

Bayesian dose–response mixed modelling

Application to perturbation screening data

Maryam Kazemi Naeini | Senior Statistician

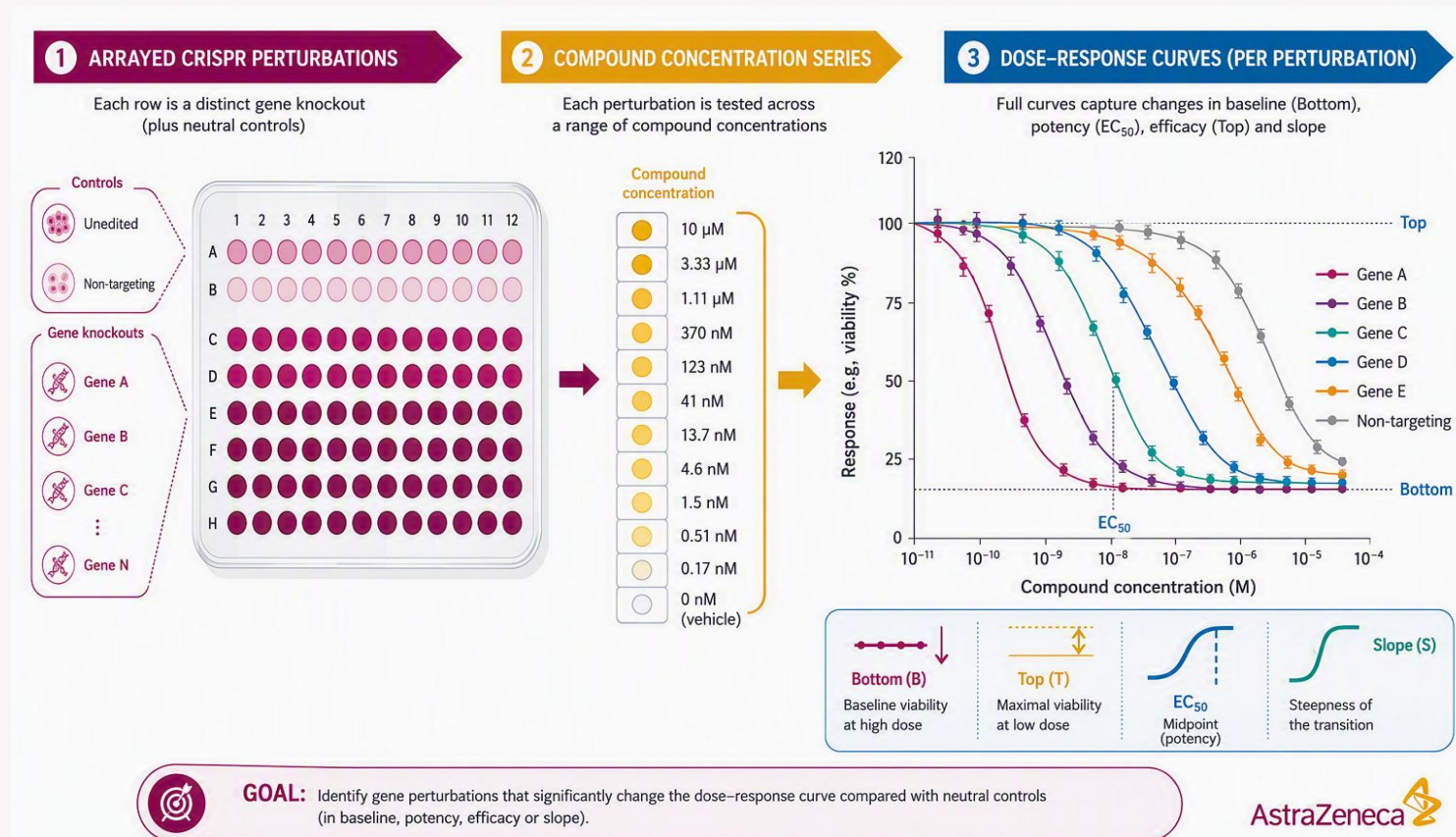
Joff Jones · Maximilian Wechsung · Salome Adam · Elissavet Kentepozidou · Bogumil Kaczkowski

Discovery Sciences, R&D, AstraZeneca, Cambridge, UK



What does an arrayed CRISPR dose–response screen measure?

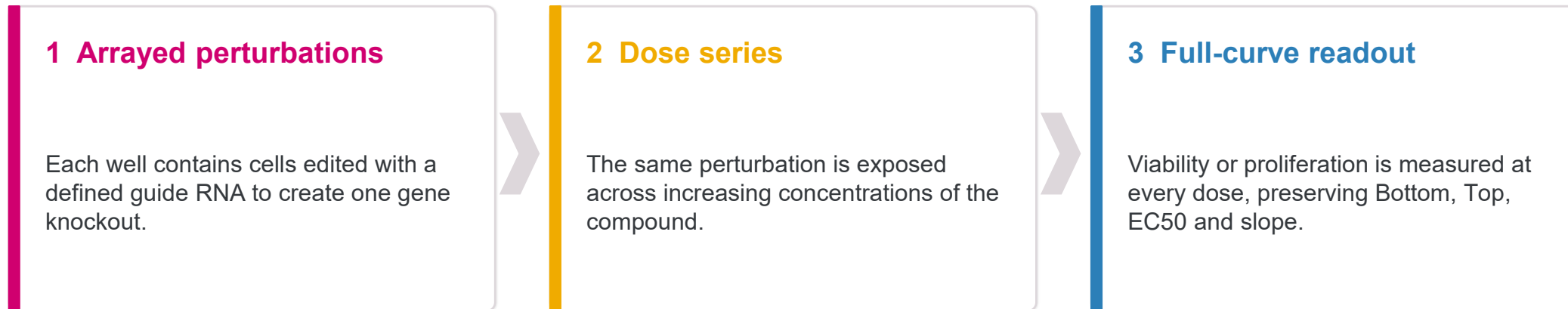
Perturbations crossed with a concentration series reveal how the whole response curve changes.



Objective: compare each perturbation with the neutral control across baseline, efficacy, potency and slope.

What does an arrayed CRISPR dose–response screen measure?

Defined gene perturbations crossed with a compound concentration series reveal how the complete response curve changes.



Analysis objective Identify perturbations that differ from the neutral control in potency, efficacy, baseline or slope.

Biological questions we typically seek to address via this type of assays are:

- To characterize genetic determinants of *sensitisation* and *resistance* to drug compounds. This is done by knocking out genes in cell lines (or potentially organoids), treating the cells with different doses of the drug(s) of interest, and monitoring the viability / proliferation of cells.

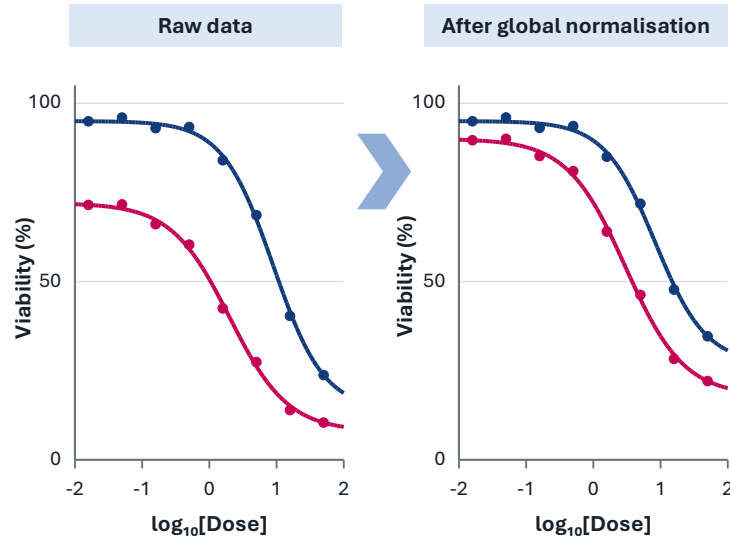


THREE MAJOR THREATS TO DOSE-RESPONSE INFERENCE

Why endpoint summaries or naïve normalisation can be misleading

1 NORMALISATION CAN DISTORT A CURVE

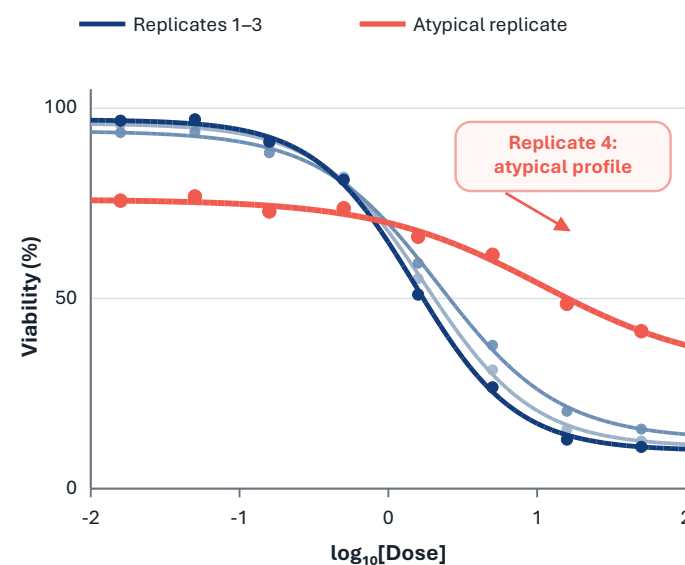
Global or plate-level scaling can shift Bottom and dynamic range.



A genuine early viability loss can be treated as a technical offset. Endpoint comparisons become sensitive to the chosen normalisation.

2 ONE REPLICATE CURVE CAN BE ATYPICAL

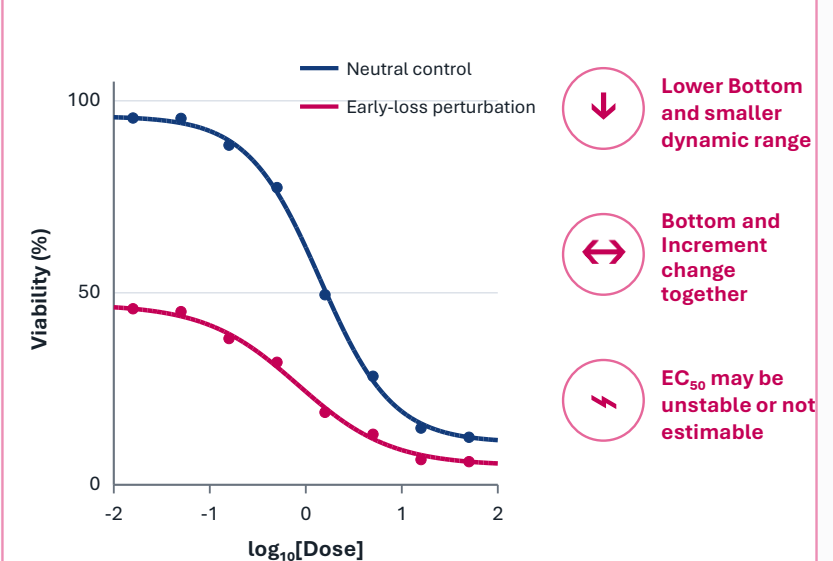
A whole dose-response profile can disagree with the other biological replicates.



- Replicate-wise EC_{50} or $E(d_{max})$ can be shifted by one atypical curve
- The disagreement spans several doses—not one isolated well
- Endpoint-only summaries hide the curve-level pattern

3 EARLY VIABILITY LOSS IS GENUINE BIOLOGY

Some perturbations suppress viability before the transition region.



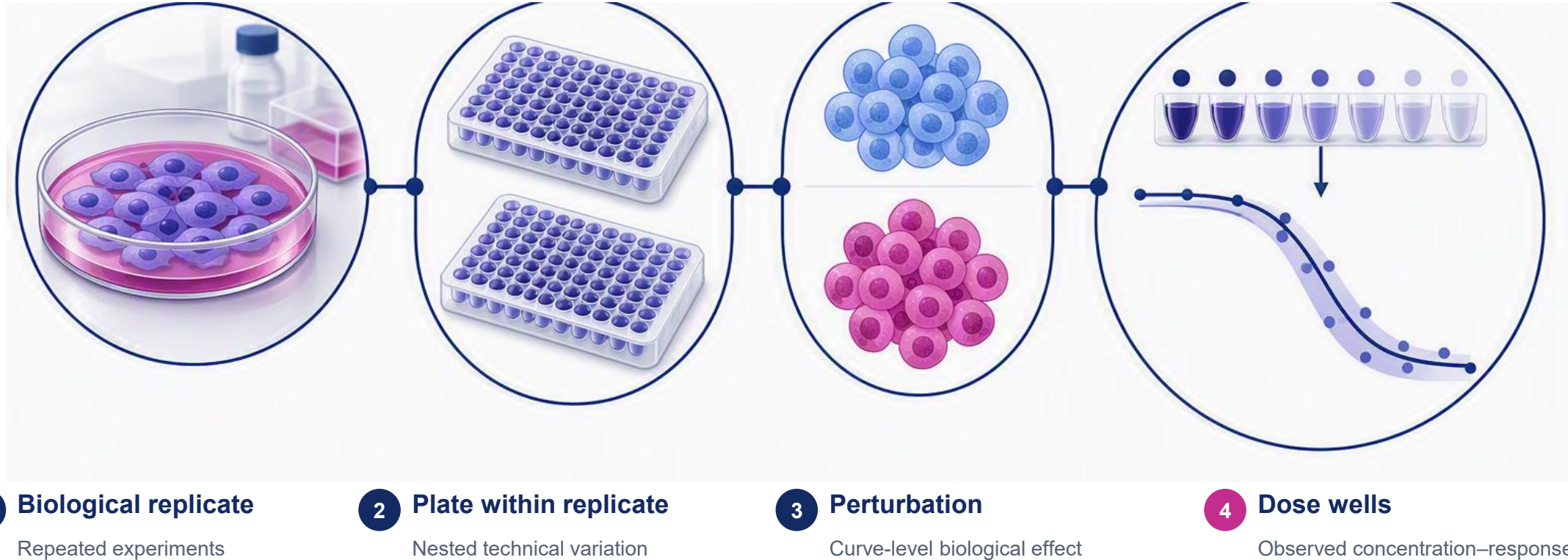
Removing or normalising this away discards real biological signal.



WE NEED A JOINT MODEL THAT SEPARATES BIOLOGY FROM TECHNICAL VARIATION AND PRESERVES THE SIGNAL.

The hierarchy exists before the curve is fitted

The model should follow how observations were generated—not correct the structure afterwards.



Joint modelling keeps all four levels connected and propagates variation through the dose–response inference.



The familiar 4PL remains biologically interpretable

The extension adds perturbation effects and technical hierarchy directly to curve parameters.

$$\mu(d) = B + \frac{T - B}{1 + (d/EC_{50})^S}$$

Log-parameterisation

keeps positive quantities in valid ranges while retaining biologically meaningful parameters.

B

Bottom: lower asymptote

T

Top: upper asymptote

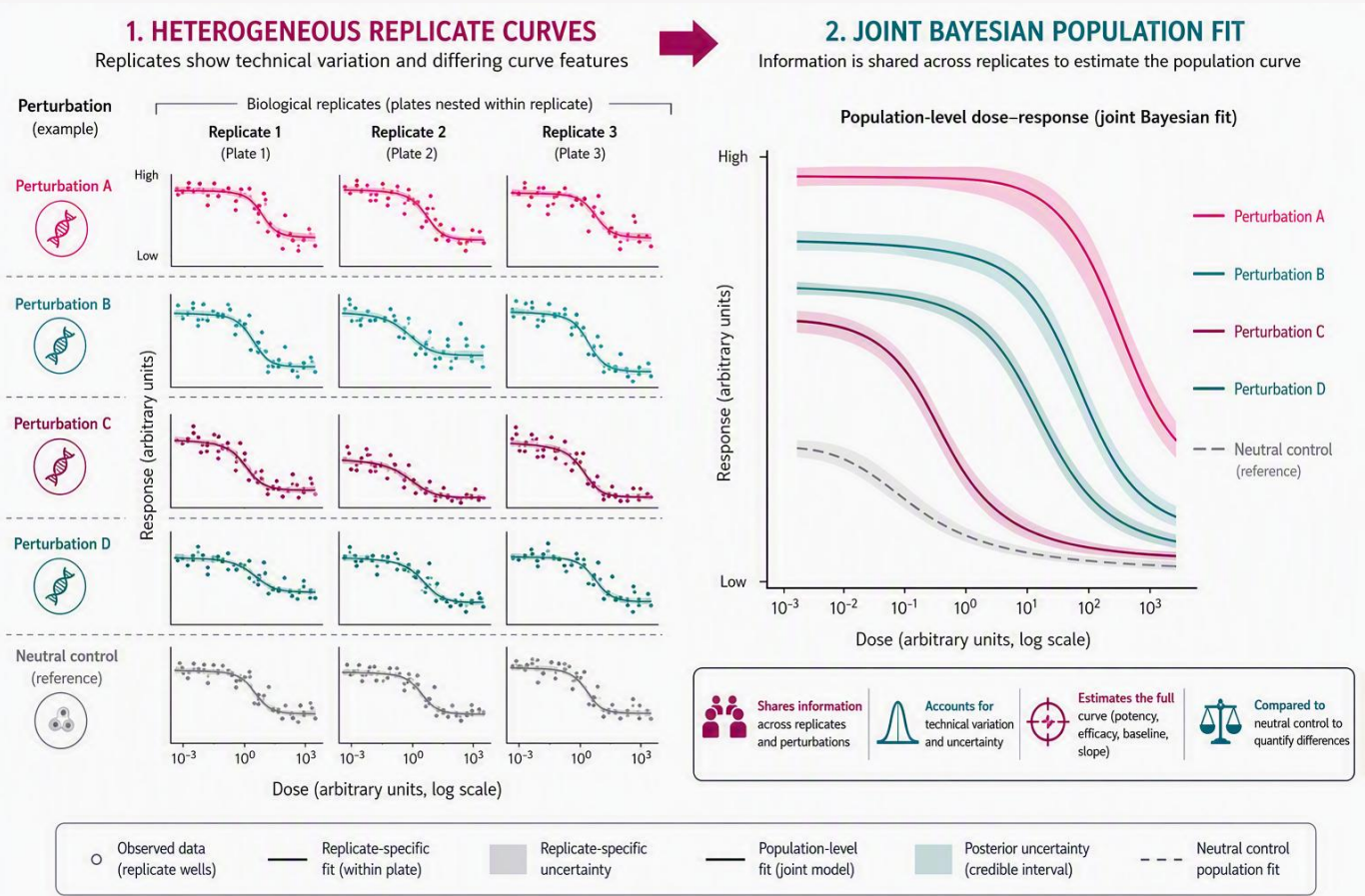
EC₅₀ Midpoint: potency**S**

Slope: transition steepness



From replicate curves to a population-level Bayesian fit

Heterogeneous replicate data become an uncertainty-aware population comparison.



All identities remain blinded; curves and uncertainty are shown schematically.

One model separates biology from technical variation

Perturbation effects and nested random effects enter the nonlinear mean function together.

$$\begin{aligned}B_{\text{KO},p} &= \exp(b_0 + b_{\text{KO}} + u_p^{(B)}) \\I_{\text{KO},p} &= \exp(a_0 + a_{\text{KO}} + u_p^{(I)}) \\EC_{50,\text{KO},p} &= \exp(g_0 + g_{\text{KO}} + u_p^{(E)}) \\S_{\text{KO}} &= \exp(d_0 + d_{\text{KO}})\end{aligned}$$

PLATE p NESTED WITHIN BIOLOGICAL REPLICATE

Biological signal

KO effects enter Bottom, Increment, EC50 and slope.

Technical hierarchy

Plate-within-replicate effects absorb structured offsets.

Posterior uncertainty

Every contrast propagates parameter, plate and residual uncertainty.



Each parameter controls an interpretable curve feature

KO effects capture biological shifts; plate effects absorb structured technical variation.

$$\log(\mu_i) = \log \left(\exp(b_0 + b_{\text{KO},i} + b_{\text{plate},i}) + \frac{\exp(a_0 + a_{\text{KO},i} + b_{\text{plate},i})}{1 + \left(\frac{\exp(\log_dose_i)}{\exp(g_0 + g_{\text{KO},i} + b_{\text{plate},i})} \right) \exp(d_0 + d_{\text{KO},i})} \right)$$

Bottom (baseline) · $\exp(b_0 + b_{\text{KO}} + b_{\text{plate}})$

b_0 global baseline · b_{KO} KO-specific baseline shift
 b_{plate} nested plate-within-replicate vertical offset

Increment (Top – Bottom) · $\exp(a_0 + a_{\text{KO}} + b_{\text{plate}})$

a_0 global log-increment · a_{KO} KO-specific shift
 b_{plate} adjusts dynamic range by plate

Slope (Hill) · $\exp(d_0 + d_{\text{KO}})$

d_0 global log-slope · d_{KO} KO-specific slope shift

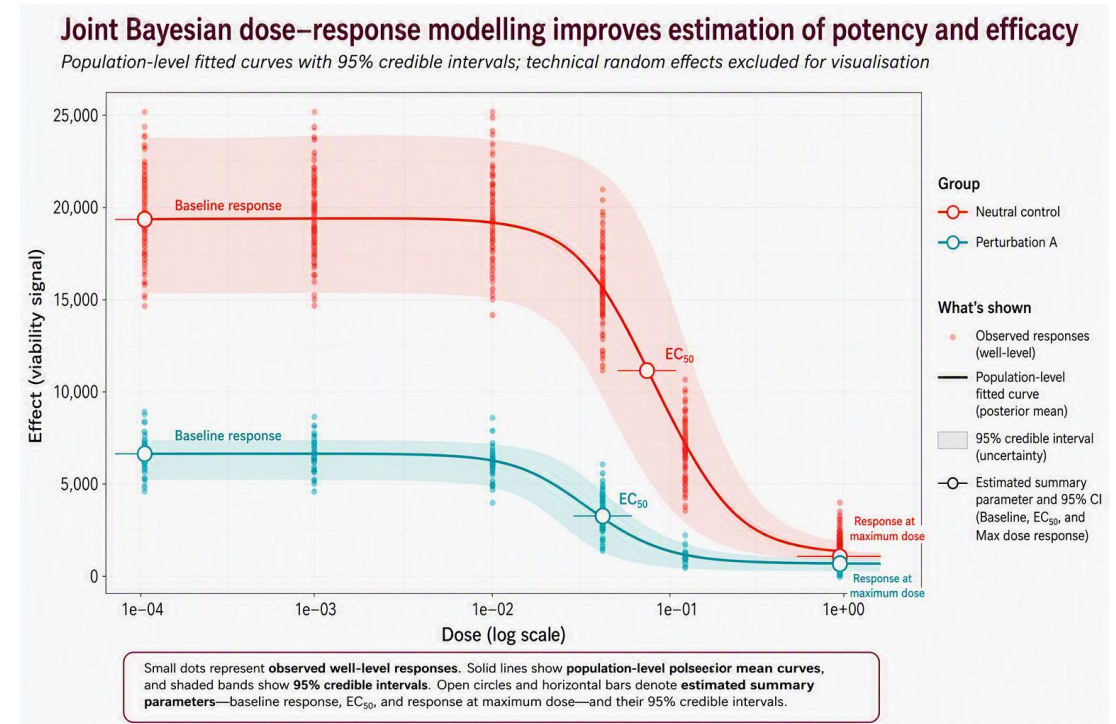
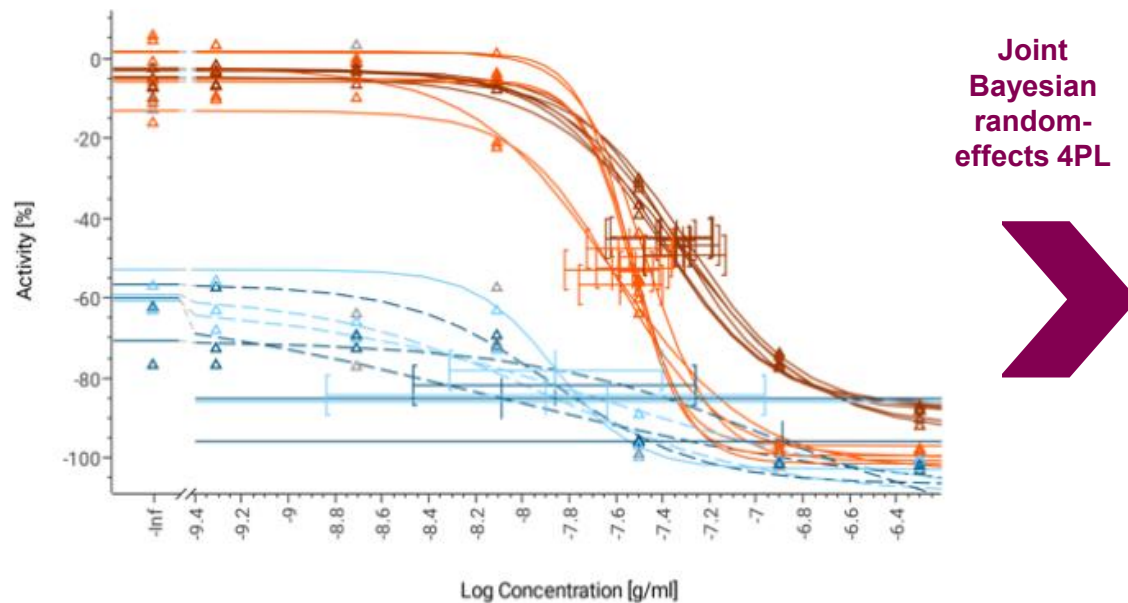
EC50 (potency) · $\exp(g_0 + g_{\text{KO}} + b_{\text{plate}})$

g_0 global log-EC50 · g_{KO} KO-specific shift
 b_{plate} adjusts horizontal position by plate



From replicate curves to a population-level Bayesian fit

The joint model converts heterogeneous replicate curves into an uncertainty-aware population comparison.



Six advantages of joint Bayesian modelling

The value is not only a better curve fit; it is a more reliable decision process.

All wells together

Uses one nonlinear mixed model rather than disconnected replicate fits.

Partial pooling

Stabilises weakly identified EC50, E(dmax) and slope estimates.

Direct uncertainty

Produces KO-versus-control contrasts directly from posterior draws.

Outlier resilience

Isolated wells have less leverage because evidence is pooled across doses and replicates.

Correction in-curve

Replicate and plate effects are handled where the observations are modelled.

Validated decisions

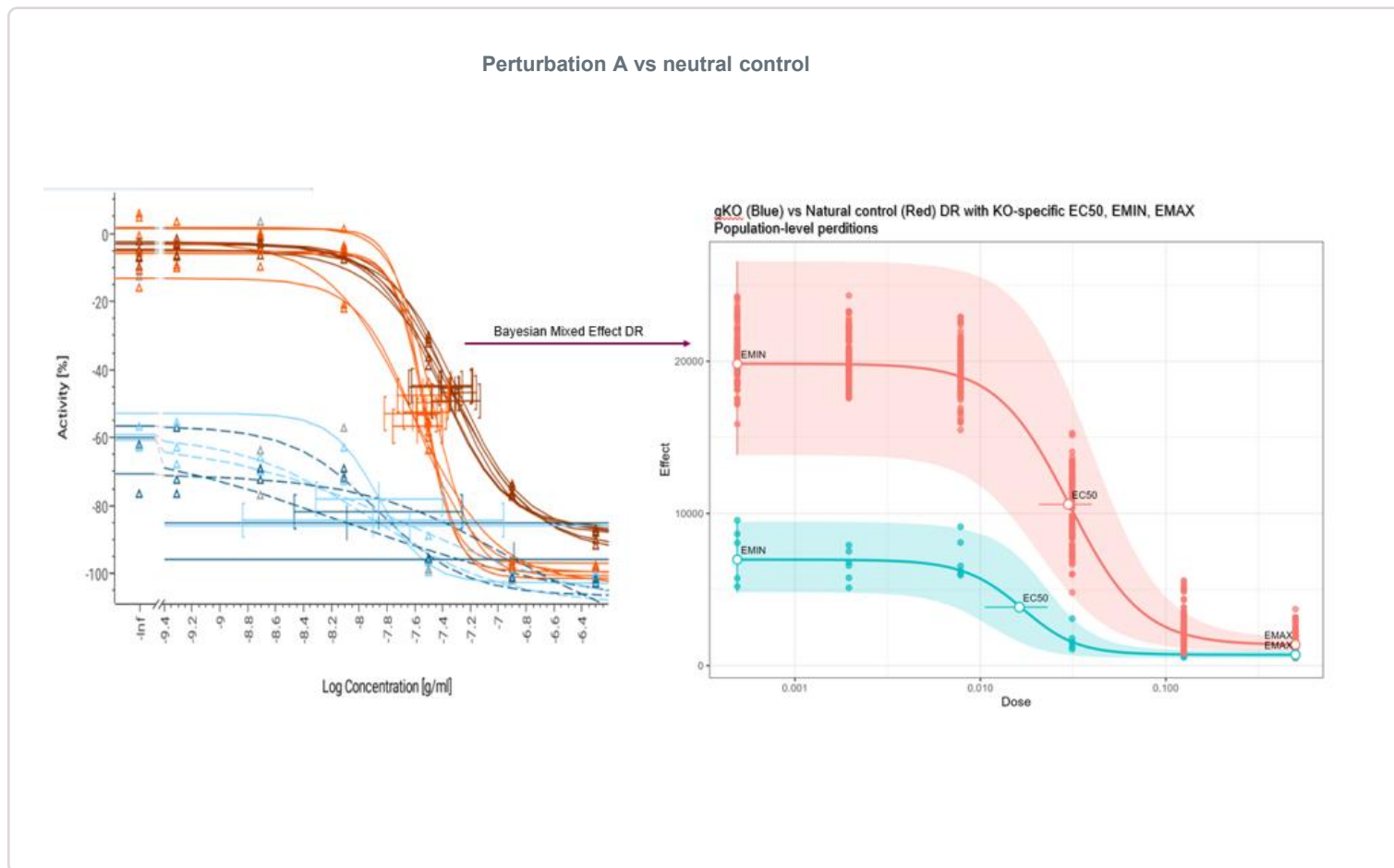
Predictive checks and uncertainty intervals support transparent hit calls.

Bottom line: more recoverable signals, fewer outlier-driven calls and uncertainty that reflects the full experiment.



Early signal loss no longer forces an “unanalysable” call

Pooling across wells and replicates stabilises the baseline and dynamic range when the midpoint is poorly bracketed.



What is recovered

An EC50 estimate with an interval even when individual replicate transitions are incomplete.

Why it works

The model shares information while retaining gene-specific curve parameters.

Decision benefit

Potential actives are not discarded solely because a replicate-level endpoint failed.



Bayesian modelling improves the robustness of E(dmax) and EMAX hit calling

Full-curve information and in-model technical adjustment reduce dependence on isolated endpoint behaviour.

WHY THE BAYESIAN APPROACH IS MORE ROBUST

1

Endpoint extremes have less leverage

E(dmax) is supported by the complete dose-response profile—not only a few wells at the highest concentration.

2

All doses and replicates inform the estimate

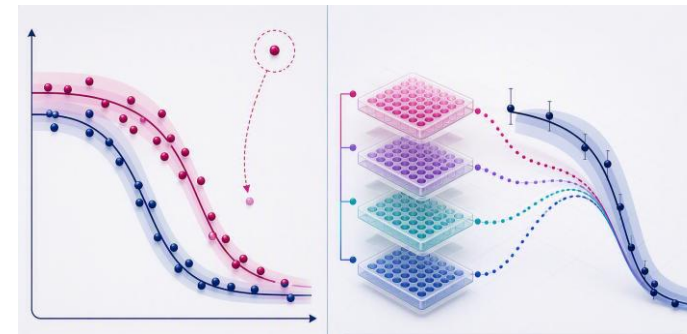
Joint fitting carries partial information forward and stabilises the curve-supported efficacy estimate.

3

Technical variation is modelled inside the curve

Plate and biological-replicate effects are accounted for before perturbation contrasts and hit calls are derived.

FROM ENDPOINT-DRIVEN TO CURVE-SUPPORTED DECISIONS



MORE RELIABLE HIT CALLING

More robust E(dmax) estimation

supports EMAX calls that reflect the full experiment

TAKEAWAY Bayesian joint modelling gives a more robust basis for E(dmax) estimation and EMAX hit calling.



Posterior contrasts turn fitted curves into transparent hit decisions

Define practically meaningful effect and probability thresholds before unblinding.

Potency

EC50 ratio or difference versus neutral control

Baseline / efficacy

Bottom, Top or E(dmax) differences versus control

Uncertainty

95% credible interval and posterior probability

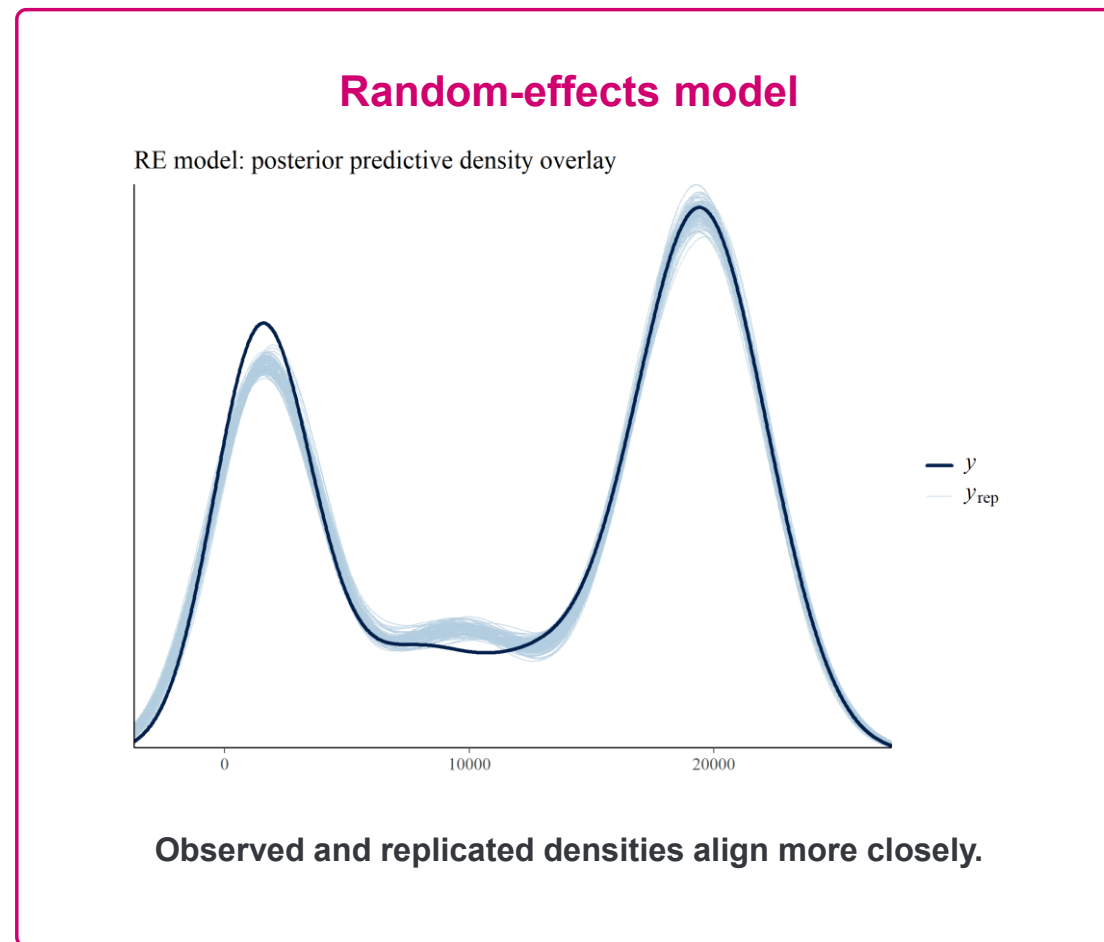
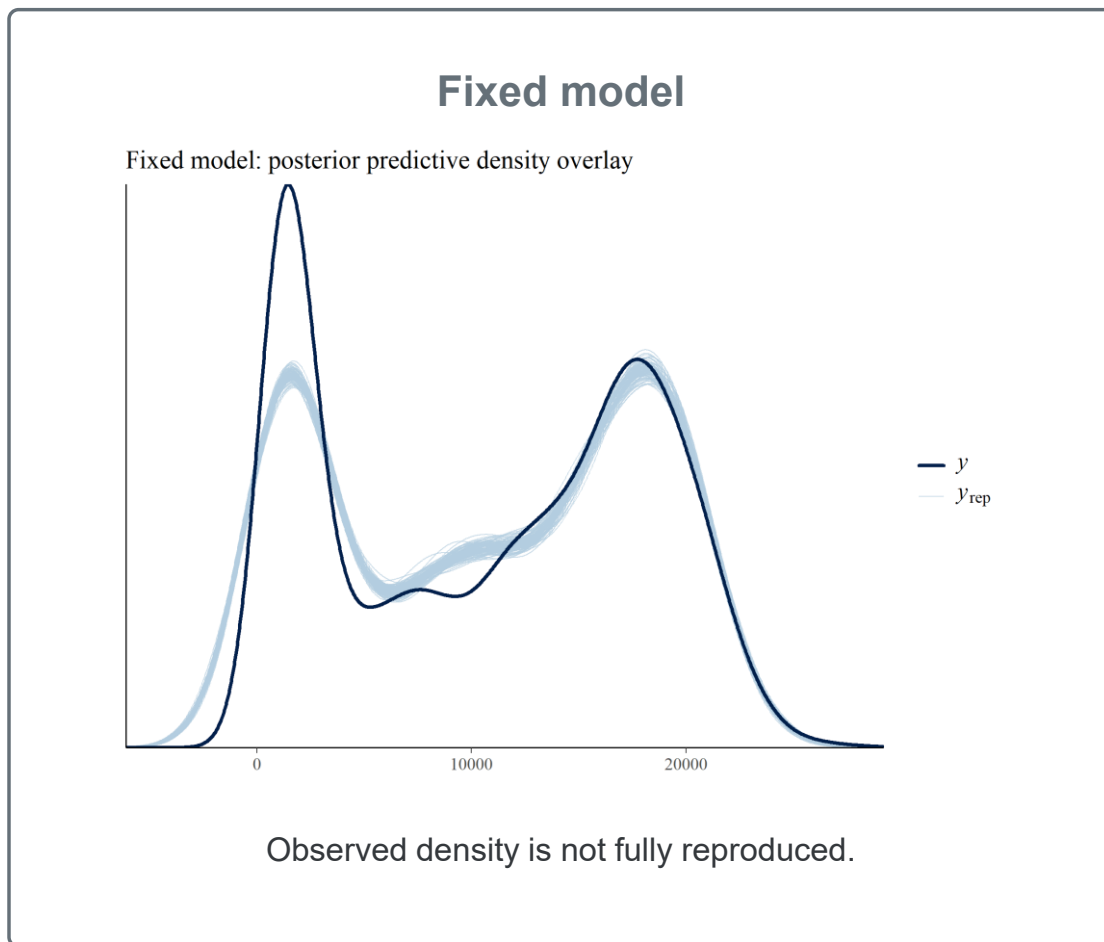
A hit is not a single fitted number.

Effect size + uncertainty + prespecified threshold



The random-effects model better reproduces the observed distribution

Posterior predictive checks assess whether replicated data resemble the assay data.



Model the biology and experimental variation together

- **Joint Bayesian dose–response modelling uses the full curve and the experimental hierarchy to make better use of the available information.**
- **Model biology directly**
Gene-KO effects are estimated on the **4PL curve parameters**, while **biological replicate and plate-within-replicate effects** capture experimental variation.
- **Recover information from difficult curves**
Joint fitting can retain information from **incomplete or early signal-loss curves**, where separate per-replicate curve fitting may fail or produce unstable estimates.
- **Preserve potency and efficacy signals**
Using the complete dose–response profile provides more stable **EC₅₀ and E(dmax)/Emax comparisons with credible intervals**, reducing the chance that potentially active perturbations are lost simply because an endpoint cannot be estimated reliably in one replicate.
- **Reduce sensitivity to atypical replicates**
Information is shared across replicates, so an atypical replicate curve has less influence on the overall perturbation-level inference.
- **Account for technical variation in-curve**
Plate and replicate effects are modelled within the dose–response model, rather than corrected only after endpoints have been extracted.

