Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Two estimation methods agree on fit but disagree up to tenfold on the binding constant: practical non-identifiability and its prediction consequences in a simulated denosumab-like quasi-steady-state target-mediated drug disposition model

Jyotheeshwar Akshay Ravikumar

jyotheeshwarakshay@gmail.com

Sri Ramachandra Institute of Higher Education and Research (Deemed to be University)

B. Anbarasan

Sri Ramachandra Institute of Higher Education and Research (Deemed to be University)

Research Article

Keywords: Practical identifiability, Target-mediated drug disposition, Quasi-steady-state approximation, Population pharmacokinetics, Profile likelihood, Denosumab

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11744808/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The quasi-steady-state (QSS) approximation to target-mediated drug disposition (TMDD) summarises drug–target binding in one constant, K_{ss}. Structural identifiability of this class is established; whether a sparse design estimates that constant precisely, and what imprecision costs, are separate questions. We simulated 24 replicate datasets from a simplified denosumab-like two-compartment QSS-TMDD model (60 mg subcutaneously at 0, 6 and 12 months; 30 subjects; 24 samples each) and fitted each by two estimators: stochastic approximation expectation–maximisation (SAEM) under multistart initialisation, and first-order conditional estimation with interaction (FOCEi). Under one common objective evaluated at fixed parameters the two estimators differ by at most 7.6 units on every replicate, with no fit identifiable as an optimisation failure, while their K_{ss} estimates differ by up to tenfold. K_{ss} fails the prespecified precision criterion under both — 58.9% and 42.0% relative standard deviation against 23.8% and 22.3% for the next-worst parameter — so the ranking is estimator-independent. The single-fit standard error reports 15.9% and 9.6%, understating the spread 3.7- and 4.4-fold, and is uninformative in two fits (relative standard error 1832% and 173%). Replacing only K_{ss} in the true vector moves predicted drug concentration by 1.4–1.5% and predicted free target by 24.8–28.0%, an approximately 18-fold asymmetry under all three estimate sets. This points to measuring the target rather than sampling the drug more densely, and agreement of two estimators on the objective with disagreement on a parameter is a direct test of practical non-identifiability.

Introduction

TMDD models describe drugs whose target binding materially affects their own disposition [1]. Under the QSS approximation the binding interaction is summarised by a constant K_{ss}, alongside a target baseline R_0 and an internalisation rate k_{int} [2, 3]. Structural identifiability of the standard TMDD approximations from drug-concentration data is established [2], and the dynamics of parameter identification in this class have been characterised analytically [4].

Three questions are routinely conflated. *Structural identifiability* asks whether parameters are uniquely recoverable in an ideal, noiseless setting, and is settled for this model class. *Practical identifiability* asks whether a finite, noisy, censored design supports estimation with acceptable precision; profile likelihood rather than the Fisher information matrix is the appropriate instrument [5, 6], and the available approaches have been compared systematically [7, 8, 9, 10], with the underlying theory in [11]. *Algorithmic recoverability* asks whether the estimator reliably finds the relevant optimum. Because the model class is structurally identifiable, anything that remains is practical or algorithmic, and fitting the same datasets by two estimators separates those two in turn.

A fourth question asks what an imprecise parameter costs: a parameter can be badly determined and harmless, if the predictions a model is used for do not depend on it, or badly determined and consequential, and a precision statistic cannot distinguish them. The question is not new — prediction profile likelihood places confidence bounds on a prediction rather than a parameter [12], sloppiness

analysis established that poor parameter precision and poor prediction accuracy are distinct properties [13], and for TMDD reduced and optimised trial designs have been evaluated by their effect on parameter uncertainty and free-target prediction precision [14]. What is added here is narrower: the cost is attributed to a single parameter rather than to a design, established by simulation and re-estimation rather than Fisher-information design evaluation [23, 24]; the model-based standard error for that parameter is anti-conservative under two estimators; and a between-start diagnostic decides whether a computed profile may be read at all. Published population analyses of denosumab report high shrinkage on the target-related parameters because circulating receptor activator of nuclear factor κB ligand (RANKL) was not measured [15], and shrinkage is a limited diagnostic of precision [16].

We therefore ask: under a realistic sparse design, what is the per-parameter recovery of a simplified denosumab-like QSS-TMDD model, and does it depend on the estimator; is any single parameter responsible; and what does that parameter's estimation error cost in prediction?

Methods

The complete specification — all units, the parameter table, the free-parameter accounting, the mass-conservation verification, the independent second implementation and the prespecified criteria — is given in Online Resource 1, S1. The model equations and the points bearing on interpretation are given here.

Model

The states are a subcutaneous depot amount A_{dep} (nmol) and three concentrations (nmol L^{-1}): total (free plus bound) central drug C_{tot}, peripheral drug C_p, and total target R_{tot}. Free drug C is the positive root of the QSS quadratic and the drug–target complex RC follows by mass balance:

$$C = \tfrac{1}{2}\left[(C_{tot} - R_{tot} - K_{ss}) + \sqrt{D} \right],\ D = (C_{tot} - R_{tot} - K_{ss})^2 + 4\,K_{ss}\,C_{tot} \quad (1)$$

$$RC = C_{tot} - C \quad (2)$$

D is a sum of a square and a non-negative term and is therefore never negative, and C is the positive root, so neither quantity requires a truncation branch. The system is

$$\mathrm{d}A_{dep}/\mathrm{d}t = -k_a\,A_{dep} \quad (3)$$

$$\mathrm{d}C_{tot}/\mathrm{d}t = k_a\,A_{dep}/V_c - (CL/V_c)\,C - (Q/V_c)\,C + (Q/V_c)\,C_p - k_{int}\,RC \quad (4)$$

$$\mathrm{d}C_p/\mathrm{d}t = (Q/V_p)\,C - (Q/V_p)\,C_p \quad (5)$$

$$\mathrm{d}R_{tot}/\mathrm{d}t = k_{syn} - k_{deg}\,(R_{tot} - RC) - k_{int}\,RC \quad (6)$$

with $R_{tot}(0) = R_0$, $A_{dep}(0) = C_{tot}(0) = C_p(0) = 0$ and $k_{deg} = k_{syn}/R_0$ derived rather than estimated. The observation model is combined additive and proportional residual error on total drug,

$DV = C_{tot}\,(1 + \varepsilon_{prop}) + \varepsilon_{add},\ \varepsilon_{prop} \sim N(0, \sigma^2_{prop}),\ \varepsilon_{add} \sim N(0, \sigma^2_{add})$ (7)

matching the parameterisation of the source analysis. All volumes and clearances are apparent (divided by bioavailability *F*).

Ground truth is a simplified model, not a published one

Parameters were taken from a published denosumab population analysis [15] with three stated departures: no CL–V_p random-effect correlation, no body-weight covariates, and Q fixed at 0.20 L h^{-1} without between-subject variability. The system studied is a *simplified denosumab-like QSS-TMDD model derived from* that analysis, and no result here characterises the published clinical model.

A units correction in the source table

Table 3 of the source analysis labels k_{syn} as h^{-1}; both the source's own steady-state relation ($k_{syn} = R_0 \times k_{deg}$) and Eq. (6) require nmol L^{-1} h^{-1}. The published value is used unchanged with the corrected unit. A consequence follows from the published numbers alone: $k_{deg} = k_{syn}/R_0 = 6.57 \times 10^{-4}$ h^{-1}, implying a RANKL half-life of 44 days, so the ground-truth target is nearly static over a 6-month dosing interval.

Simulated population, design and estimation

Thirty virtual subjects received 60 mg subcutaneously at 0, 6 and 12 months (60 mg of a 147 kDa antibody is 408.1633 nmol), with eight samples per dose window (12, 24, 72, 168, 336, 720, 1440 and 2160 h after each dose). The assay limit of quantification was 0.136 nmol L^{-1} and below-limit observations were discarded (M1); the behaviour of M1 against the censored-likelihood alternative as the censored fraction grows is documented elsewhere [17, 18, 19]. Eight fixed effects, eight diagonal variance components and two residual terms were estimated in nlmixr2 [20].

Each of the 24 replicate datasets was fitted by two estimators. *SAEM* [21] was run under best-of-four multistart initialisation: three log-normally perturbed starts ($\sigma = 0.5$ on the log scale) plus a true-value start at a matched budget, with the lowest-objective start refit at 150 burn-in and 200 EM iterations. *FOCEi* was run once per replicate from that replicate's SAEM estimate, so the two are compared on identical data from identical starting values. A third set initialised at the true values is reported as a comparison; because the true values are not a maximum-likelihood start, that set bounds achievable recovery rather than estimating it. SAEM iteration budgets differ between analyses and objectives are compared only within an analysis (Online Resource 1, Table S1). The simulated subjects carry no sex, gender, age or ethnicity attribute, so sex- and gender-based analysis is not applicable and the absence of demographic covariates limits generalisability ("Limitations").

Prespecified criteria

A parameter is adequately recovered when |relative bias| ≤ 20% and RSD ≤ 30%, where relative bias is $100 \cdot [\text{median}(\theta) - \theta_{true}]/\theta_{true}$ and RSD is the empirical across-replicate $100 \cdot \text{SD}(\theta)/\text{mean}(\theta)$. These are

operational thresholds, not drawn from a published standard, and they were declared before the profile, design-arm, sample-size and calibration results existed but after the true-value-start estimates were in hand, so for that comparison set they are post hoc. Convergence is defined by five operational criteria recorded per fit (Online Resource 1, S1.6.3).

Comparing two estimators on one objective

Two estimates of one dataset can differ because the likelihood is flat between them or because an estimator failed to optimise; separating these requires evaluating both vectors under a single method. For each replicate the objective was evaluated at fixed parameters — every typical value, every variance component and both residual terms fixed, with only the individual random effects estimated — for the SAEM vector, the FOCEi vector and the true vector. Thresholds were fixed before the evaluations ran: objectives within 4 units indicate a likelihood flat in K_{ss} over the range the vectors span, a gap above 10 units identifies the higher vector as an optimisation failure, and between 4 and 10 neither statement is made. The fitted variance matrices were not recorded for every fit, so the variance components are held at common values for all three vectors; the comparison is therefore of fixed-effect vectors under one random-effects model, and the objectives are not the fits' own reported values (Online Resource 1, S1.6.5).

Prediction impact used no new fits

The typical-subject system (all random effects zero) was solved for each fitted vector and for the truth on a 1,501-point grid over 18 months with an independent Python/SciPy LSODA implementation of Eqs. (1)–(7) (rtol 10^{-9}, atol 10^{-11}). Free RANKL was recovered from the QSS algebra. Error is the mean over the grid of |prediction − truth|/truth, evaluated identically for all three estimate sets.

Inventory of analyses

315 fits in total, counted from the per-fit tables rather than from a plan (Online Resource 1, Table S12): 24 datasets fitted once each at true-value starts; 120 multistart fits in three phases; 24 FOCEi fits; 16 fits for the convergence check; 56 profile fits (2 parameters × 7 grid points × 4 starts); 18 design-arm fits; 6 fits at $N = 80$; 6 with per-parameter uncertainty and η-shrinkage captured; 8 reproducibility fits; 13 supporting fits for seed sensitivity, random-effects structure and the initialisation pilot; and the 24-replicate re-analysis with full convergence capture. The 72 fixed-parameter objective evaluations, the Fisher-information eigenstructure and the prediction-impact analysis required no estimation. Every data table is one row per fit.

Results

Two estimators agree on fit and disagree on K_{ss}

All 24 datasets were fitted successfully by both estimators, and the covariance step succeeded on all 24 FOCEi fits. The K_{ss} estimates do not agree: the FOCEi-to-SAEM ratio has median 1.17 but spans 0.59 to

10.28, differing by more than 25% in 13 of 24 replicates and more than two-fold in 4 (Table 1).

Table 1
Per-replicate comparison of the two estimators across 24 independent 30-subject datasets

Replicate	K_{ss} SAEM[a]	K_{ss} FOCEi[a]	Ratio[b]	ΔOFV (FOCEi − SAEM)[c]
1	1.771	2.642	1.49	-5.53
2	1.996	1.614	0.81	-2.98
3	1.318	1.949	1.48	-3.07
4	1.470	1.708	1.16	-3.05
5	1.265	1.811	1.43	-0.45
6	1.957	1.757	0.90	+ 0.11
7	1.473	2.460	1.67	-3.06
8	1.107	3.020	2.73	-5.43
9	1.450	1.577	1.09	+ 1.46
10	0.486	0.473	0.97	+ 1.23
11	0.201	2.066	10.28	-5.57
12	0.703	0.704	1.00	-1.20
13	4.332	2.568	0.59	-5.01
14	2.676	3.402	1.27	-0.91
15	1.666	2.862	1.72	+ 4.39
16	0.848	0.542	0.64	+ 0.89
17	1.521	1.625	1.07	-0.60
18	2.033	2.094	1.03	-7.60
19	1.136	1.082	0.95	+ 1.62
20	1.464	1.728	1.18	-0.75
21	1.124	1.906	1.69	-6.16
22	0.566	1.464	2.59	-4.22
23	1.018	0.954	0.94	+ 2.07
24	0.660	1.394	2.11	+ 2.67

[a] Population estimate, nmol L^{-1}. [b] FOCEi estimate divided by the SAEM estimate. [c] Difference in objective function value between the FOCEi and the SAEM parameter vector, both evaluated at fixed parameters under one common method (Methods, “Comparing two estimators on one objective”). A negative value means FOCEi reaches the lower objective. No replicate reaches the prespecified 10-unit threshold for an optimisation failure.

Evaluated under one common objective at fixed parameters (“Comparing two estimators on one objective”), those disagreements cost almost nothing. The absolute objective difference between the two vectors is at most 7.6 units on any replicate, median 2.8, within 4 units in 16 of 24 and between 4 and 10 in the remaining 8. *The 10-unit threshold that would identify an optimisation failure is reached on no replicate*, and FOCEi takes the lower objective in 16 of 24, so neither estimator is systematically failing. Two estimators starting from the same values on the same data settle up to tenfold apart in K_{ss} at a fit the objective cannot distinguish: practical non-identifiability of K_{ss} under this design, shown without reference to any one algorithm’s behaviour.

The consequence for recovery is estimator-independent. K_{ss} fails the precision criterion and is rank-last of the ten parameters by RSD under both estimators — 58.9% under SAEM and 42.0% under FOCEi, against 23.8% and 22.3% for the next-worst parameter, R_0 (Table 2). The two disagree about the sign of the bias (− 11.3% and + 11.7%) and both pass the bias criterion, so under either estimator the failure is of precision, not accuracy. The other seven structural parameters and both residual terms pass both criteria under both estimators. FOCEi is the less dispersed on every target-related parameter, narrowing the gap without closing it, and the ordering is unchanged. The 24 true-value-start fits give 35.1% RSD at − 24.6% bias: holding the start at the truth suppresses the spread and inflates the bias, which is why that set bounds recovery rather than estimating it.

Table 2
Per-parameter recovery under two estimators, each fitted from the same starting values on the same 24 datasets

Parameter	True	SAEM multistart median	Rel. bias (%)[a]	RSD (%)[a]	FOCEi median	Rel. bias (%)[a]	RSD (%)[a]	Both criteria[b]
V_p/F (L)	6.06	6.09	0.6	2.7	6.12	1.0	2.8	pass
σ_{prop}	0.07	0.0708	1.2	4.6	0.0710	1.4	4.2	pass
CL/F (L h^{-1})	0.006	0.0061	1.5	5.9	0.0061	2.1	5.7	pass
k_a (h^{-1})	0.0078	0.0077	-0.7	8.0	0.0078	0.1	8.1	pass
σ_{add} (nmol L^{-1})	0.72	0.73	1.7	11.1	0.71	-1.9	11.2	pass
V_c/F (L)	1.58	1.62	2.3	11.7	1.65	4.4	11.8	pass
k_{int} (h^{-1})	0.022	0.0220	0.1	12.6	0.0221	0.6	6.9	pass
k_{syn} (nmol L^{-1} h^{-1})	0.01	0.0094	-5.8	14.2	0.0093	-6.6	12.7	pass
R_0 (nmol L^{-1})	15.23	15.45	1.4	23.8	16.10	5.7	22.3	pass
K_{ss} (nmol L^{-1})	1.56	1.38	-11.3	58.9	1.74	11.7	42.0	fail (RSD)

[a] Relative bias and relative standard deviation as defined in Methods, “Prespecified criteria”. [b] A parameter passes when |relative bias| ≤ 20% and RSD ≤ 30%. Rows are ordered by SAEM RSD. The corresponding true-value-start comparison set is Online Resource 1, Table S14.

The variance components follow the same ordering and are harder to recover: only two of eight meet the criteria and the three worst are again the target parameters, $\omega^2(k_{syn})$, $\omega^2(K_{ss})$ and $\omega^2(k_{int})$ (Online Resource 1, Table S11).

Convergence, chain length and parameter boundaries

Because the two estimators disagree, the possibility that either had not converged was tested rather than assumed away. Eight replicates, selected in advance, were refit by SAEM at a roughly six-fold longer chain (1000 burn-in, 1000 EM against 150 and 200) from their own estimates, under a rule fixed before the fits ran.

The longer chain did not stabilise K_{ss}: it changed by more than 5% in all eight (median absolute change 27.2%, range 12.7–279.6%) and agreed with FOCEi to within 10% in only one. The traces explain why.

During burn-in K_{ss} wanders over an order of magnitude; from the first EM iteration the trace is flat and does not move again, because the EM phase averages the stochastic approximation and retains wherever burn-in left the chain. *A flat parameter trace over the second half of an SAEM run is therefore not evidence of convergence in this model*, and a longer chain lengthens the frozen segment rather than extending the search. The long chain also drove the between-subject variance on K_{ss} to a median 2% of its true value, against 64% in the fits reported here, so chain length reshapes the variance structure as well.

Two FOCEi fits returned $\omega^2(\eta_{Kss})$ at exactly the upper bound the estimator imposes on a diagonal variance, five times its initial value (replicates 11 and 23); the convergence definition as first declared tested only a lower bound and could not detect this, and has been made two-sided and estimator-aware (Online Resource 1, S1.6.3). Two further fits exceed the 90% η-shrinkage ceiling (replicates 10 and 16, at 91.5% and 94.8%). All four are reported rather than excluded; 20 of 24 FOCEi fits meet all five criteria.

Reported uncertainty understates the spread, and is sometimes uninformative

The standard error from a single fit does not reveal the problem under either estimator (Table 3). For K_{ss} the model-based relative standard error has median 15.9% under SAEM and 9.6% under FOCEi against across-replicate spreads of 58.9% and 42.0% — understatement factors of 3.7 and 4.4 — and coverage of the true value by the nominal 95% interval is 12 of 24 and 9 of 24. Under SAEM the three target parameters are the three whose model-based uncertainty most understates the empirical spread (ratios 0.45, 0.31, 0.27 against 0.82–1.19 for the five disposition parameters), and K_{ss} is worst under FOCEi at 0.23. With 24 replicates a coverage figure carries roughly ± 10 percentage points and is read as a direction, but the direction is consistent across the three parameters and matches the ratios.

Table 3
Model-based against empirical uncertainty under the two estimators, the same 24 datasets

Parameter	SAEM rel. SE (%)[a]	SAEM empirical RSD (%)	Ratio[b]	Coverage[c]	FOCEi rel. SE (%)[a]	FOCEi empirical RSD (%)	Ratio[b]	Coverage[c]
k_a (h^{-1})	9.6	8.0	1.19	23/24	10.3	8.1	1.26	22/24
V_p/F (L)	3.0	2.7	1.10	24/24	3.2	2.8	1.13	23/24
V_c/F (L)	12.0	11.7	1.02	23/24	7.4	11.8	0.63	13/24
R_0 (nmol L^{-1})	22.8	23.8	0.96	22/24	29.1	22.3	1.30	21/24
CL/F (L h^{-1})	4.8	5.9	0.82	19/24	5.1	5.7	0.90	21/24
k_{syn} (nmol L^{-1} h^{-1})	6.4	14.2	0.45	14/24	8.4	12.7	0.66	17/24
k_{int} (h^{-1})	3.9	12.6	0.31	14/24	5.9	6.9	0.85	21/24
K_{ss} (nmol L^{-1})	15.9	58.9	0.27	12/24	9.6	42.0	0.23	9/24

[a] Model-based relative standard error of θ on its natural scale, that is SE(log θ). [b] Model-based relative SE divided by the empirical RSD. A value below 1 means the covariance step reports more precision than the estimates have. Rows are ordered by the SAEM ratio. [c] Number of the 24 replicates whose nominal 95% interval contains the true value. Per-parameter η-shrinkage is in Online Resource 1, Table S6.

In two FOCEi fits the covariance step is not optimistic but uninformative: the relative standard error on K_{ss} is 1832% (replicate 4) and 173% (replicate 24). Both are finite and so satisfy a finiteness test — the opposite failure mode to the understatement above, on the same parameter. The source analysis shows the same understatement on clinical data, reporting K_{ss} with a relative standard error of 12.4% alongside high shrinkage on K_{ss}, R_0 and k_{int} and the statement that drug-concentration data alone were insufficient to characterise target and complex turnover [15].

Table 3 reports SE(log θ), which is scale-free; the software's %RSE column divides it by |log θ| and is therefore not comparable across parameters (Online Resource 1, S1.6.3).

What the K_{ss} error costs in prediction

Perturbing the true vector by nothing but the fitted K_{ss} error — isolating it from the other seven parameters — moves predicted drug concentration by a median 1.4–1.5% and predicted free RANKL by 24.8–28.0%

(Table 4, Fig. 1). The ratio is 17.7, 17.5 and 18.5 for the true-value-start, SAEM and FOCEi sets, computed identically; under FOCEi no replicate exceeds 5% on drug and 14 of 24 exceed 20% on free RANKL.

Table 4
Median absolute prediction error over 18 months for a typical subject, from the true parameter vector with only K_{ss} replaced by each fitted value

Estimate set	Total drug (measured)[a]	Free RANKL (not measured)[a]	Ratio[b]
24 true-value-start fits	1.5	26.4	17.7
24 SAEM multistart winners	1.4	24.8	17.5
24 FOCEi estimates	1.5	28.0	18.5

[a] Median across the 24 vectors of the mean absolute percentage error over an 18-month horizon, per cent.
[b] Free RANKL error divided by total drug error. No new fits were run for this table.

This asymmetry is substantially a consequence of model structure and should be read as such: free RANKL is obtained from the QSS algebra of Eqs. (1)–(2), in which K_{ss} appears explicitly, whereas total drug is buffered by the disposition parameters, so the regression of predicted maximum RANKL suppression on the K_{ss} error (slope 1.14, SE 0.10, $r = 0.92$) confirms propagation rather than providing independent evidence. What simulation–estimation adds over a sensitivity analysis is the pairing of that sensitivity with a realised error distribution: a sensitivity analysis can report that free target responds more strongly to K_{ss} than drug does, but not that the error a sparse design actually produces moves the unmeasured prediction by a quarter while leaving the measured one inside 2%, nor that the ratio holds to within 6% across two estimators and three initialisation schemes. The individual percentages inherit the start-, budget- and estimator-conditionality of the fitted errors; the ratio does not.

The isolated effect must not be read as a whole-model accuracy claim: the complete fitted vector carries a median 11.5% error on total drug and 30.8% on free RANKL, and substituting the true K_{ss} removes 58% of the free-RANKL error against 0.2 percentage points of the drug error.

The weak direction of the information matrix is essentially K_{ss} itself: a sensitivity-based Fisher information at the typical subject has condition number 9.4×10^4 and its smallest eigenvector loads + 0.99 on log K_{ss} with no other loading above 0.08, so the poorly determined direction is one parameter rather than a composite — which is what makes a per-parameter answer meaningful. Its ordering agrees with the observed RSD ($\rho = 0.74$, $p = 0.037$) but its level does not (“Mechanism and sample size”). Fixed-seed fits in this build are bit-identical across seeds, repeats and thread counts, so movement under changes of seed or start is a property of the objective surface, not floating-point noise.

Profile likelihood, and a diagnostic for when one is usable

Profiles for K_{ss} and for R_0 as comparator were computed on 7-point grids spanning ± 1.5 log units about the truth, at four log-normally perturbed starts per grid point (Fig. 2; all 56 objectives in Online Resource 1,

S3). *Single-start profiling failed for both parameters*: all 14 grid points lay above the best of four starts, with three extra starts recovering 8.3–27.3 objective units for K_{ss} and 5.1–38.9 for R_0. A smooth-looking single-start R_0 curve was not evidence that one start sufficed, and a preprint deriving exact total-QSS formulas for competitive TMDD independently reports a false objective-function minimum in a reduced TMDD model [22].

A between-start diagnostic separates a usable profile from an unusable one without knowing the true optima. Dividing the median spread of the four objective values at one fixed value by the span of the whole four-start curve gives 0.36 for R_0 (29.3/80.3) and 1.36 for K_{ss} (22.6/16.6). A profile is a lower envelope over starts, so more starts can only lower it; when the ratio exceeds one the envelope is not a stable estimate and no interval or flatness statement can be read from it. R_0's profile is usable at four starts and K_{ss}'s is not — its minimum sits at 1.65 × truth with ΔOFV 11.0 at the truth, but that displacement cannot be interpreted.

Mechanism and sample size

Two mechanisms could explain the deficit: the schedule stops at 2160 h of a 4383 h interval, or the target's 44-day half-life leaves it nearly static. Three arms of six paired datasets each (identical subject-level random effects across arms) discriminate them (Table 5, Fig. 3). These arms and the N = 80 arm were fitted by SAEM from true-value starts only, so they carry the start-conditionality of "Two estimators agree on fit and disagree on K_{ss}". Accelerating target turnover 100-fold — scaling k_{syn} so k_{deg} rises by the same factor, RANKL half-life 44 days → 10.6 h — moved K_{ss} from failing both criteria to passing both and improved the estimate in 6 of 6 paired datasets (exact sign test p = 0.016), surviving a Bonferroni correction for two comparisons; adding samples at 3600 and 4380 h improved 3 of 6 (p = 0.66). At n = 6 the smallest attainable p-value is 0.0156, so this is a direction rather than a magnitude. The turnover arm changes target timescale, flux and censored fraction together (median 107 below-limit observations per dataset against 0.5 in the reference arm), so it is diagnostic rather than actionable, and the attribution that survives is negative: it is *not* the sampling schedule.

Table 5
K_{ss} recovery by design arm and sample size

Arm	Median K_{ss}[a]	Rel. bias (%)	RSD (%)	Criteria	Paired vs reference[b]
Reference design	1.17	−24.7	35.6	fail (both)	—
+ late samples (3600, 4380 h)	1.74	+ 11.7	22.5	pass	3/6 closer, p = 0.66
100× faster target turnover	1.43	−8.1	16.8	pass	6/6 closer, p = 0.016
Reference design, N = 80	1.63	+ 4.8	21.8	pass (both)	—

[a] nmol L⁻¹; true K_{ss} 1.56 nmol L⁻¹. All fits in all four rows were initialised at the true parameter values. [b] Exact one-sided sign test on the within-dataset change in $|K_{ss}$ − truth|/truth. The first three rows are six paired datasets per arm, with the same subject-level random effects in all three arms of each replicate. The *N* = 80 row is a separate six-dataset experiment, not paired with the other three, which is why it carries no paired entry.

Six datasets of 80 subjects settled the sample-size question directly: the observed RSD was 21.8% against 21.5% from $N^{-1/2}$ scaling of the empirical value, and the median bias moved to + 4.8%, so K_{ss} passes both criteria at *N* = 80 (Table 5). The Fisher information predicted 129% at *N* = 30 and 78.8% at *N* = 80 against 35.1% and 21.8% observed, because it is not the same quantity: a fixed-effects matrix at the typical subject carries no information from the random-effects structure, and its *N* = 80 figure is the √(30/80) rescaling of its own *N* = 30 value rather than an independent check.

Discussion

What the two-estimator result establishes

The central observation is negative and therefore hard to argue with: two estimators, run on identical data from identical starting values, settle up to tenfold apart in K_{ss} at objectives that never differ by more than 7.6 units, with no fit identifiable as an optimisation failure. Neither algorithm is broken; the likelihood does not distinguish the answers they return. A single-estimator study cannot make that statement, because a disagreement between starts can always be attributed to the estimator — which is why both are reported here as peers rather than one as primary and the other as a robustness check. The precision failure is therefore a property of the design, not the method.

What makes this consequential in practice is that none of the diagnostics an analyst would reach for detects it. Trace flatness and chain length are the most readily available, and in this model a flat SAEM trace over the second half of a run is an artefact of the averaging phase; lengthening the chain roughly six-fold moved K_{ss} in all eight replicates tested while driving its between-subject variance towards zero. The covariance step understates the spread 3.7- to 4.4-fold, and in the two fits where it does signal trouble it does so through a relative standard error of 1832% and 173% — finite, and so passing a finiteness test. An analyst running either estimator alone, inspecting the convergence flags and reading the reported standard error would see nothing wrong.

On mechanism, the static-target explanation is supported and the sampling-location explanation is not, at this power, and that inverts the practical recommendation. If the deficit were sampling location the fix would be a longer schedule, cheap and available to any analyst; because it is the target's own timescale, no rearrangement of drug-concentration samples supplies the missing information. The interventions worth costing are a RANKL or bound-complex assay, a reparameterisation that estimates an identifiable combination in place of K_{ss}, or an informative prior. Increasing the cohort to 80 works at these parameter values but pays $N^{-1/2}$ to shrink one side of an 18-fold asymmetry, where a target assay attacks the

asymmetry itself. The variance components sharpen the point: six of eight fail the recovery criteria and the three worst are the three target parameters, so the population model misreports not only the typical value of the binding constant but its spread across subjects — the quantity any individualised target-suppression prediction would be built on.

Two findings that stand independently of the drug

Single-start profile likelihoods of this model are unreliable for both parameters tested, including every point of a curve that looked smooth; and the between-start spread relative to a curve's own span distinguishes a usable profile from an unusable one without knowledge of the true optima. Any SAEM profile of a TMDD model should report the between-start spread at each grid point alongside the profile, because without it a smooth curve carries no information about whether its points are optima.

Limitations

The ground truth is a simplified model ("Ground truth is a simplified model, not a published one") with no demographic covariates and no sex or gender dimension ("Simulated population, design and estimation"), so no result transfers directly to a real population. The ground-truth turnover implies a RANKL half-life of 44 days, which follows arithmetically from the published k_{syn} and R_0 but is long for a cytokine-like ligand; if the true turnover is faster, the design studied here is harder than reality and the difficulty is overstated. Below-limit observations were discarded (M1) rather than modelled with a censored likelihood (M3), the less efficient choice, and this interacts with the turnover arm, where the censored fraction is far larger. One parameter set was studied, so nothing here speaks to parameter regions away from the published denosumab values. The design arms and the sample-size arm rest on six datasets each, and the turnover arm is not separable into timescale, flux and censoring. The Fisher information is a fixed-effects matrix at the typical subject rather than a population Fisher information from a dedicated design tool [24], and its level should be used only as an ordering. The winning multistart start was identified at the lower 100/150 budget and only then refit at 150/200, so the selection step is cheaper than the estimate it selects; selection margins were large relative to budget effects (a median 28.6 objective units in the replicates examined), but invariance of the winning start to that budget has not been shown. Prediction impact is a typical-subject calculation against a single truth with no propagated estimation uncertainty, and no population predictive check was run. Profiles are on one dataset and no k_{int} profile was computed. The full list is in Online Resource 1, S4.

A note on the workflow

The simulation and estimation models are generated from a single parameter definition and their agreement is asserted before any fit; the correction history is in Online Resource 1, S5. The general point is practical: a simulation–estimation study should report its consistency check in the same way a fit reports convergence status.

Conclusions

Under the sparse drug-concentration design tested here, at 30 subjects, for a target-mediated antibody whose target turns over slowly relative to the dosing interval, K_{ss} is practically non-identifiable: two estimators settle up to tenfold apart at objectives differing by at most 7.6 units with no optimisation failure, and K_{ss} is rank-last of the ten parameters by precision under both. The difficulty is design- and cohort-dependent rather than a property of the model class — six datasets of 80 subjects brought K_{ss} inside both criteria. The same estimation error is approximately 18 times more expensive on predicted target suppression than on predicted exposure, stably across both estimators and three initialisation schemes, while the model-based standard error understates the K_{ss} spread 3.7- to 4.4-fold.

Three consequences are general. A per-parameter identifiability result should be expressed in units of the prediction the model will be used for, because precision alone cannot separate a parameter that is badly determined and harmless from one that is badly determined and consequential. Agreement of two estimators on the objective combined with disagreement on a parameter is a direct, algorithm-independent demonstration of practical non-identifiability, and is cheaper than a profile likelihood. And neither a flat SAEM trace nor a longer chain is evidence of convergence in this model class; single-start profile likelihoods are unreliable even where the curve looks smooth, and the between-start spread relative to the curve's own span tests whether a profile may be read at all.

Declarations

Funding. No funding was received for conducting this study.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval. Not applicable. This study did not involve human participants, human data, human biological material or animals; all data analysed were generated by computer simulation from the model specified under Methods.

Consent to participate and consent for publication. Not applicable, for the reason given under Ethics approval.

Data availability. All data in this study are simulated and are fully regenerated by the deposited code. The simulation and estimation code, the job specification and one-row-per-fit result tables for every analysis reported are deposited at Zenodo under DOI 10.5281/zenodo.23761636 (https://doi.org/10.5281/zenodo.23761636) and cited as reference [25]. The deposit contains 00_truth.R (the single parameter definition generating both the simulation and the estimation model, with an automated consistency assertion), the workers with their job specifications (one row per fit, resumable), and the result tables underlying all five tables and all three figures; the file-by-file inventory is in Online Resource 1.

Code availability. As under Data availability. The exact package set is pinned in renv.lock (R 4.5.3, 179 packages); the software environment and the reproduction procedure are described in Online Resource 1.

Author contributions. Jyotheeshwar Akshay Ravikumar: conceptualization, methodology, software, formal analysis, investigation, data curation, validation, visualization, writing — original draft, writing — review and editing. B. Anbarasan: supervision, validation, writing — review and editing. Both authors read and approved the final manuscript.

Use of artificial intelligence. During the preparation of this manuscript the authors used Claude (Anthropic), a large language model assistant, to draft and revise manuscript text, to write and review analysis and plotting code, and as a critical review aid, including the diagnosis of pipeline and convergence problems reported under “A note on the workflow” and “Convergence, chain length and parameter boundaries”. The authors reviewed and edited all output, verified that every reference is real and checked, and take full responsibility for the content of this article. No content was generated without author oversight, and the tool is not listed as an author. All figures are data visualisations rendered by the deposited code directly from the deposited result tables, so their underlying data, methods and computational outputs can be inspected and verified; no figure contains AI-generated image content, and no image representing primary observed or experimental data appears in this article.

References

1. Mager DE, Jusko WJ (2001) General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. J Pharmacokinet Pharmacodyn 28:507–532. https://doi.org/10.1023/a:1014414520282
2. Gibiansky L, Gibiansky E, Kakkar T, Ma P (2008) Approximations of the target-mediated drug disposition model and identifiability of model parameters. J Pharmacokinet Pharmacodyn 35:573–591. https://doi.org/10.1007/s10928-008-9102-8
3. Gibiansky L, Gibiansky E (2009) Target-mediated drug disposition model: relationships with indirect response models and application to population PK–PD analysis. J Pharmacokinet Pharmacodyn 36:341–351. https://doi.org/10.1007/s10928-009-9125-9
4. Peletier LA, Gabrielsson J (2012) Dynamics of target-mediated drug disposition: characteristic profiles and parameter identification. J Pharmacokinet Pharmacodyn 39:429–451. https://doi.org/10.1007/s10928-012-9260-6
5. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25:1923–1929. https://doi.org/10.1093/bioinformatics/btp358
6. Wieland FG, Hauber AL, Rosenblatt M, Tönsing C, Timmer J (2021) On structural and practical identifiability. Curr Opin Syst Biol 25:60–69. https://doi.org/10.1016/j.coisb.2021.03.005
7. Chis OT, Banga JR, Balsa-Canto E (2011) Structural identifiability of systems biology models: a critical comparison of methods. PLoS ONE 6:e27755. https://doi.org/10.1371/journal.pone.0027755

8. Kreutz C (2018) An easy and efficient approach for testing identifiability. Bioinformatics 34:1913–1921. https://doi.org/10.1093/bioinformatics/bty035
9. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J (2014) Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30:1440–1448. https://doi.org/10.1093/bioinformatics/btu006
10. Villaverde AF, Barreiro A, Papachristodoulou A (2016) Structural identifiability of dynamic systems biology models. PLoS Comput Biol 12:e1005153. https://doi.org/10.1371/journal.pcbi.1005153
11. Murphy SA, Van der Vaart AW (2000) On profile likelihood. J Am Stat Assoc 95:449–465. https://doi.org/10.1080/01621459.2000.10474219
12. Kreutz C, Raue A, Timmer J (2012) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Syst Biol 6:120. https://doi.org/10.1186/1752-0509-6-120
13. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol 3:e189. https://doi.org/10.1371/journal.pcbi.0030189
14. Brekkan A, Jönsson S, Karlsson MO, Hooker AC (2018) Reduced and optimized trial designs for drugs described by a target mediated drug disposition model. J Pharmacokinet Pharmacodyn 45:637–647. https://doi.org/10.1007/s10928-018-9594-9
15. Choi S, Park S, Jung J, Baek S, Lim HS (2025) Population pharmacokinetics/pharmacodynamics analysis confirming biosimilarity of SB16 to reference denosumab. Front Pharmacol 16:1631034. https://doi.org/10.3389/fphar.2025.1631034
16. Savic RM, Karlsson MO (2009) Importance of shrinkage in empirical Bayes estimates for diagnostics: problems and solutions. AAPS J 11:558–569. https://doi.org/10.1208/s12248-009-9133-0
17. Ahn JE, Karlsson MO, Dunne A, Ludden TM (2008) Likelihood based approaches to handling data below the quantification limit using NONMEM VI. J Pharmacokinet Pharmacodyn 35:401–421. https://doi.org/10.1007/s10928-008-9094-4
18. Beal SL (2001) Ways to fit a PK model with some data below the quantification limit. J Pharmacokinet Pharmacodyn 28:481–504. https://doi.org/10.1023/a:1012299115260
19. Jaber MM, Cheng S, Brundage RC (2021) Evaluation of bias in weighted residual calculations when handling below the limit of quantification data using Beal's M3 method. CPT Pharmacometrics Syst Pharmacol 10:275–278. https://doi.org/10.1002/psp4.12616
20. Fidler M, Wilkins JJ, Hooijmaijers R, Post TM, Schoemaker R, Trame MN, Xiong Y, Wang W (2019) Nonlinear mixed-effects model development and simulation using nlmixr and related R open-source packages. CPT Pharmacometrics Syst Pharmacol 8:621–633. https://doi.org/10.1002/psp4.12445
21. Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Comput Stat Data Anal 49:1020–1038. https://doi.org/10.1016/j.csda.2004.07.002
22. Kim T, Cha J, Jeon H, Yun H-Y, Lim D, Kim JK (2026) Exact formula of the total quasi-steady state approximation in competitive target-mediated drug disposition. bioRxiv (preprint). https://doi.org/10.64898/2026.01.10.698771

23. Mentré F (1997) Optimal design in random-effects regression models. Biometrika 84:429–442. https://doi.org/10.1093/biomet/84.2.429
24. Nyberg J, Bazzoli C, Ogungbenro K, Aliev A, Leonov S, Duffull S, Hooker AC, Mentré F (2015) Methods and software tools for design evaluation in population pharmacokinetics–pharmacodynamics studies. Br J Clin Pharmacol 79:6–17. https://doi.org/10.1111/bcp.12352
25. Ravikumar JA, Anbarasan B (2026) Simulation and estimation code and per-fit result tables for a simplified denosumab-like quasi-steady-state TMDD identifiability study [Data set]. Zenodo. https://doi.org/10.5281/zenodo.23761636

Figures

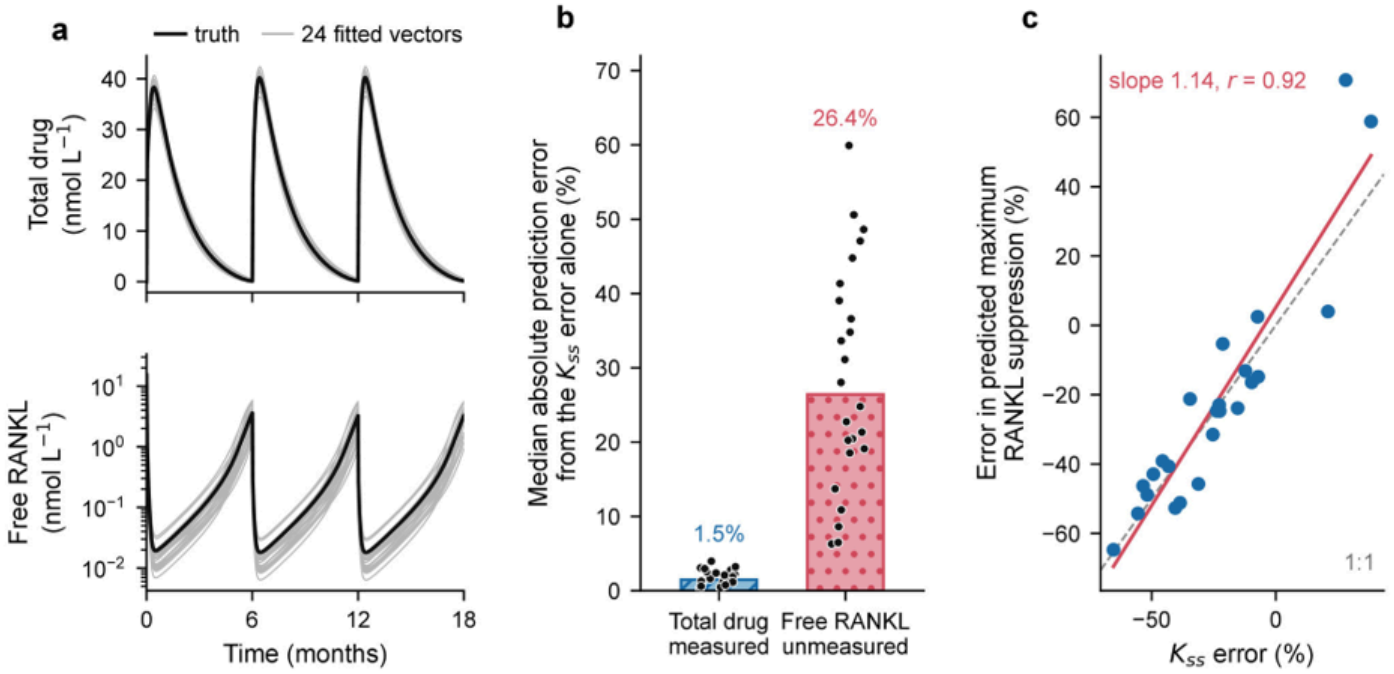


Figure 1

Prediction impact of the K_{SS} error. Throughout, the 24 fitted vectors are the 24 true-value-start fits of Online Resource 1, Table S14, the set underpinning the first row of Table 4. (a) Typical-subject total drug concentration and free RANKL under those 24 vectors (light traces) against the truth (heavy); no random effects, free RANKL on a logarithmic axis. (b) Median absolute prediction error caused by the K_{SS} error alone — the true vector with only K_{SS} replaced by each fitted value; bars are medians, points the 24 replicates. (c) Error in predicted maximum RANKL suppression against the K_{SS} error, with the fitted line and a 1:1 reference. n = 24 fitted vectors; no new fits

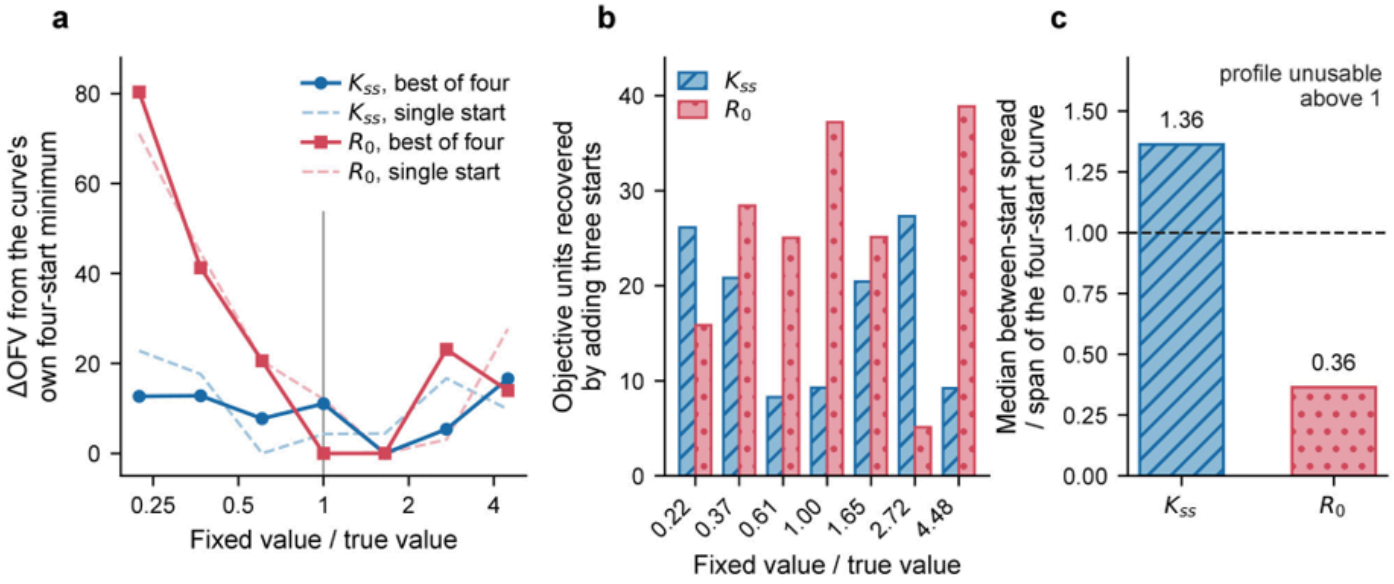


Figure 2

Multistart profile likelihoods for K_{ss} and R_0. (a) Objective against the fixed value, relative to each parameter's own four-start minimum; solid lines with markers are the best of four starts, the pale dashed lines the single start at each point. (b) Objective units recovered by adding three starts, at each of the seven grid points for each parameter. (c) Median between-start spread at one grid point divided by the span of that parameter's four-start curve; a ratio above one means the curve is not a profile at that start count. n = 4 starts at each of 7 grid points per parameter, one dataset

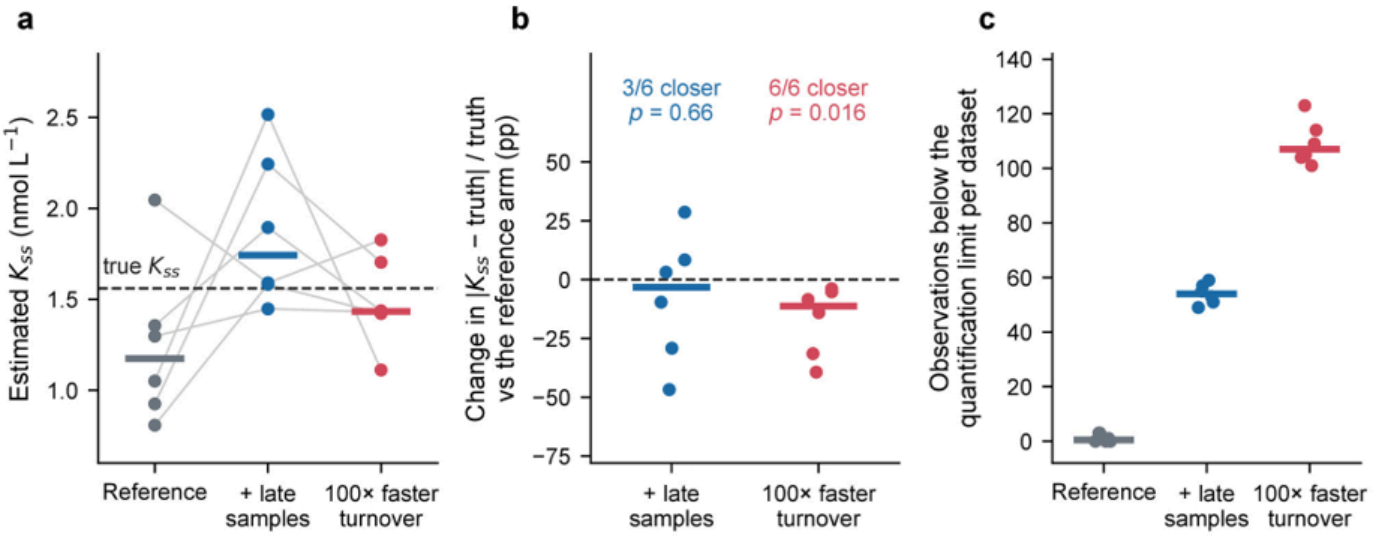


Figure 3

Design arms. (a) Estimated K_{ss} in each arm; grey lines join the three fits of one replicate, which share subject-level random effects, and heavy bars are arm medians. (b) Within-dataset change in |K_{ss} − truth|/truth relative to the reference arm, in percentage points; p values are exact one-sided sign tests on

six paired differences. (c) Observations below the quantification limit per dataset — the reference and turnover arms have 720 nominal observations and the late arm 900. Arm labels abbreviate the three designs of Table 5. *n* = 6 paired datasets per arm

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf