CD3 and antigen-specific tolerance in type 1 diabetes: a within-host modeling study

Author: Seth Cope[1] *([1] Independent researcher. ORCID: 0009-0000-5520-915X)*

Preprint draft — illustrative modeling study / hypothesis generation. Not validated experimental or clinical findings. Convert to PDF (e.g. `pandoc MANUSCRIPT.md -o manuscript.pdf` *) for posting.*

Abstract

Background. Two disease-modifying immunotherapies are advancing in early (pre-symptomatic) type 1 diabetes (T1D): anti-CD3 monoclonal antibody (teplizumab-class), which delays clinical onset, and antigen-specific tolerance therapy (peptide / mRNA / tolerogenic constructs), which re-educates the autoimmune response. Both are being pushed toward combination use, so *how to combine them* is a live question. Foster et al. (2025, NOD mice) reported an unexplained result — anti-CD3 *reduces* the efficacy of antigen-specific immunotherapy — that no mechanistic model addresses. **Methods.** We build a three-state within-host ODE in which autoreactive effector T cells (E) and antigen-specific regulatory T cells (R) form a mutual-repression **bistable switch** that drives β-cell mass (B). Antigen-specific tolerance *converts* effectors into Tregs (a flux that requires effectors to be present); anti-CD3 is lymphodepleting (it removes effectors and partly depletes Tregs). We score a cohort durable-control fraction and sweep the order and inter-drug interval of the two therapies. **Results.** The model reproduces the Foster antagonism and resolves it into a **sequencing rule**. Tolerance monotherapy controls 100% of the cohort; giving anti-CD3 simultaneously drops this to 59%, and anti-CD3-*first* to 41%, whereas tolerance-*first* fully rescues (100%). The antagonism arises because anti-CD3 removes the effector substrate that tolerance must convert into protective Tregs. The optimal protocol is tolerance-first or, if anti-CD3 must precede, a sufficient inter-drug gap (anti-CD3-first recovers 59%→100% as the gap grows to ~1.5 yr). The rule is robust: tolerance-first ≥ simultaneous in 171/171 viable parameter sets in the Treg-sparing regime, never worse. A second, falsifiable sub-prediction: the optimal order **inverts** if anti-CD3 strongly depletes Tregs. **Conclusions.** We predict that combining anti-CD3 with

antigen-specific tolerance should be done **tolerance-first**, testable directly in a NOD sequencing experiment. Illustrative modeling hypothesis, conditional on the stated assumptions.

---

1. Introduction

Type 1 diabetes results from autoimmune destruction of insulin-producing β-cells, with a long pre-symptomatic prodrome (stage 1–2) during which intervention can delay or prevent clinical (stage 3) disease. Anti-CD3 monoclonal antibody (teplizumab) delayed progression from stage 2 to stage 3 by a median of roughly two years in the TN10 trial — the first disease-modifying therapy approved for this window. A complementary strategy, antigen-specific tolerance (peptide, mRNA, or tolerogenic nanoparticle constructs), aims to re-educate autoreactive T cells into a regulatory phenotype rather than broadly debulk them. Because the two act by different mechanisms, combining them is an obvious next step, and several groups are advancing antigen-specific tolerance explicitly as a combination partner.

Against this momentum, Foster et al. (2025, NOD mice; ADA abstract 2136-LB) reported a surprising result: adding anti-CD3 *reduced* the efficacy of antigen-specific immunotherapy. An earlier NOD study (Stewart et al., 2020) had likewise found that antigen-specific microparticles plus anti-CD3 “fail to synergize.” The phenomenon is therefore reproducibly observed but **unexplained**, and the T1D “modeling” literature is statistical / machine-learning, with no mechanistic within-host model of the effector–Treg–β-cell system under these two therapies. We build the smallest such model that reproduces the antagonism and ask what it implies for *how to combine* the therapies.

2. Model and methods

Three states (time in years): β-cell mass `B` (fraction of healthy, ~ proportional to C-peptide), autoreactive effector burden `E`, and antigen-specific Treg burden `R`. `E` and `R` form a **mutual-repression bistable switch** (each self-promotes via a Hill term, each represses the other — the canonical motif for immune cell-fate decisions; cf. Alexander & Wahl 2011), and β-cell mass grows logistically and is killed in proportion to effector burden:

```
dE/dt = bE + VE·E^n/(K^n+E^n)/(1+(R/Ki)^n) − dE·E − u_a3(t)·E − u_tol(t)·E
dR/dt = bR + VR·R^n/(K^n+R^n)/(1+(E/Ki)^n) − dR·R + φ·u_tol(t)·E − ρ·u_a3(t)·R
dB/dt = ρB·B·(1−B) − κ·E·B
```

The switch has two stable basins: **autoimmune** (E high, R low → β-cells lost; `B → 0.002`) and **tolerant** (E low, R high → β-cells preserved; `B → 0.931`). Late stage 2 sits in the autoimmune basin. The two interventions deliver the same total drug; only order/overlap differs. Antigen-

specific **tolerance** converts effectors to Tregs (the flux `φ·u_tol·E`; it *needs* effectors present to convert). **Anti-CD3** is lymphodepleting: it removes effectors (`u_a3·E`) but also depletes Tregs (`ρ·u_a3·R`), so on its own it transiently debulks E and the switch reverts. Outcomes are reported as a **cohort durable-control fraction** (the bistable switch makes per-patient outcomes binary, so the antagonism is read across a severity cohort). Equations, parameters, the cohort sampling, and numerics are documented in `analysis/METHODS.md`; every headline number below is re-derived and asserted by `analysis/verify_claims.py` (24/24 checks pass).

3. Results

3.1 The model reproduces the antagonism (Fig. 1, Fig. 2). Across a severity cohort, durable control (β-cell mass above threshold at 5 years) was: untreated 0%, anti-CD3 monotherapy 0% (it delays but does not durably control), **antigen-specific tolerance monotherapy 100%**. Adding anti-CD3 **simultaneously dropped control to 59%** (a 41-point loss), reproducing the Foster antagonism. Mechanistically, anti-CD3 removes the effector pool that tolerance must convert into self-stabilizing Tregs, so the switch fails to flip.

3.2 The antagonism resolves into a sequencing rule (Fig. 2). Order matters: anti-CD3-*first* gave only **41%** durable control, whereas tolerance-*first* fully rescued at **100%** — the ordering `tolerance-first ≥ tolerance-only > simultaneous > anti-CD3-first` is the mechanism's signature. The falsifiable operational prediction is an **optimal inter-drug interval** (Fig. 3): tolerance-first is flat at ~97–100% for any gap, while anti-CD3-first recovers monotonically with the gap (59% at gap 0 → 100% at a ~1.5-year gap), i.e. if anti-CD3 must precede, one must wait for the effector pool to recover before giving tolerance. Simultaneous dosing is the worst protocol.

3.3 Robustness, and a discovered two-channel caveat. Over a five-parameter grid, in the **Treg-sparing regime** (anti-CD3 spares Tregs, ρ<1 — matching teplizumab's documented profile), tolerance-first ≥ simultaneous in **171/171 viable parameter sets (100%), never worse** (strictly better in 26; the remainder saturate, so order is irrelevant there). Including strongly Treg-*depleting* anti-CD3 (ρ up to 1.1), the optimal **order can invert** (17/228 sets, all at the highest ρ). Two antagonism channels are at work: **substrate-depletion** (anti-CD3 deletes the effectors tolerance needs → favors tolerance-first) and **Treg-destruction** (anti-CD3 destroys freshly-built Tregs → favors anti-CD3-not-last). Which dominates is set by how Treg-depleting anti-CD3 is — a second, sharply falsifiable prediction.

3.4 Calibration (Fig. 4). A stage-2 cohort heterogeneous in effector severity and residual β-cell mass reproduces the untreated progression curve: **median time-to-diagnosis 2.06 years** (TN10 placebo ~2.0 yr) with **~45% progressed by 2 years** (TrialNet stage-2 ~50%). The sequencing

antagonism persists on this clinically-anchored cohort (tolerance-only 100% → simultaneous 65% → anti-CD3-first 35% → tolerance-first 97%).

3.5 A derived criterion that replicates the model (Fig. 5). Because each course (~2–4 weeks) is near-impulsive relative to the year-scale switch, it integrates *exactly* to a linear map on (E, R): tolerance gives `E→T·E, R→R+φ(1−T)E` and anti-CD3 gives `E→A·E, R→A^ρ·R`, with `T=e^{−σ_tol·τ_tol}` and `A=e^{−σ_a3·τ_a3}`. Two closed-form results follow, both verified against the full ODE. **(i) An antagonism factor:** co-administration retains only $\mathscr{A}$ = `Y_sim/Y_tol = A^ρ·σ_tol/(σ_tol+(1−ρ)σ_a3) ≈ 0.47` of the tolerogenic Treg yield — and this factorizes into exactly the two mechanisms found numerically, `A^ρ` (Treg-destruction) × `σ_tol/(σ_tol+(1−ρ)σ_a3)` (substrate-competition). **(ii) An order-inversion law:** tolerance-first nets `~A^ρ·φ(1−T)E₀` Tregs versus anti-CD3-first's `~A·φ(1−T)E₀`, so **tolerance-first is optimal ⟺ A^ρ > A ⟺ ρ < 1**, with crossover **ρ* = 1**. This single inequality *predicts* the numerically-observed two-channel inversion: the ODE benefit of going tolerance-first over simultaneous flips sign as ρ passes 1 (+41 points at ρ=0.9 → −6 points at ρ=1.1). The criterion explains the whole result — why co-dosing antagonizes ($\mathscr{A}$<1), why tolerance- first is best when anti-CD3 spares Tregs (ρ<1), and exactly where that advice reverses (ρ>1).

4. Discussion

To our knowledge this is the first mechanistic within-host model of the anti-CD3 × antigen-specific-tolerance interaction, and the first to state a sequencing rule for it. It does not claim to discover the antagonism — the *phenomenon* was anticipated empirically (Stewart 2020; Foster 2025) — but to **explain** it and convert it into an actionable prediction. The bistable Treg/effector motif itself is established (Alexander & Wahl 2011); a sequencing-matters precedent exists in cancer immunotherapy with different drug classes (Messenheimer et al. 2017); neither models this combination or states this rule. The mechanism is intuitive: antigen-specific tolerance and anti-CD3 are not interchangeable "more immunosuppression" — tolerance *needs* the effector substrate that anti-CD3 removes, so the two cooperate only when tolerance acts first.

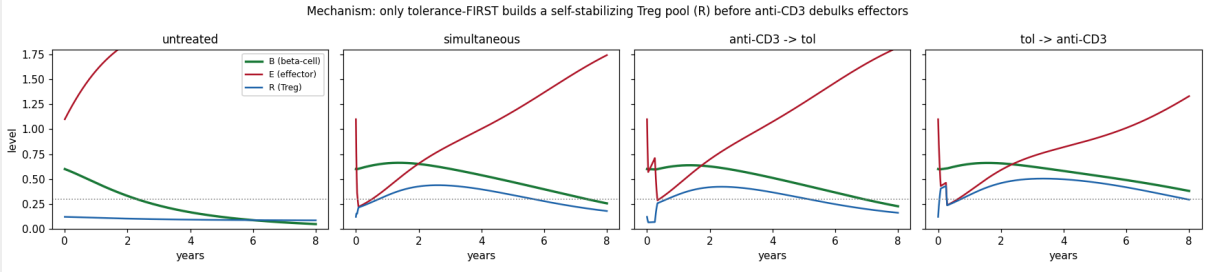
Falsifiable prediction. Combining the two therapies should be done **tolerance-first**, not simultaneously or anti-CD3-first; if anti-CD3 must precede, a sufficient inter-drug gap is required. Direct test: a NOD (or pre-clinical) sequencing experiment comparing (a) tolerance-first → anti-CD3, (b) simultaneous, (c) anti-CD3-first → tolerance, measuring diabetes incidence / banked C-peptide; the model predicts (a) > (b) > (c). A second prediction: if anti-CD3 is found to strongly deplete Tregs in vivo, the optimal order inverts toward anti-CD3-first — so the experiment should also measure Treg dynamics.

Limitations. This is an illustrative within-host model. (i) The antagonism magnitude is conditional on the Hill cooperativity n=2; at n≥3 the cross-repression sharpens and the antagonism

can reverse — the robustness sweep did not vary n, so the result should be read for the moderate-cooperativity switch. (ii) “Tolerance-first = 100% protection” is specific to the 5-year evaluation horizon (it remains the best arm, but is not exactly 100%, at 6–10 yr). (iii) The model **under-produces** the teplizumab *monotherapy* delay magnitude (+0.68 yr vs TN10’s ~+2 yr) — the robust contribution is the *sequencing* antagonism, not the monotherapy-delay magnitude. (iv) Outcomes are cohort fractions because the underlying switch is bistable. (v) Parameters are illustrative beyond the calibrated progression timing, and the Foster result is a conference abstract. A self-audit (in the repository) added caveats (i)–(iii) and is the reason the framing is “a sequencing rule for a known antagonism,” not a discovery.

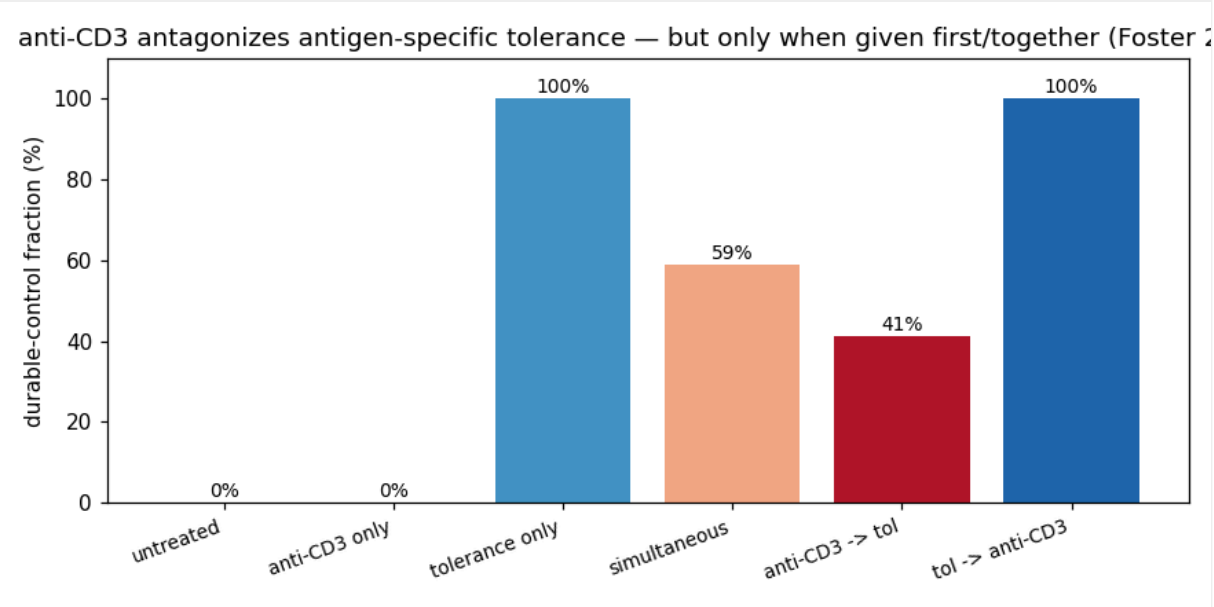
Figures

Figure 1. The mechanism — only tolerance-first builds a self-stabilizing Treg pool. Timecourses of β-cell mass (B), effectors (E) and Tregs (R) for the untreated, simultaneous, anti-CD3 → tolerance, and tolerance → anti-CD3 arms. Tolerance-first is the only schedule in which the regulatory pool R is established and held, flipping the switch to the tolerant basin.



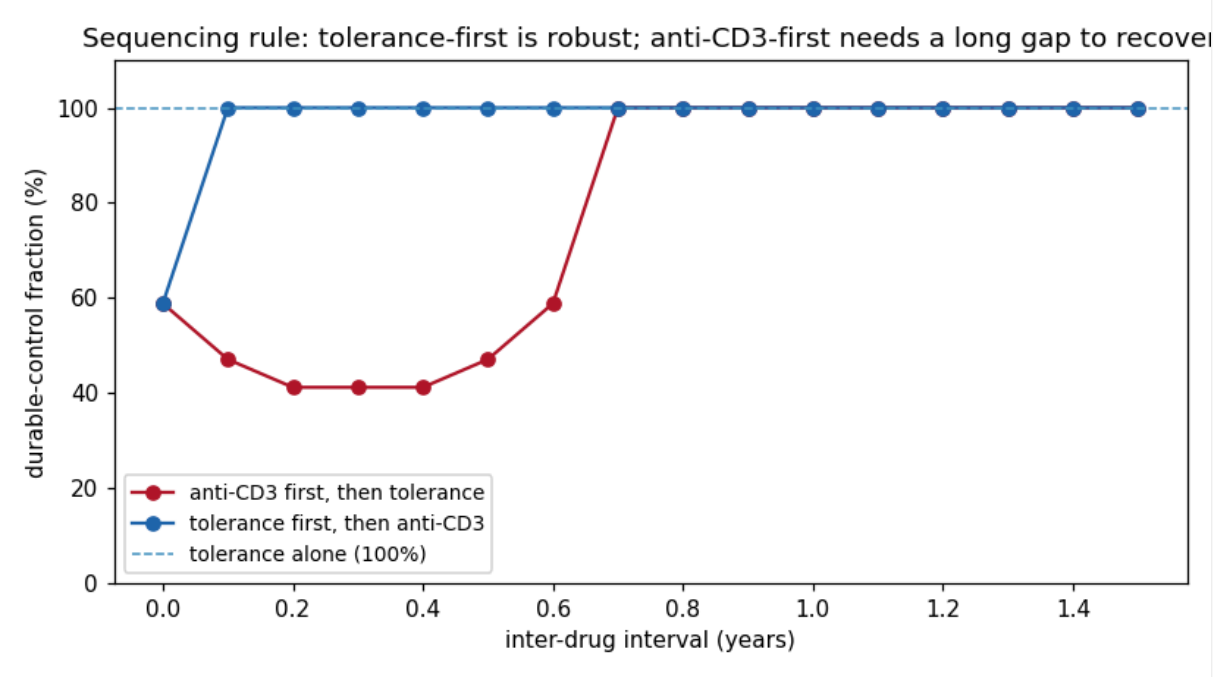
analysis/t1d_mechanism.png

Figure 2. anti-CD3 antagonises antigen-specific tolerance, and order rescues it. Cohort durable-control fraction (β-cell mass > 0.45 at 5 yr) by arm: tolerance-only 100%, simultaneous 59%, anti-CD3-first 41%, tolerance-first 100%.



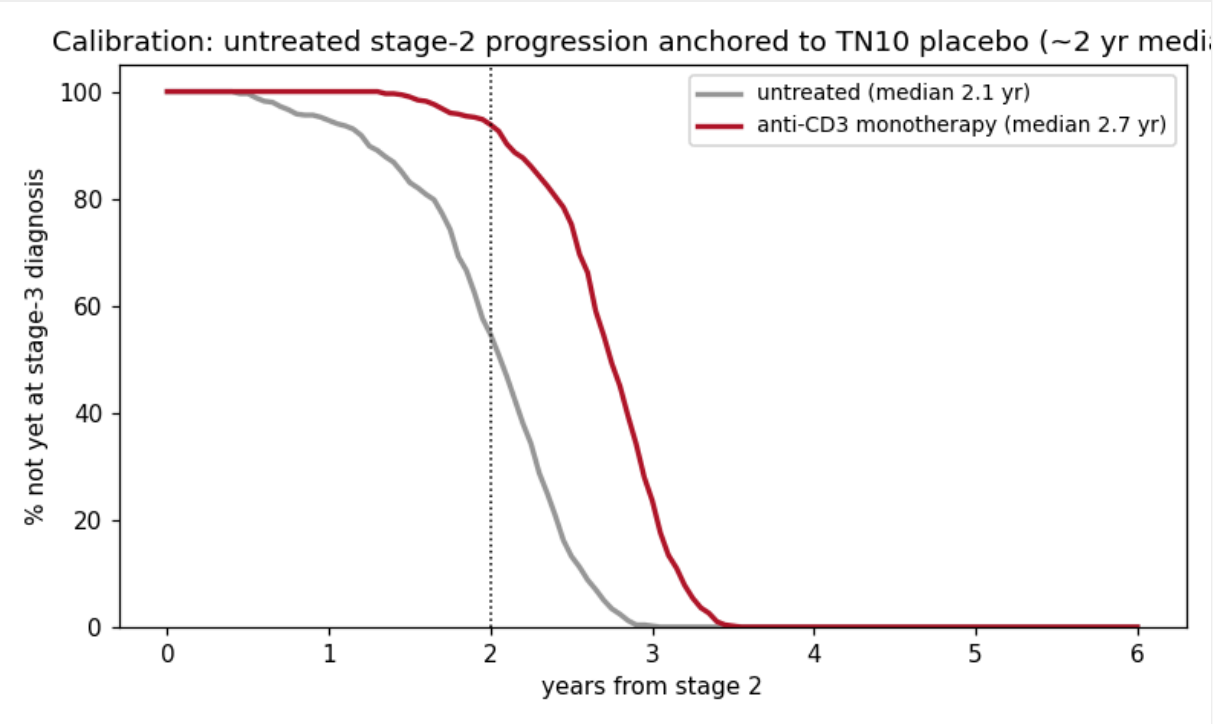
analysis/t1d_cohort.png

Figure 3. The falsifiable prediction — an optimal inter-drug interval. Durable-control fraction versus the inter-drug gap: tolerance-first is flat at ~97–100% for any gap, whereas anti-CD3-first recovers monotonically (59%→100%) only as the gap grows to ~1.5 yr.



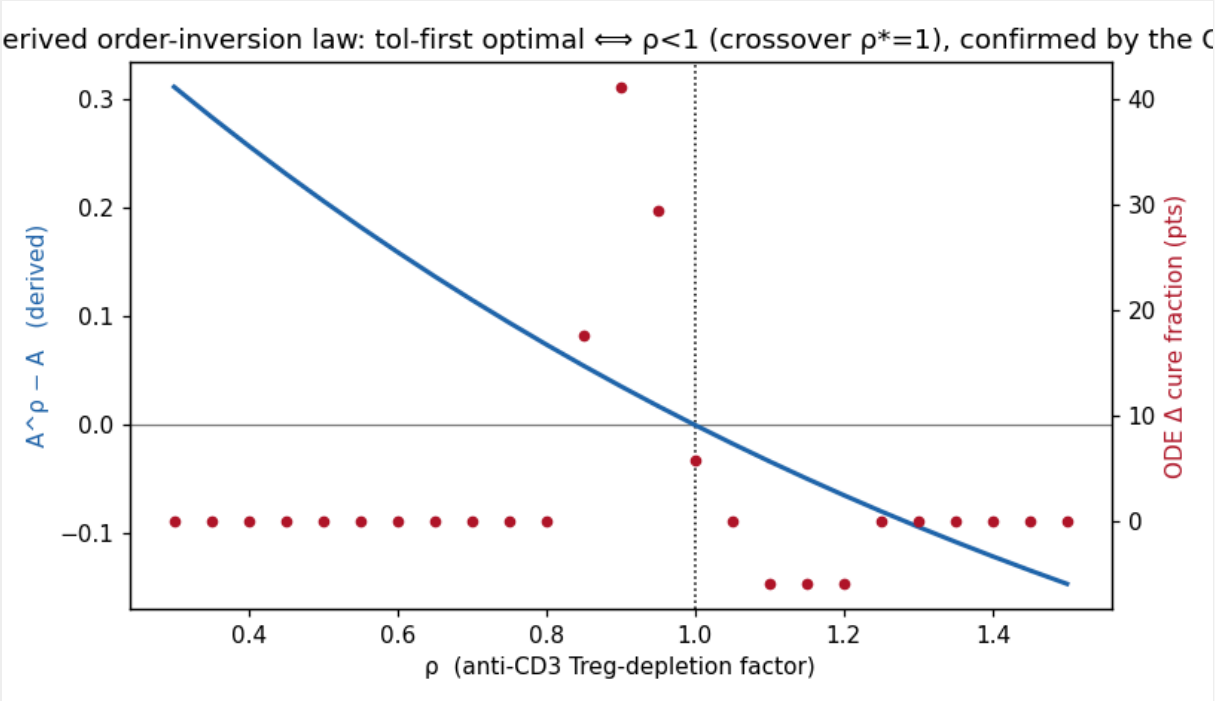
analysis/t1d_gap.png

Figure 4. Calibration to the untreated progression curve. Stage-2 cohort (heterogeneous in effector severity and residual β-cell mass): untreated median time-to-diagnosis 2.06 yr (TN10 placebo ~2.0 yr), ~45% progressed by 2 yr (TrialNet stage-2 ~50%).



analysis/t1d_calibration.png

Figure 5. The derived order-inversion law. The closed-form `A^ρ − A` (whose sign decides the optimal order) and the ODE benefit of tolerance-first over simultaneous dosing, both crossing zero at the predicted crossover ρ* = 1.



analysis/t1d_analytic.png

Data and code availability

All code, figures, the literature corpus, the verification harness, and the audit trail are openly available: https://github.com/sethc555/type1-diabetes-research (to be archived at Zenodo with a citable DOI). Running `cd analysis && python3 verify_claims.py` re-derives every headline number (24/24 checks pass).

Disclosures

Status: illustrative within-host modeling / hypothesis generation; not validated experimental or clinical findings, not medical advice. **AI assistance:** this work was developed with substantial help from an AI coding/analysis assistant (Anthropic Claude) for implementation, derivation, audit, and drafting, under the author's direction; AI tools are not authors. **Competing interests:** none. **Funding:** none.

References (key — full prior-art assessment in `analysis/NOVELTY.md`)

1. Foster J, Kelly C, et al. Anti-CD3 reduces the efficacy of antigen-specific immunotherapy in NOD mice. *Diabetes* 2025; 74 (Suppl. 1), abstract 2136-LB.
2. Stewart JM, Posgai AL, et al. Combination treatment with antigen-specific dual-sized microparticle system plus anti-CD3 immunotherapy fails to synergize to improve late-stage T1D prevention in NOD mice. *ACS Biomater Sci Eng* 2020; DOI 10.1021/acsbiomaterials.0c01075.
3. Alexander HK, Wahl LM. Self-tolerance and autoimmunity in a regulatory T cell model. *Bull Math Biol* 2011; DOI 10.1007/s11538-010-9519-2.
4. Messenheimer DJ, et al. Timing of PD-1 blockade is critical to effective combination immunotherapy with anti-OX40. *Clin Cancer Res* 2017; 23:6165–6177.
5. Herold KC, Bundy BN, et al. (TN10) An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N Engl J Med* 2019; 381:603–613.
6. Mueller SN, Ahmed R. High antigen levels are the cause of T cell exhaustion during chronic viral infection. *PNAS* 2009; 106:8623–8628.