The New England Journal of Statistics in Data Science Volume 0, 1–16 (2026)
DOI: https://doi.org/10.51387/26-NEJSDS109

Bayesian Design and Analysis of Mouse Clinical Trials with MiniPDX and PDX Models for Translational Oncology

Jiaxin LIU[1], Diandong JIANG[1], Xinmeng ZHANG, Hongkui CHEN, Shizhu ZHAO, Danyi WEN, and Yuan JI*

Abstract

The MiniPDX model is a cost-effective preclinical platform that replicates human tumor microenvironments in mice, enabling rapid drug response evaluation in precision oncology. In contrast, the patient-derived xenograft (PDX) model offers a more comprehensive but slower and costlier approach by engrafting human tumors into immunodeficient mice to evaluate efficacy in a physiologically relevant context. In this study, we present an integrated Bayesian framework that supports both separate and joint analyses of MiniPDX and PDX data. Our models borrow information across biomarker-defined subgroups and between assays, improving the efficiency and precision of drug effect estimation. Furthermore, we implement an adaptive sample size re-estimation procedure for MiniPDX experiments, allowing for more flexible designs and optimized resource utilization. Together, these methods provide a Bayesian framework that improves estimation efficiency for small-sample MiniPDX and PDX experiments and offers a principled adaptive sample-size rule, helping to identify biomarker-defined subpopulations with distinct predictive profiles for treatment response.

KEYWORDS AND PHRASES: Drug development, Preclinical, Sample size re-estimation.

1. INTRODUCTION

Patient-derived xenograft (PDX) models[3, 7, 2] have become a cornerstone of translational oncology research due to their ability to recapitulate patient tumor heterogeneity and clinical drug responses. Expanding on this concept, mouse clinical trials (MCTs)[13], in which large cohorts of PDX models are treated in a manner analogous to human clinical trials (Figure 1), allow rigorous evaluation of therapeutic efficacy across diverse tumor populations, thereby improving the predictive power and clinical translatability of preclinical findings. However, conventional PDX studies are time-intensive, often limiting their utility for rapid decision-making. To address this limitation, the MiniPDX model [17] has been developed as a rapid and clinically relevant platform that preserves the predictive value of PDX while significantly shortening study duration. Together, MiniPDX and MCT approaches enhance the translational relevance of preclinical studies and may increase the likelihood of successful clinical outcomes.

Figure 1 is a design schematic illustrating the per-cohort allocation principle of the MCT: patients (or their derived models) are allocated to cohorts by HER2 expression level, and within each cohort the tumor is evaluated against vehicle control. Cohort sizes in Figure 1 are illustrative only — they are not pre-specified by the schematic, and the "Cohort 4" depiction with a single mouse is a visual placeholder rather than a sample-size statement. The actual analysis sample sizes for our study are reported in Table 1 (enrollment) and Table 2 (analyzed models). The MCT setting we consider does not use within-cohort randomization between treatment and control in the way a human trial would; allocation is by HER2 cohort, and each model contributes paired treatment and control arms.

PDX and MiniPDX studies differ in design and biological fidelity. A PDX study engrafts an intact patient-tumor fragment subcutaneously into immunodeficient mice and follows tumor volume over weeks to months, preserving tumor architecture and a partial microenvironment but at substantial time and cost. A MiniPDX study [17] dissociates the same tumor into a cell suspension, encapsulates the cells, implants the capsules, and reads out cell viability after roughly ten days, trading architectural and immune-stromal fidelity for speed and lower cost. The two assays are therefore complementary rather than interchangeable: MiniPDX enables rapid screening but typically with smaller per-cohort sample sizes and without the immune compartment, while PDX provides a more physiologically faithful but slower readout. Differences in fidelity and in time-to-readout motivate joint analysis of the two assays, which is one of the contributions of this paper. However, preclinical studies often involve small sample sizes, imposing challenges for robust statistical inference and sample size planning. Simple metrics such as mean response or p-values may fail to support meaningful decision-making, and advanced approaches such as Bayesian models have not been extensively applied in this setting.

*Corresponding author.
[1]These authors contributed equally to this work.

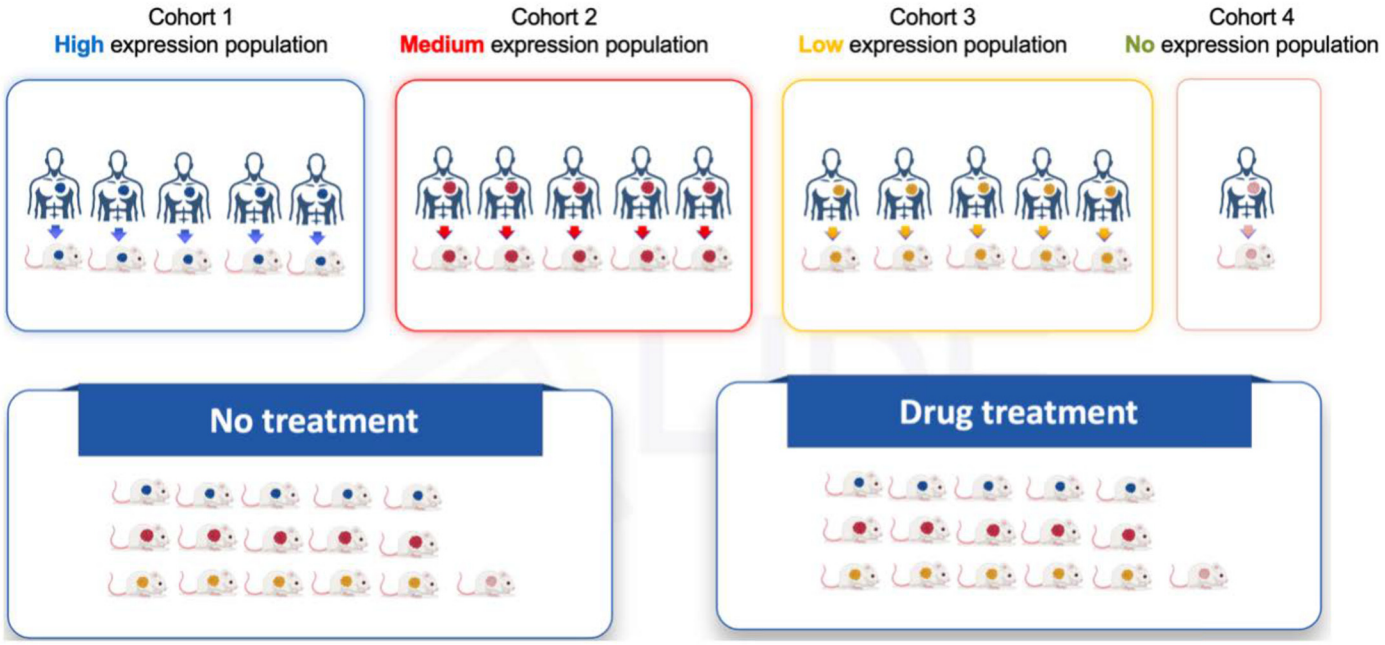


Figure 1: Fundamental Design of Mouse Clinical Trial.

In recent years, Bayesian methods and models have gained popularity in clinical trial designs [1, 9] and analysis due to their flexibility, ability to incorporate prior knowledge, and robustness in handling complex data structures. These methods are now widely used in various aspects of clinical trials, such as adaptive trial designs, dose-finding studies, and borrowing information from historical data to improve the efficiency and reliability of clinical research. However, Bayesian applications in preclinical research have not been widely implemented with few exceptions. Walley et al. [14] utilized a Bayesian meta-analytic predictive (MAP) approach to incorporate historical control data from repeated preclinical trials, either as an external control or to inform a prior distribution. Rokita [12] conducted genomic profiling of childhood tumor patient-derived xenograft (PDX) models to inform rational clinical trial design, employing a Bayesian hierarchical model to analyze gene expression across different histologies. Kelter [8] explored how Bayesian statistical decision criteria can effectively reduce the false discovery rate in preclinical animal research by comparing four types of Bayesian decision criteria with traditional p-value-based methods. Zhang et al. [16] introduced a three-step Bayesian modeling approach for analyzing heterogeneous responses in preclinical mouse tumor models. This approach utilized Bayesian mixed-effects models and Bayesian ordered logistic regression to predict individual tumor response categories, showcasing the flexibility and power of Bayesian frameworks in preclinical settings. However, none of these works jointly model MiniPDX and PDX assays under a single hierarchical structure with explicit assay-specific and group-specific effects, which is the gap we address in Section 2.3. The resulting two-assay model lets the smaller MiniPDX cohorts borrow strength from the larger PDX cohorts while still allowing the assays to differ systematically through an assay-specific effect α_i.

Adaptive designs are structured decision-making frameworks in clinical trials, enabling modifications such as adjusting treatment assignments, altering sample sizes, adding or removing treatments, or deciding trial continuation based on accumulated data [5]. In Bayesian adaptive designs, these modifications are guided by continuously updated posterior distributions of parameters, allowing optimal decisions informed by current data and predefined utility criteria. In some settings, adaptive designs may offer more flexibility [10], and is more efficient [6] and cost-effective [15] while providing high-power results. Therefore, adaptive and Beyesian methods have been acknowledged and recommended by regulatory authorities, including the US Food and Drug Administration (FDA) [11].

In this paper, we present a Bayesian framework that improves estimation efficiency for small-sample MiniPDX and PDX experiments by jointly modeling the two assays under a single hierarchical structure with explicit assay-specific and group-specific effects, and that pairs this analysis with a principled adaptive sample-size rule. The Bayesian framing is well suited to small, heterogeneous preclinical samples, where partial pooling across HER2 strata and across assays reduces posterior variance and provides interpretable posterior probabilities to guide go/no-go decisions.

Table 1. Models with various HER2 expression.

Batch	Target	Animal ID	Cancer Type	IHC Grade
Batch one	HER2	LD1-2009-361973	Breast Cancer	+++
	HER2	LD1-2009-362721	Breast Cancer	+++
	HER2	LD1-0038-361928	Colorectal Cancer	+++
	HER2	LD1-0017-200702	Gastric Cancer	+++
	HER2	LD1-0009-410668	Breast Cancer	++
	HER2	LD1-2012-362429	Colorectal Cancer	++
	HER2	LD1-2017-361430	Gastric Cancer	++
	HER2	LD2-0032-200651	Ovarian Cancer	++
	HER2	LD1-2033-362331	Pancreatic Cancer	++
	HER2	LD1-2009-362153	Breast Cancer	+
	HER2	LD1-0012-200620	Colorectal Cancer	+
	HER2	LD1-0017-200697	Gastric Cancer	+
	HER2	LD1-9025-360768	Lung Cancer	+
	HER2	LD1-0033-390792	Pancreatic Cancer	+
	HER2	LD1-2009-362543	Breast Cancer	–
	HER2	LD1-0017-200688	Gastric Cancer	–
Batch two	HER2	LD1-0038-361855	Colorectal Cancer	+++
	HER2	LD1-2009-361836	Breast Cancer	+++
	HER2	LD2-0017-201149	Adenocarcinoma	+++
	HER2	LD2-0017-201204	Gastric Cancer	+++

2. MATERIALS AND METHODS

2.1 Animal Experiments and Models

2.1.0.1. Animals All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at LIDE Biotech and conducted in accordance with applicable national and international regulations for laboratory animal welfare. Mouse strains, vendors, and housing at the AAALAC-accredited LIDE Biotech facility are reported in Appendix C.

2.1.0.2. Establishing the PDX Model A total of 16 patient-derived xenograft (PDX) models were selected for this study, comprising five breast cancer, three colorectal cancer, four gastric cancer, one lung cancer, two pancreatic cancer, and one ovarian cancer model (Table 1). Model selection was based on the expression of human HER2, as determined by immunohistochemistry (IHC) using H-score assessment. The selected models represented a spectrum of HER2 expression levels, including high (+++), medium (++), low (+), and negative (0). Mice were administered a single intravenous dose of trastuzumab deruxtecan (also known as Enhertu, DS-8201) at 10 mg/kg.

2.1.0.3. PDX Drug Sensitivity Assays The Tumour Growth Inhibition (TGI) value is calculated as follows:

$$\%\,\mathrm{TGI} = 1 - \frac{\text{Tumor Volume at treatment group at day 21} - \text{Tumor Volume at treatment group at day 0}}{\text{Tumor Volume at control group at day 21} - \text{Tumor Volume at control group at day 0}} \tag{2.1}$$

In equation (2.1), "Day 21" denotes the protocol-specified efficacy endpoint, which for the present study is fixed at 21 days after the start of treatment, and "Day 0" is the start-of-treatment baseline. TGI is therefore computed once at the single Day 21 endpoint by comparing the per-arm mean tumor volumes at Day 21 to baseline at Day 0; we do not model the longitudinal volume trajectory in this paper. The PDX experimental unit is the mouse: each PDX model contributes 6 mice in total, split as 3 in the treatment arm and 3 in the vehicle (control) arm.

2.1.0.4. Establishing the MiniPDX Model Tumor tissue from each PDX model was dissociated and encapsulated in OncoVee MiniPDX® capsules (LIDE Biotech), and the capsules were implanted subcutaneously in 5-week-old nu/nu mice with two capsules (tubes) per mouse. The full wet-lab protocol — tumor selection criteria, HBSS washes, collagenase digestion, magnetic-bead depletion of stromal and blood cells, and the encapsulation procedure — is given in Appendix C.

2.1.0.5. MiniPDX Drug Sensitivity Assays Mice bearing MiniPDX capsules were treated with trastuzumab deruxtecan for 10 days. After 10 days the capsules were removed and tumor cell proliferation was quantified by relative luminance unit (RLU) using a luminescent cell-viability assay; reagent and instrument details are reported in Appendix C. Tumor cell growth inhibition (TCGI) (%) was calculated using the formula:

$$\text{TCGI (\%)} = \left(1 - \frac{\text{Mean RLU of the treatment group on day 10} - \text{Mean RLU on day 0}}{\text{Mean RLU of the vehicle group on day 10} - \text{Mean RLU on day 0}}\right) \times 100\% \tag{2.2}$$

Each experiment was done in sextuplicate and mean values were reported. Concretely, each MiniPDX patient was run with 6 mice in total: 3 in the vehicle arm and 3 in the treatment arm, with 2 capsules (tubes) per mouse, giving 6 capsules per arm. The TCGI in equation (2.2) is computed from the mean RLU per arm across these capsules, so the response unit at the readout level is the capsule and "sextuplicate" refers to the six capsules per arm (3 mice × 2 capsules). A positive or negative drug response was considered present if TCGI was $\geq 45\%$ or $< 45\%$, respectively.

We interpret %TGI and %TCGI in drug-sensitivity terms as follows. Values greater than 1 (i.e., greater than 100%) indicate net tumor regression in the treatment arm relative to baseline, that is, cytoreduction beyond growth arrest; a value of 1 corresponds to tumor stasis, in which the treated tumor showed no net growth from baseline during the treatment window; values strictly between 0 and 1 indicate partial growth inhibition, in which the treated tumor grew but more slowly than control; a value of 0 indicates no treatment effect (the treated tumor grew at the same rate as control); and negative values correspond to tumor escape, where the treated tumor grew faster than control. The threshold TCGI $\geq 45\%$ used here is a positive-response cutoff that is standard in the MiniPDX literature; we apply the same threshold to TGI for consistency across assays in the joint model.

2.2 Bayesian Hierarchical Model for Single Assay

Let $j = 1, \ldots, J$ index the HER2 groups. Within each group j, let y_{jk} denote the average tumor proliferation outcome across mouse replicates under the treatment (y_{jk1}) and control (y_{jk0}) for patient $k = 1, \ldots, n_j$. Specifically, y_{jk1} is the average treatment-arm response for patient k in HER2 group j: for PDX it is the mean tumor volume at Day 21 across the 3 treatment-arm mice, and for MiniPDX it is the mean RLU at Day 10 across the 6 treatment-arm capsules (3 mice × 2 capsules each). y_{jk0} is the corresponding vehicle-arm (control) average computed the same way. Each patient therefore contributes both a treatment arm and a vehicle arm, and the two arms are paired within patient: y_{jk1} and y_{jk0} come from the same tumor source under the same protocol, with allocation to cohorts determined by HER2 expression rather than by within-cohort randomization.

Let z_{jk} denote the outcome corresponding to TGI or TCGI, as defined in (2.1) and (2.2), respectively. It follows that

$$z_{jk} = \left(1 - \frac{y_{jk1}}{y_{jk0}}\right) \times 100\%.$$

Here z_{jk} is the patient-specific drug response, summarizing relative tumor inhibition for patient k in HER2 group j: $z_{jk} > 1$ corresponds to net tumor regression, $z_{jk} = 1$ to tumor stasis (no net growth from baseline), $0 < z_{jk} < 1$ to partial growth inhibition, $z_{jk} = 0$ to no treatment effect, and $z_{jk} < 0$ to tumor escape. We take $z_{jk} \geq 0.45$ as a positive-response threshold consistent with the per-tumor TCGI cutoff used in the MiniPDX literature; this corresponds to roughly 55% of the control growth being suppressed in the treatment arm. We propose a Bayesian hierarchical model for analyzing the data $\{z_{jk}\}$. We assume

$$z_{jk} \sim N(z_{jk}; \mu_j, \sigma_j^2) \tag{2.3}$$

where $N(\cdot; \mu_j, \sigma_j^2)$ is a normal distribution with mean μ_j and variance σ_j^2. Moreover, to enable shrinkage and partial pooling, we specify the patient-level parameters to follow hyperpriors

$$(\mu_j, \sigma_j^2) \sim N(\mu_j; \mu_0, \tau^2) \cdot \text{TCauchy}(\sigma_j; a, b, (0, \infty)). \tag{2.4}$$

Here, $\text{TCauchy}(\cdot; a, b, (0, \infty))$ is a half-Cauchy distribution. In addition, let

$$(\mu_0, \tau^2) \sim N(\mu_0; s_0, r_0^2)) \cdot \text{TCauchy}(\tau; a_\tau, b_\tau, (0, \infty)) \tag{2.5}$$

where s_0, r_0^2, a, b, a_τ and b_τ are specified to enable vague but realistic hyperpriors, ensuring flexibility without imposing strong assumptions.

The prior choices in equations (2.3)–(2.4) are weakly informative. We use a half-Cauchy prior on σ_j and on the between-patient scale τ, following the recommendation of Gelman [4] for variance components in hierarchical models with a small number of groups, which gives a heavier-tailed alternative to a half-normal and behaves well when the number of groups is small. We use a vague normal prior on μ_0 centered at a mid-range TCGI value and with realistic spread, so that the prior is non-informative within the support of the data without sending posterior mass to implausible negative-inhibition extremes.

2.3 Bayesian Hierarchical Model for Two Assays

Let $i = 1, 2$ index experimental assays (MiniPDX and PDX), $j = 1, \ldots, J$ index HER2 expression groups, and $k = 1, \ldots, n_{ij}$ index patients within each assay-by-group combination. We propose a Bayesian hierarchical model designed to integrate information from both the MiniPDX and PDX assays simultaneously. This structure facilitates borrowing strength across assays, allowing the MiniPDX assay to leverage richer information from the more established PDX data. Specifically, we define assay-specific effects as α_i and HER2 group-specific effects as β_j, thus capturing systematic variations attributable to assay conditions and biological differences between HER2 groups.

Formally, our hierarchical model for the TCGI or TGI (μ_{ij}) outcome z_{ijk} is specified as:

$$z_{ijk} \sim \mathcal{N}\left(\alpha_i + \beta_j, \sigma_{ij}^2\right) \quad k = 1, \ldots, n_{ij}, \quad i = 1, 2, \quad j = 1, \ldots, J, \tag{2.6}$$

where the mean structure explicitly partitions the expected outcome into assay-specific(α_i) and group-specific(β_j) contributions and define $\mu_{ij} = \alpha_i + \beta_j$ to be the total effect. Without loss of generality, we use μ_{ij} to represent the treatment effect in both single-assay and multi-assay contexts. To ensure meaningful interpretation and identifiability, we impose the sum-to-zero constraint $\sum_{i=1}^{I} \alpha_i = 0$ on the assay effects.

We further model the variance parameters hierarchically. To stabilize variance estimation and allow for flexible heterogeneity within and between assays, we introduce a hierarchical log-variance structure:

$$\begin{aligned} \log(\sigma_{ij}^2) \mid \log(\sigma_i^2) &\sim \mathcal{N}(\log(\sigma_i^2), \tau_{\text{within}}^2), \\ \log(\sigma_i^2) \mid \log(\sigma_0^2) &\sim \mathcal{N}(\log(\sigma_0^2), \tau_{\text{between}}^2) \end{aligned} \tag{2.7}$$

where τ_{within}^2 and τ_{between}^2 captures the variability within assay and between assay, and $\sigma_0 \sim \text{TCauchy}(a_{\sigma_0}, b_{\sigma_0}, (0, \infty))$.

At the top level, we place hyperpriors on the assay-level parameters α_i and group-level mean parameters β_j, modeling their variances directly:

$$\alpha_i \mid \tau_\alpha^2 \sim \mathcal{N}(\alpha_0, \tau_\alpha^2), \tag{2.8}$$

$$\beta_j \mid \beta_0, \tau_\beta^2 \sim \mathcal{N}(\beta_0, \tau_\beta^2), \tag{2.9}$$

$$\text{with constraint:} \quad \sum_{i=1}^{I} \alpha_i = 0. \tag{2.10}$$

Since we have imposed the constraint on the assay effects, we do not assign a hyperprior to α_0, but instead fix its value at 0. To robustly model uncertainty around hyperparameters, we employ weakly informative hierarchical priors. Specifically, we set:

$$\tau_\alpha \sim \text{TCauchy}(a_\alpha, b_\alpha, (0, \infty)), \tag{2.11}$$

$$(\beta_0, \tau_\beta^2) \sim \mathcal{N}(s_\beta, r_\beta^2) \cdot \text{TCauchy}(\tau_\beta; a_\beta, b_\beta, (0, \infty)) \tag{2.12}$$

where s_β, r_β^2, a_α, b_α, a_{σ_0}, b_{σ_0}, a_β and b_β are specified to enable vague but realistic hyperpriors, ensuring flexibility without imposing strong assumptions.

The two-assay model imposes the sum-to-zero constraint $\sum_i \alpha_i = 0$ to make the assay effects identifiable under the additive coding $\mu_{ij} = \alpha_i + \beta_j$. The hierarchical log-variance specification $\log(\sigma_{ij}^2) \sim \mathcal{N}(\log(\sigma_i^2), \tau_{\text{within}}^2)$ and $\log(\sigma_i^2) \sim \mathcal{N}(\log(\sigma_0^2), \tau_{\text{between}}^2)$ stabilizes variance estimation when each assay-by-group cell has only a handful of patients. As in the single-assay model, the variance scales $\tau_{\text{within}}, \tau_{\text{between}}, \sigma_0, \tau_\beta$ carry weakly informative half-Cauchy priors per Gelman [4], and the top-level mean β_0 has the same vague-normal treatment we used for μ_0 in the single-assay model. The assay-effect SD τ_α is not identifiable from the data with only $I = 2$ assays under the sum-to-zero constraint and is therefore fixed at $\tau_\alpha = 2.5$ in the present analysis (see Appendix B for the sensitivity check).

2.4 Estimation and Inference

We fit all Bayesian hierarchical models in R 4.4 via the `rjags` package (version 4.x), which interfaces to JAGS for Markov chain Monte Carlo (MCMC) sampling. For both the single-assay and two-assay models we run 3 parallel chains with an adaptation phase of 10,000 iterations, a burn-in of 10,000 iterations discarded, and 100,000 retained iterations per chain thinned by 10, yielding 10,000 effective post-thin draws per chain and 30,000 draws in total. Random-number seeds 1001, 1002, and 1003 are used for the three chains so that all results in this paper are reproducible. Convergence is assessed using the Gelman–Rubin potential scale reduction factor $\hat{R}$ (target < 1.05) and effective sample size, computed via `coda::gelman.diag` and `coda::effectiveSize` on the saved 3-chain `mcmc.list`, and trace plots are inspected for all monitored parameters. Posterior probabilities of the form $\Pr(\mu_{ij} \geq t_0 \mid \text{data})$ are computed as the empirical proportion of post-burn-in, thinned MCMC draws satisfying the inequality, and credible intervals are reported as 95% equal-tailed quantile intervals of the posterior draws.

2.5 Adaptive Design and Sample Size Re-Estimation

An initial and usually small number of tumor samples is often used to investigate therapeutics effects using MiniPDX or PDX models. After the data are observed and treatment effects estimated based on the proposed Bayesian hierarchical model, it may be the case that treatment effects for some tumors may exhibit large posterior variabilities and require additional investigation. For others, the treatment effects might be more homogeneous without further exploration. Therefore, we consider an adaptive design and sample size re-estimation (SSR) to allow additional tumor samples to be tested. We define efficacy as the posterior probability

greater than a threshold η that the mean TCGI or TGI (μ_{ij}) is above a pre-determined value t_0.

The proposed design employs a Bayesian adaptive procedure to efficiently estimate the required number of additional patients. To guide this adaptive enrollment, we simulate future patient outcomes assuming the future TCGI or TGI values follow some distribution f, which reflects a realistic anticipated effect. Importantly, as new information becomes available, we can adjust this assumed distribution accordingly, underscoring the advantage of the Bayesian adaptive framework.

We iteratively update the posterior probability $\Pr(\mu_{ij} \geq t_0 \mid \text{data})$ with each simulated set of additional patients. Enrollment simulations continue until this posterior probability surpasses our predefined threshold η. The detailed adaptive sampling procedure is described in Algorithm 1.

Algorithm 1 Sample size re-estimation based on observed data.

Specify $ss = 1$. Let $f =$ some distribution .
for Sim $= 1$ to n_{sim} **do**
  $n(\text{Sim}) \leftarrow 0$
  repeat
    $n(\text{Sim}) \leftarrow n(\text{Sim}) + ss$
    Assume the outcomes $\sim f()$; generate ss new data points y_{ss}.
    Add the new data points to the existing data: $y_{\text{old}} \leftarrow \{y_{\text{old}}, y_{ss}\}$.
    Compute $\xi = \Pr(\mu_{ij} \geq t_0 \mid y_{\text{old}})$.
  until $\xi > \eta$
  Record $n(\text{Sim})$.
end for
Report $\{n(1), \ldots, n(n_{sim})\}$ — compute mean, standard deviation (SD), and density.

In the application below (Section 3), we take $f = \text{InvGamma}(2, 5) - 1$ as a transparent, anticipated-response distribution that reflects an optimistic but plausible projection of future TCGI given the encouraging stage-1 results in the HER2-high group. By placing more mass on larger response values, this choice shrinks the projected required additional sample size relative to a more diffuse alternative, which is appropriate for a feasibility-driven sample-size re-estimation when the stage-1 signal is already favorable. A natural alternative for fully prospective use is the posterior predictive distribution induced by the fitted hierarchical model; we use the parametric f here for clarity of exposition and to make the optimism of the projection explicit.

3. RESULTS

3.1 Efficacy of Mouse Clinical Trial Using MiniPDX and PDX

The dataset consists of measurements evaluating the efficacy of trastuzumab deruxtecan relative to a control in inhibiting tumor growth. Tumors are categorized into three biomarker subgroups based on HER2 expression: High, Medium, and Low. Within each subgroup, tumor tissues from different patients underwent MiniPDX and PDX assays.

The sample flow for the analyses below reflects a two-stage adaptive design and is summarized in Figure 2. In Stage 1 we enrolled 16 PDX models (Table 1, batch one), each assayed on both the PDX (TGI) and MiniPDX (TCGI) platforms. Two of the Stage 1 models have HER2-negative IHC status (HER2 −) and are not included in the Bayesian hierarchical analyses; the mechanism of action of trastuzumab deruxtecan is HER2-driven, so the no-expression cohort is not expected to share the same response distribution as the HER2-positive strata, and pooling it into the same hierarchical structure would mis-specify the borrowing across groups. We retain the HER2-negative cases in the design schematic (Figure 1) and in Table 1 for completeness of the enrolled cohort, but they do not enter equations (2.3)–(2.4) or the two-assay model. The 14 Stage 1 HER2-positive models (HER2 +, ++, +++) contribute paired PDX TGI and MiniPDX TCGI measurements (Table 2) and form the primary analysis cohort for the single-assay (Section 2.2) and two-assay (Section 2.3) hierarchical models. After the Stage 1 analysis, the *MiniPDX* HER2-high posterior probability of efficacy was $\Pr(\mu_1 \geq 0.45 \mid \text{data}) = 0.853$, below our pre-specified target of 0.9, whereas the PDX HER2-high probability was already ≈ 1.00. The MiniPDX assay was therefore the bottleneck for confidence, and the adaptive sample-size re-estimation procedure described in Section 2.5 was triggered specifically for MiniPDX. Under that procedure we then prospectively enrolled 4 additional HER2-high models in Stage 2 (Table 1, batch two), assayed on the MiniPDX platform only (no PDX TGI was collected for Stage 2; Table 7 reports the MiniPDX TCGI values). These 4 Stage 2 models enter the analysis only in Section 3.4, where they illustrate the MiniPDX sample-size re-estimation under the adaptive design.

Table 2 and Figures 3 - 4 present the MiniPDX and PDX data across the HER2 subgroups.

3.2 Single Assay Analyses

We applied the Bayesian hierarchical model presented in Section 2.2 to analyze the MiniPDX and PDX datasets separately. We set the parameters $s_0 = 0.5$, $r_0 = 1$, $a = a_\tau = 0$, $b = b_\tau = 2.5$ for the single assay Bayesian hierarchical model.

3.2.0.1. MiniPDX Data Analysis Based on the probability model and data, we perform MCMC sampling using Rjags and provide the inference in Table 3 and Figure 5. Convergence diagnostics for both single-assay fits are within standard bounds (max $\hat{R} < 1.005$, min ESS $> 6{,}500$ across all monitored parameters). Figure 5 illustrates the posterior

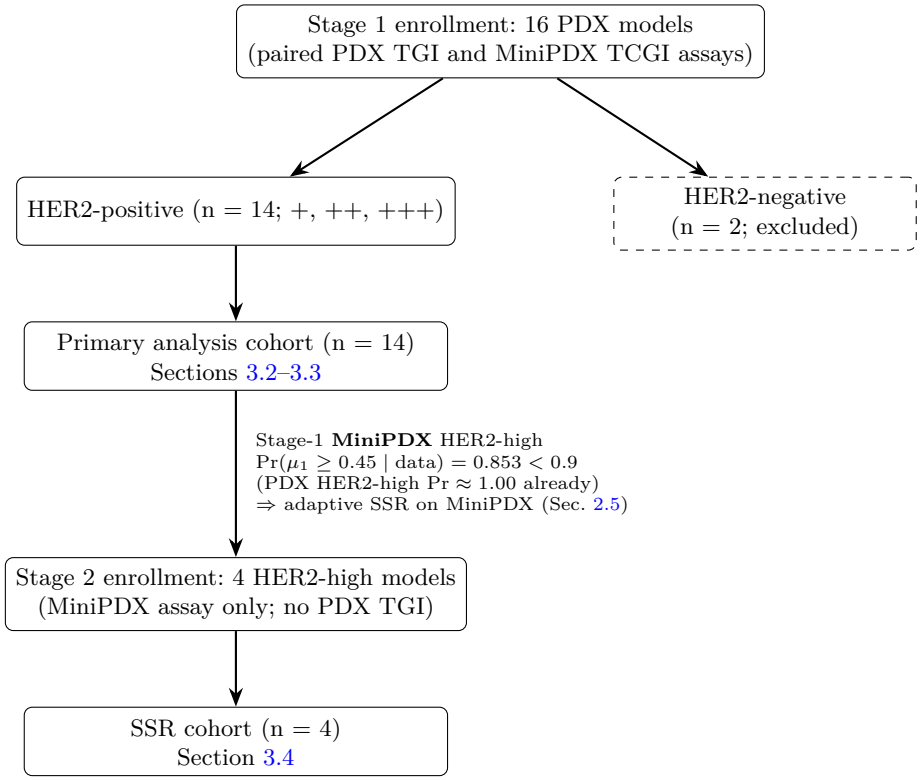


Figure 2: Sample-flow diagram for the analyses. Stage 1 enrolled 16 PDX models, each assayed on both the PDX (TGI) and MiniPDX (TCGI) platforms (Table 1, batch one). The 14 HER2-positive Stage 1 models (HER2 +, ++, +++) form the primary analysis cohort for the single-assay and two-assay hierarchical analyses; the 2 HER2-negative Stage 1 models are excluded from the hierarchical analyses because the trastuzumab deruxtecan mechanism is HER2-driven and the no-expression cohort is not expected to share the same response distribution. After Stage 1, the *MiniPDX* HER2-high posterior probability of efficacy was $\Pr(\mu_1 \geq 0.45 \mid \text{data}) = 0.853$, below the 0.9 threshold, while the PDX HER2-high probability was already ≈ 1.00; the SSR was therefore triggered specifically for the MiniPDX assay (Section 2.5), and Stage 2 prospectively enrolled 4 additional HER2-high models on MiniPDX only (no PDX TGI was collected for Stage 2; Table 1, batch two). These 4 Stage 2 models form the SSR cohort analyzed in Section 3.4. The excluded cohort is shown with a dashed border.

Table 2. TCGI or TGI percentages for MiniPDX and PDX assays, stratified by HER2 expression groups. Each row corresponds to a tumor tissue; blank entries indicate missing measurements.

Tumor	HER2 Group	PDX TGI (%)	MiniPDX TCGI (%)
1	High	134.9	177.00
2	High	119.8	62.17
3	High	117.7	68.50
4	High	105.7	79.33
5	Medium	111.4	65.17
6	Medium	116.1	86.83
7	Medium	94.4	51.50
8	Medium	99.8	
9	Medium	108.3	10.17
10	Low	111.5	6.17
11	Low	105.7	60.00
12	Low	117.7	32.33
13	Low	78.7	48.17
14	Low	53.6	89.33

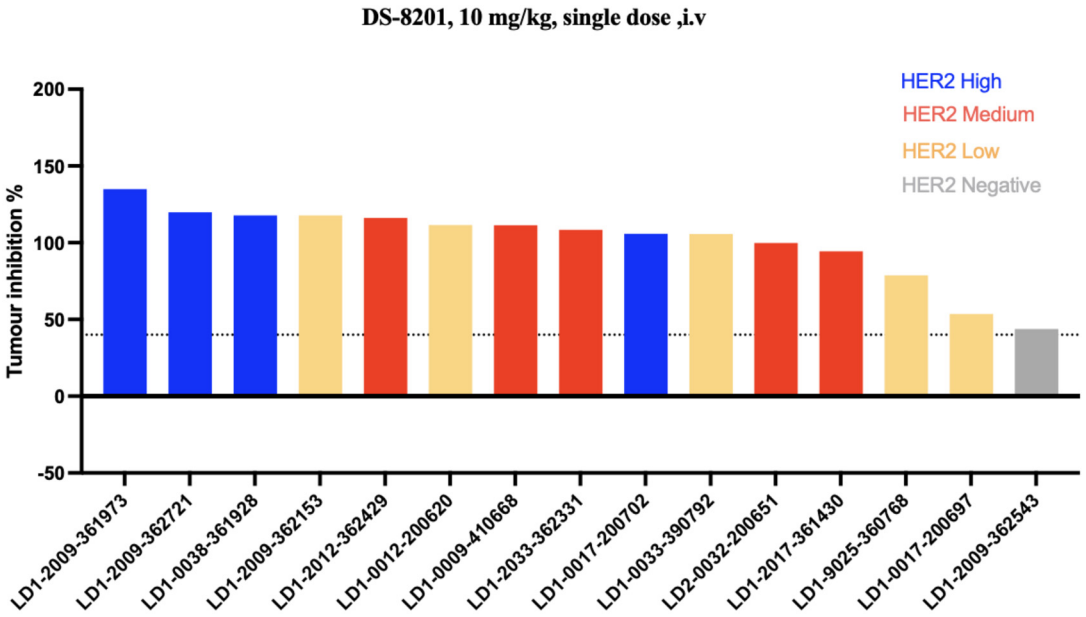


Figure 3: Tumor Inhibition (%) in PDX Models.

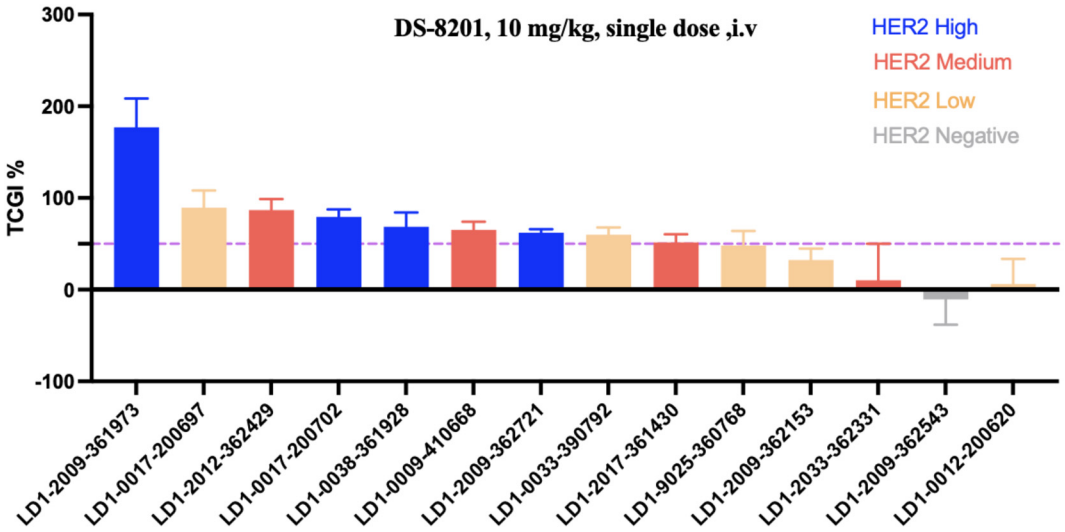


Figure 4: TCGI (%) in MiniPDX Models.

means and Bayesian Credible Intervals for each HER2 subgroup alongside traditional confidence intervals derived from frequentist z-tests from the MiniPDX data. Bayesian intervals consistently demonstrate greater precision, reflected by narrower widths. The HER2 High group has the highest posterior mean TCGI ($\mu_1 = 74.52\%$), indicating greater efficacy. However, the substantial variability, seen by largest posterior variance of μ_1 which is shown in Table 3, suggests the influence of exceptional cases, notably Tumor 1, which has an unusually large TCGI value.

3.2.0.2. PDX Data Analysis Similarly, Figure 6 presents results using the Bayesian hierarchical model from the PDX data.

Table 3. Posterior inference for μ_j under different HER2 groups for MiniPDX data.

HER2 Group	High (μ_1)	Medium (μ_2)	Low (μ_3)	Overall (μ_0)
$\Pr(\mu_j \geq 0.45 \mid \text{data})$	0.853	0.724	0.656	0.729
Posterior Mean of μ_j	0.745	0.555	0.513	0.592
Posterior s.d. of μ_j	0.326	0.225	0.185	0.342

We observe that the HER2 High group again exhibits the highest posterior mean TGI rate ($\mu_1 = 115.10\%$). By borrowing strength across groups through the Bayesian model,

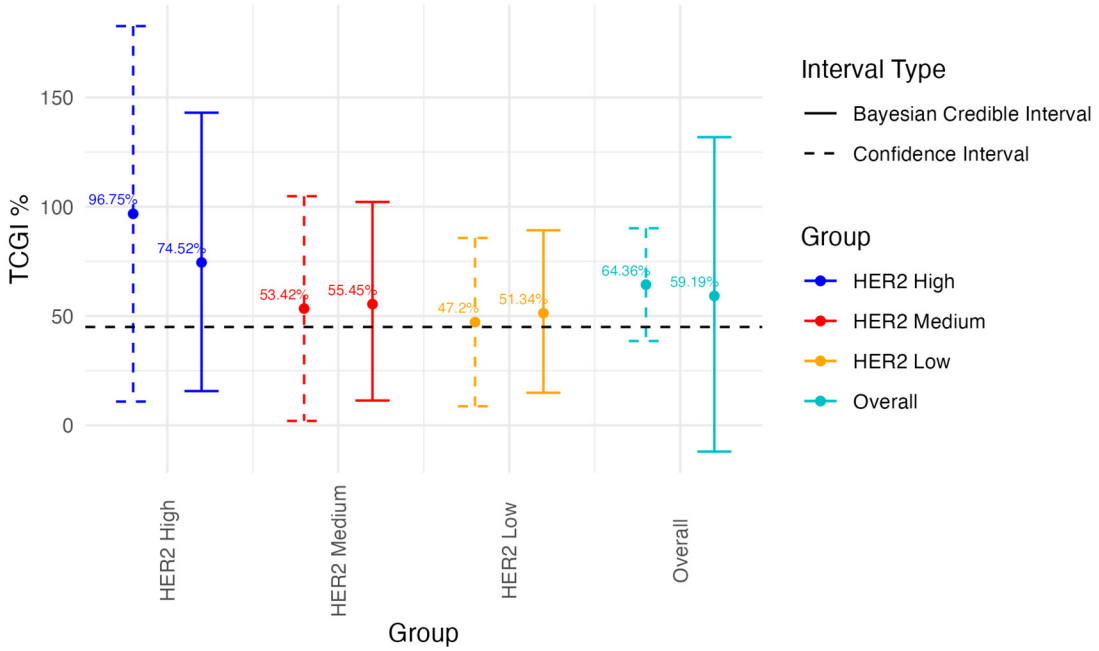


Figure 5: Bayesian Credible Intervals and frequentist Confidence Intervals for the analysis of MiniPDX data. The dots are the posterior means (Bayesian) and sample means (frequentist). The shorter the length of an interval, the more precise the statistical inference.

Table 4. Posterior inference for μ_j under different HER2 groups for PDX data.

HER2 Group	High (μ_1)	Medium (μ_2)	Low (μ_3)	Overall (μ_0)
$\Pr(\mu_j \geq 0.45 \mid \text{data})$	1.00	1.00	0.997	0.977
Posterior Mean of μ_j	1.151	1.064	0.999	1.043
Posterior s.d. of μ_j	0.112	0.063	0.147	0.237

all three subgroups demonstrate a significant treatment effect, as the lower bounds of their credible intervals all exceed 45%. The shrinkage effect is most pronounced in the HER2 Low group. As shown in Table 4, all posterior probabilities exceed 0.9, providing strong evidence of trastuzumab deruxtecan's effectiveness in this assay.

Tables 3 and 4 show that the results from the MiniPDX data are broadly consistent with those from the PDX analysis. In both datasets, the posterior probabilities for μ_1, μ_2, and μ_3 exceeding the efficacy threshold of 0.45 decrease across groups, with the HER2 High group demonstrating the strongest treatment effect. These findings support similar inferences regarding trastuzumab deruxtecan's effectiveness across HER2 expression groups. Although the posterior means differ in magnitude due to differences in assay characteristics, the direction of effect and overall statistical conclusions remain consistent.

3.3 Two Assay Analysis

Based on the results from one assay: across all HER2 groups, the posterior mean TGI values derived from the PDX assay are consistently higher compared with those obtained from the MiniPDX assay. Moreover, the credible intervals for the HER2 High and Medium groups from the PDX assay are narrower, reflecting lower uncertainty. Therefore, leveraging a two assay hierarchical model, as described in Section 2.3, is particularly beneficial. After separately modeling the MiniPDX and PDX data, we apply the model described in Section 2.3 to borrow information across both groups and assays.

For the two assay model, τ_{within} and τ_{between} are estimated from the data (see Appendix A for details), and we set the hyperparameters as $s_\beta = a_\beta = a_{\sigma_0} = 0$, $r_\beta = 1$ and $b_\beta = b_{\sigma_0} = 2.5$. With only $I = 2$ assays, the assay-effect SD τ_α is not identifiable from the data; we therefore fix $\tau_\alpha = 2.5$ rather than sampling it under a half-Cauchy prior. The justification and a sensitivity check at $\tau_\alpha \in \{0.5, 1, 2.5, 5\}$ are reported in Appendix B.

The results are presented in Table 5 and Figure 7. By jointly analyzing the PDX and MiniPDX data, we observe a notable reduction in the posterior variance of μ_{ij}, reflecting that borrowing information across assays improves the precision of the inference. In particular, the posterior probability $\Pr(\mu_{1j} \geq 0.45 \mid \text{data})$ for the HER2 High group increases from 0.853 (when using only MiniPDX data) to 0.958 under the combined model, indicating substantially greater confidence in the treatment effect for this subgroup. Overall, these findings highlight the advantage of hierarchical modeling in leveraging information across assays to yield more robust and reliable estimates. Convergence diagnostics for the

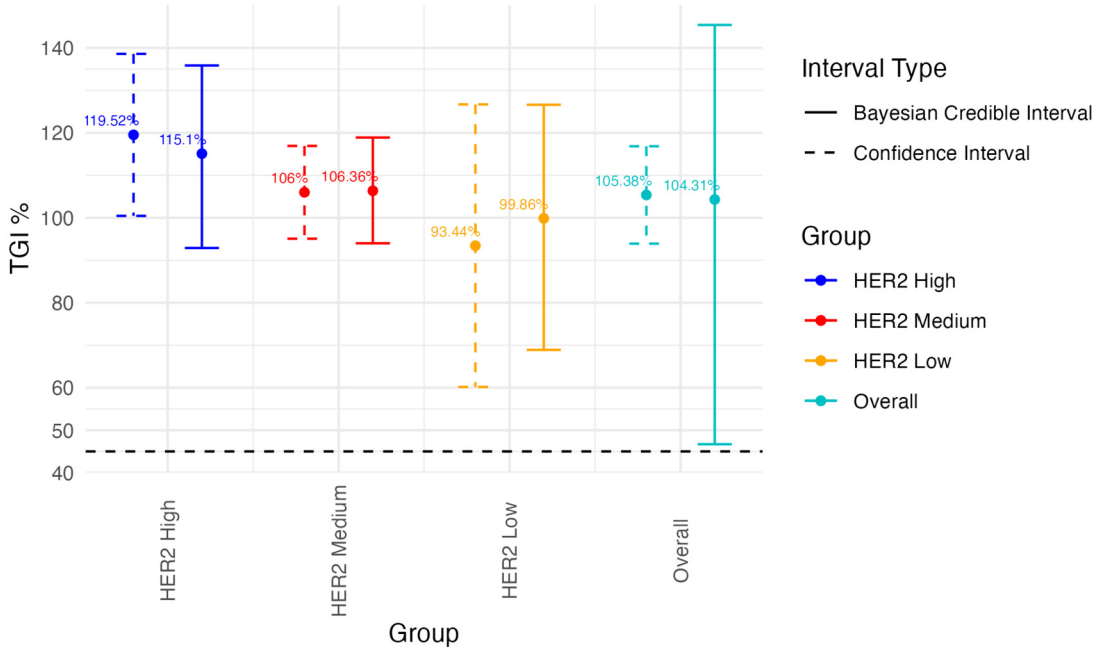


Figure 6: Bayesian Credible Intervals and frequentist Confidence Intervals for the analysis of PDX data. The dots are the posterior means (Bayesian) and sample means (frequentist). The shorter the length of an interval, the more precise the statistical inference.

Table 5. Posterior inference for μ_{ij} under different HER2 groups for MiniPDX and PDX data.

Assay	Statistic	High	Medium	Low
MiniPDX	$\Pr(\mu_{1j} \geq 0.45 \mid \text{data})$	0.958	0.867	0.672
	Post. Mean of μ_{1j}	0.709	0.586	0.504
	Post. s.d. of μ_{1j}	0.151	0.127	0.125
PDX	$\Pr(\mu_{2j} \geq 0.45 \mid \text{data})$	1.000	1.000	1.000
	Post. Mean of μ_{2j}	1.181	1.058	0.976
	Post. s.d. of μ_{2j}	0.083	0.051	0.110

two-assay production fit are within standard bounds for all parameters of scientific interest ($\hat{R} \leq 1.001$, ESS $> 17{,}000$ for α_i, β_j, β_0, and μ_{ij}); the worst $\hat{R}$ across the model is 1.029 on a single σ^2_{ij} cell, well within the < 1.05 target.

To make the efficiency contribution of the Bayesian hierarchical model concrete, we contrast its posterior estimates with simple per-group frequentist sample means and 95% Student-t confidence intervals computed from the same data shown in Table 2. Table 6 reports posterior mean, 95% credible-interval half-width (the half-length of the central 95% posterior quantile interval, computed from the saved MCMC draws), sample mean, and 95% confidence-interval half-width ($t_{n-1,\,0.975} \times s/\sqrt{n}$) for each HER2 group in each assay, under both the single-assay model of Section 2.2 and the joint two-assay model of this section. In the MiniPDX assay, where per-group sample sizes are small and the data are heterogeneous, the single-assay Bayesian credible intervals are already narrower than the frequentist confidence intervals across all three HER2 groups, and the two-assay model tightens them further by roughly a factor of two (for example, the HER2 High half-width drops from 0.637 to 0.296). In the PDX assay the single-assay Bayesian and frequentist widths are comparable, while the two-assay model again tightens the credible intervals across groups (the HER2 Medium half-width drops from 0.124 to 0.102). This pattern is consistent with the intuition that hierarchical shrinkage helps most when per-group information is weakest, and that joint modeling across the two assays buys further efficiency on top of within-assay borrowing.

We confirmed the robustness of the two-assay inference along two prior axes: the variance multiplier on $(\tau_{\text{within}}, \tau_{\text{between}})$ swept over $\{1, 2, 5, 10\}$, and the fixed assay-effect SD τ_α swept over $\{0.5, 1, 2.5, 5\}$. Across both sweeps the cell-mean posterior estimates μ_{ij} shift by at most ≈ 0.03 on the TGI/TCGI scale, the rank ordering High $>$ Medium $>$ Low is preserved within each assay, and the HER2-High threshold-exceedance probability $\Pr(\mu_{1,\text{High}} \geq 0.45 \mid \text{data})$ remains ≥ 0.95 in MiniPDX and ≈ 1.00 in PDX. Full numerical tables and the per-run prior scales are reported in Appendix B.

3.4 Sample Size Re-Estimation for MiniPDX Data

After completing the one & two assays analyses, we conducted a sample size re-estimation for the MiniPDX data

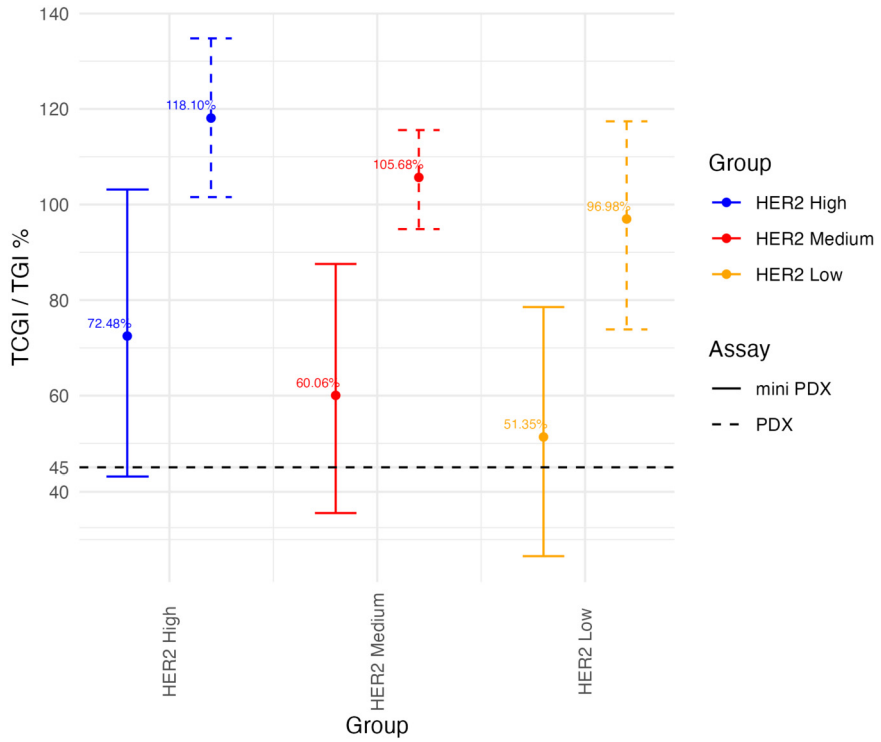


Figure 7: Bayesian credible intervals from the joint MiniPDX/PDX two-assay model for μ_{ij} in each HER2 group (HER2 High/Medium/Low). Each color denotes a HER2 group; solid intervals are for the MiniPDX assay and dashed intervals are for the PDX assay. Dots are posterior means. The dashed horizontal line marks the 0.45 efficacy threshold.

as a demonstration. Based on initial findings, the drug's efficacy is most promising in the HER2 High subgroup for the MiniPDX assay, although this subgroup also shows the largest variability given the unusual value of Tumor 1. The posterior probability $\Pr(\mu_1 \geq 0.45 \mid \text{data}) = 0.853$ at the end of stage 1, and we set a target posterior probability of 0.9 for concluding that the drug is efficacious in this subgroup. Therefore, following Algorithm 1, we conducted adaptive sample size re-estimation and the parameters were established as $f = \text{InvGamma}(2, 5) - 1$, with $\eta = 0.9$, $t_0 = 0.45$, and $n_{sim} = 1000$. We then calculate the values of $\{n(1), \ldots, n(n_{sim})\}$ for the HER2 High group. Figure 8 shows that the simulated extra number n^* of subjects needed is highly skewed, ranging from 1 to 138, with the kernel-density mode at 1.49. Because n^* is a discrete sample size that takes only integer values ≥ 1, the kernel-density mode at 1.49 is a smoothing artifact and is not directly interpretable as a sample-size recommendation; we therefore summarize the simulated distribution by its empirical quantiles. The median of n^* is 3, the 25th and 75th percentiles are 1 and 9, the 90th percentile is 18, and the empirical mass at small integers is $\Pr(n^* = 1) = 0.292$, $\Pr(n^* \leq 2) = 0.419$, and $\Pr(n^* \leq 3) = 0.506$. The distribution is dominated by the smallest discrete values with a long right tail (95th percentile at 25, only 0.4% of simulations exceed 50). The median value $n^* = 3$ is the most defensible single-number recommendation under this design; we also note that the modal integer is $n^* = 1$, while a slightly more conservative integer choice $n^* = 2$ adds a small safety buffer over the mode without committing to the long-tail upper quantiles. In the actual study we prospectively collected 4 additional HER2-high patients (Table 7), which sits between the median and the upper quartile of the simulated distribution; these four observations were collected after the stage-1 analysis indicated that additional data were needed under the adaptive design, and they are not retrospectively chosen for illustration. We acknowledge that summarizing n^* by quantiles, rather than by a kernel-density mode of a discrete distribution, is the standard reporting we would recommend in any prospective application of this procedure.

Based on the 4 additional MiniPDX data points, we reran the one assay hierarchical model using the same hyperparameters and conducted inference, as summarized in Table 8.

From Table 8, we observe that with 4 additional MiniPDX data points, $\Pr(\mu_1 \geq 0.45 \mid \text{data}) = 0.935$, which exceeds the threshold of 0.9. In addition, the posterior standard deviation of μ_1 is reduced to 0.170 from 0.326 previously. These results indicate that adding four additional HER2-high observations sharpened the posterior: the posterior standard deviation decreased ($0.326 \rightarrow 0.170$) and the posterior probability of efficacy crossed the 0.9 threshold,

Table 6. Single-assay Bayesian, two-assay Bayesian, and per-group frequentist estimates of the mean response μ_{ij} in each HER2 group (HER2 High/Medium/Low), for the MiniPDX and PDX assays. The two-assay Bayesian rows use the joint hierarchical model presented above. Sample sizes are $n = 4$ for MiniPDX High, MiniPDX Medium, and PDX High, and $n = 5$ for MiniPDX Low, PDX Medium, and PDX Low. The frequentist columns are identical in the single-assay and two-assay blocks because they depend only on the raw per-group data.

Model	Assay	HER2 Group	Bayes Mean	Bayes 95% CrI half-width	Freq. Mean	Freq. 95% CI half-width
Single-assay Bayes	MiniPDX	High	0.745	0.637	0.968	0.859
Single-assay Bayes	MiniPDX	Medium	0.555	0.454	0.534	0.514
Single-assay Bayes	MiniPDX	Low	0.513	0.372	0.472	0.385
Single-assay Bayes	PDX	High	1.151	0.215	1.195	0.191
Single-assay Bayes	PDX	Medium	1.064	0.124	1.060	0.109
Single-assay Bayes	PDX	Low	0.999	0.288	0.934	0.333
Two-assay Bayes	MiniPDX	High	0.709	0.296	0.968	0.859
Two-assay Bayes	MiniPDX	Medium	0.586	0.254	0.534	0.514
Two-assay Bayes	MiniPDX	Low	0.504	0.250	0.472	0.385
Two-assay Bayes	PDX	High	1.181	0.168	1.195	0.191
Two-assay Bayes	PDX	Medium	1.058	0.102	1.060	0.109
Two-assay Bayes	PDX	Low	0.976	0.215	0.934	0.333

Table 7. Additional TCGI percentages for MiniPDX assays for HER2 High groups. Each row corresponds to a tumor tissue.

Tumor	HER2 Group	MiniPDX TCGI (%)
1	High	83.5
2	High	56.0
3	High	48.5
4	High	43.26

Table 8. Posterior inference for μ_j under different HER2 groups for Stage 1 & 2 MiniPDX data.

HER2 Group	High (μ_1)	Medium (μ_2)	Low (μ_3)	Overall (μ_0)
$\Pr(\mu_j \geq 0.45 \mid \text{data})$	0.935	0.762	0.702	0.773
Posterior Mean of μ_j	0.693	0.563	0.528	0.586
Posterior s.d. of μ_j	0.170	0.205	0.174	0.290

even though the posterior mean shifted slightly downward ($0.745 \rightarrow 0.693$). The sample size re-estimation thus delivered increased certainty about the existing efficacy signal in the HER2 High group, rather than a larger point estimate of efficacy.

4. DISCUSSION

To summarize, this study presents a Bayesian hierarchical modeling framework combined with an adaptive sample size strategy to improve analysis and decision-making in MiniPDX and PDX preclinical trials. By incorporating prior knowledge and accounting for variation across HER2 biomarker subgroups, our approach enables interpretable, flexible, and data-efficient inference even with small sample sizes. Posterior analyses indicated that trastuzumab deruxtecan was most effective in the HER2 High subgroup, with sharpened posterior inference (rather than a larger point estimate) after adaptively adding four new data points. The sample size re-estimation algorithm provided a principled, cost-effective method for determining additional patients needed to achieve a predefined posterior confidence, demonstrating the practical utility of Bayesian adaptive designs in preclinical research.

Integrating PDX and MiniPDX data improved the precision of the MiniPDX estimates. PDX assays in our data showed stronger drug efficacy signals and lower variability, consistent with preservation of the tumor's architecture and microenvironment, and produced higher posterior TGI means and narrower credible intervals. Borrowing information from the PDX assay tightened the MiniPDX posteriors, illustrating the value of joint modeling when one assay carries richer information. The two-assay model improved efficiency in this setting and may help reduce the number of additional animal experiments required to reach a target posterior confidence, though we caution that the magnitude of any such reduction is study-specific.

Despite these strengths, several limitations should be noted. First, initial sample sizes for HER2 subgroups were chosen arbitrarily rather than determined by formal design principles. While the Bayesian adaptive procedure improved resource allocation later, incorporating sample size considerations at the outset could improve trial efficiency. Future work could explore optimal design strategies under prior uncertainty.

Second, while MiniPDX offers faster and more cost-effective evaluation than PDX, it replicates only selected aspects of the tumor microenvironment and lacks full immune interactions, which may limit generalizability to human patients. In addition, technical variation in engraftment efficiency and tumor processing may introduce noise not fully modeled here.

Finally, although we focused on HER2 stratification, more complex biomarker information (e.g., gene expression

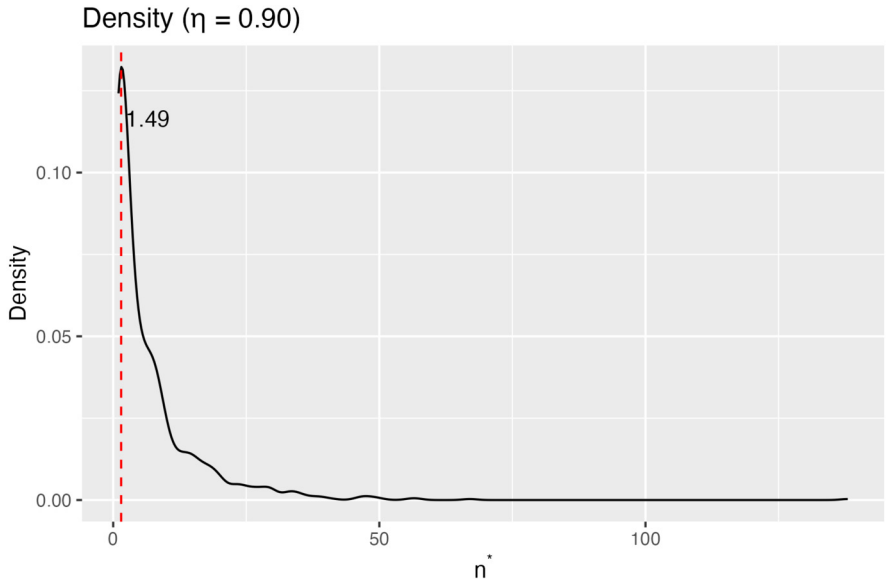


Figure 8: Density for n(sim) using Inverse Gamma prior.

or mutation profiles) could be incorporated to better identify subpopulations with differential responses. This would require more advanced modeling, such as hierarchical latent variable models or Bayesian nonparametric approaches. Furthermore, our two-assay framework can readily extend to accommodate more than two assays, increasing flexibility for future studies.

Overall, this study illustrates how Bayesian hierarchical modeling and adaptive design can be applied to MiniPDX and PDX trials in translational oncology. It offers an interpretable, data-efficient analytic strategy for early-phase preclinical screening and supports the identification of biomarker-defined subgroups with differential treatment response. We do not claim direct clinical translation: confirming external validity would require formal operating-characteristic studies and prospective preclinical-to-clinical bridging, which we leave to future work.

CONFLICT OF INTEREST

Jiaxin Liu is an employee of Shanghai Bayes Data LLC. Diandong Jiang is an employee of Shanghai LIDE Biotech Co., Ltd. ("the Company"), which develops and provides the MiniPDX and PDX technologies and related services that are the subject of the research reported in this manuscript. Xinmeng Zhang is a contractor for Bayesoft, Inc. Hongkui Chen is an employee of the Company. Shizhu Zhao is an employee of the Company. Danyi Wen is the Founder, President, and CEO of the Company and holds equity in the Company. Yuan Ji is a co-founder of Bayesoft, Inc.; serves as an Independent Data Monitoring Committee (IDMC) member for Astellas and Lyell; and is a consultant for Eli Lilly and Astellas.

APPENDIX A. CALCULATION OF τ_{within} AND τ_{between}

To obtain empirical estimates of the between-assay and within-assay variability on the log-variance scale, we proceed as follows. Let z_{ijk} denote observations from assay i and group j. For each combination, we compute the sample variance: $s^2_{ij} = \text{Var}_k(z_{ijk})$ and define the log-variance: $\ell_{ij} = \log s^2_{ij}$.

The grand mean of the log-variances is: $\bar{\ell} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\ell_{ij}$ and the assay-specific mean log-variance is: $\bar{\ell}_i = \frac{1}{J}\sum_{j=1}^{J}\ell_{ij}$.

The between-assay variability is then estimated as:

$$\hat{\tau}_{\text{between}} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(\bar{\ell}_i - \bar{\ell})^2}$$

and the within-assay variability is estimated as:

$$\hat{\tau}_{\text{within}} = \sqrt{\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}(\ell_{ij} - \bar{\ell}_i)^2}$$

We multiply the empirical estimates by a factor of 5 to construct a deliberately wide, weakly informative prior on the log-variance hierarchy: $\tau_{\text{between}} = 5 \times \hat{\tau}_{\text{between}}$ and $\tau_{\text{within}} = 5 \times \hat{\tau}_{\text{within}}$. The plug-in estimates $\hat{\tau}_{\text{within}}$ and $\hat{\tau}_{\text{between}}$ are themselves computed from only $I \times J = 6$ assay-by-group cells, so a multiplier of 1 would yield an over-precise prior that under-states uncertainty in the variance components. The factor of 5 is a conservative default; we report sensitivity to this choice over $\{1, 2, 5, 10\}$ in Appendix B and find that posterior conclusions are stable across the range.

Table A.1. Sensitivity of two-assay posterior inference to the variance multiplier on $(\tau_{within}, \tau_{between})$ for the HER2 High group, computed under the production specification with $\tau_\alpha = 2.5$ fixed. Across multipliers in $\{1, 2, 5, 10\}$ the posterior mean shifts by at most ≈ 0.025 on the TGI/TCGI scale and the threshold-exceedance probability remains ≥ 0.95 in MiniPDX and ≈ 1.00 in PDX.

Assay	Quantity	Mult. 1	Mult. 2	Mult. 5	Mult. 10
MiniPDX	Post. mean (sd) of $\mu_{1,\text{High}}$	0.733 (0.175)	0.718 (0.155)	0.709 (0.151)	0.708 (0.149)
MiniPDX	$\Pr(\mu_{1,\text{High}} \geq 0.45 \mid \text{data})$	0.949	0.960	0.958	0.960
PDX	Post. mean (sd) of $\mu_{2,\text{High}}$	1.179 (0.115)	1.182 (0.090)	1.181 (0.083)	1.182 (0.083)
PDX	$\Pr(\mu_{2,\text{High}} \geq 0.45 \mid \text{data})$	1.000	1.000	1.000	1.000

APPENDIX B. PRIOR-SENSITIVITY ANALYSES

To check that the conclusions in Section 3.3 are not driven by the choice of variance multiplier in Appendix A, we re-fit the two-assay model at multipliers $\{1, 2, 5, 10\}$ applied jointly to $\hat{\tau}_{\text{within}}$ and $\hat{\tau}_{\text{between}}$, holding the data, JAGS model, MCMC controller, and RNG seeds fixed across runs. Empirical estimates were $\hat{\tau}_{\text{within}} = 0.753$ and $\hat{\tau}_{\text{between}} = 0.984$, so the prior scales across the four runs were $(\tau_{\text{within}}, \tau_{\text{between}}) \in \{(0.75, 0.98), (1.51, 1.97), (3.77, 4.92), (7.54, 9.84)\}$. Table A.1 summarizes posterior mean (and standard deviation) of μ_{ij} and the threshold posterior probability $\Pr(\mu_{ij} \geq 0.45 \mid \text{data})$ for the HER2 High row in each assay across the four multipliers. Posterior means shift by at most about 0.025 across the range, the rank ordering High > Medium > Low within each assay is preserved, and the headline conclusion that the HER2-high group exceeds the 0.45 efficacy threshold survives at every multiplier: $\Pr(\mu \geq 0.45 \mid \text{data}) \geq 0.95$ in MiniPDX and ≈ 1.00 in PDX in all four runs. Moving from multiplier 5 to multiplier 10 produces essentially no further change, indicating that the original choice 5 already sits in the near-flat regime of the prior. The fits in this table use the production specification with τ_α fixed at 2.5, matching the headline two-assay results in Section 3.3; at multiplier 5 the MiniPDX HER2-High posterior probability (0.958) coincides with that reported in Table 5.

As a complementary sensitivity check on the fixed assay-effect SD $\tau_\alpha = 2.5$ (Section 3.3), we also re-fit the production model at $\tau_\alpha \in \{0.5, 1, 2.5, 5\}$, holding all other settings fixed. Across this range the cell-mean posterior estimates μ_{ij} shift by at most 0.02 on the TGI/TCGI scale and the threshold posterior probabilities $\Pr(\mu_{ij} \geq 0.45 \mid \text{data})$ stay within ≈ 0.01, so the choice of $\tau_\alpha = 2.5$ does not drive the inference.

APPENDIX C. SUPPLEMENTARY METHODS (WET-LAB PROTOCOLS)

This appendix collects the wet-lab protocols moved out of Section 2 so that the main text focuses on the statistical study design, data structure, and analytic methods. The content here is for reproducibility of the experimental work and is not required to follow the statistical results.

C.0.0.1. Animals 6–8 week-old CB17-SCID mice (Charles River Co., Beijing, China) and nu/nu nude mice (Zhejiang Vital River, Zhejiang, China) were housed at the AAALAC-accredited animal facility at LIDE Biotech (Shanghai, China). CB17-SCID mice were used for PDX model recovery and nu/nu mice were used for drug efficacy tests. All study protocols were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at LIDE Biotech and conducted in accordance with established national and international regulations for laboratory animal welfare.

C.0.0.2. MiniPDX Establishment We developed an in vivo drug sensitivity assay, the MiniPDX, using a modified microencapsulation and hollow fiber culture system (OncoVee MiniPDX®, LIDE Biotech) following the manufacturer's instructions. Tumors with a size of $\geq 500\,\text{mm}^3$ and necrotic areas $< 30\%$ were selected. Tumor tissues were washed with Hank's Balanced Salt Solution (HBSS) in a biosafety cabinet to remove non-tumor and necrotic tissue. Tissues from PDX models were then minced and digested with collagenase at 37 °C for 1–4 hours. After digestion, cells were pelleted by centrifugation at $600\,g$ for 5 minutes, and blood cells and fibroblasts were removed using magnetic beads. Purified tumor cells were washed with HBSS and encapsulated in OncoVee® capsules. Capsules were implanted subcutaneously into 5-week-old nu/nu mice through a small skin incision, with two capsules per mouse.

C.0.0.3. MiniPDX Viability Readout After 10 days of treatment the implanted capsules were removed and tumor cell proliferation was evaluated using the CellTiter Glo Luminescent Cell Viability Assay kit (G7571, Promega, Madison, WI, US) according to the manufacturer's instructions. Luminescence was measured in relative luminance units (RLU) on a SpectraMax M3 spectrophotometer (Molecular Devices, Sunnyvale, CA, US). The RLU values feed directly into the TCGI calculation defined in equation (2.2) of the main text.

Accepted 17 July 2026

REFERENCES

[1] Berry, D. A. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery* **5**(1) 27–36. Publisher: Nature Publishing Group TLDR: The rationale underlying Bayesian clinical trials is explained, the potential of such trials to improve the effectiveness of drug development is discussed, and the potential for smaller more informative trials and for patients to receive better treatment is discussed. https://doi.org/10.1038/nrd1927. Accessed 2025-05-08.

[2] Bertotti, A., Migliardi, G., Galimi, F., Sassi, F., Torti, D., Isella, C., Corà, D., Di Nicolantonio, F., Buscarino, M., Petti, C. et al. (2011). A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discovery* **1**(6) 508–523.

[3] Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y. et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine* **21**(11) 1318–1325.

[4] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**(3) 515–534. https://doi.org/10.1214/06-BA117A. MR2221284

[5] Giovagnoli, A. (2021). The Bayesian Design of Adaptive Clinical Trials. *International Journal of Environmental Research and Public Health* **18**(2) 530. TLDR: A brief overview of the recent literature on adaptive design of clinical trials from a Bayesian perspective for statistically not so sophisticated readers, pointing at further interesting reading material. https://doi.org/10.3390/ijerph18020530. Accessed 2025-05-19.

[6] Golchi, S., Willard, J. J., Pullenayegum, E., Bassani, D. G., Pell, L. G., Thorlund, K. and Roth, D. E. (2022). A Bayesian adaptive design for clinical trials of rare efficacy outcomes with multiple definitions. *Clinical Trials (London, England)* **19**(6) 613–622. TLDR: A Bayesian adaptive design is proposed that enables monitoring and utilizing multiple composite outcomes based on rare events to optimize the trial design operating characteristics and maintain a realistic probability of stopping in trials with low event rates. https://doi.org/10.1177/17407745221118366. https://doi.org/10.1002/cjs.11699. MR4429845

[7] Hidalgo, M., Amant, F., Biankin, A. V., Budinská, E., Byrne, A. T., Caldas, C., Clarke, R. B., de Jong, S., Jonkers, J., Mælandsmo, G. M. et al. (2014). Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discovery* **4**(9) 998–1013.

[8] Kelter, R. (2023). Reducing the false discovery rate of preclinical animal research with Bayesian statistical decision criteria. *Statistical Methods in Medical Research* **32**(10) 1880–1901. Publisher: SAGE Publications Ltd STM TLDR: Simulation shows that a shift towards statistical approaches which explicitly incorporate the minimum clinically important difference reduces the false discovery rate of frequentist approaches and that Bayesian statistical decision criteria can improve the reliability of preclinical animal research by reducing the number of false-positive findings. https://doi.org/10.1177/09622802231184636. Accessed 2025-05-19. https://doi.org/10.1177/09622802231184636. MR4651763

[9] Neuenschwander, B., Branson, M. and Gsponer, T. (2008). Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine* **27**(13) 2420–2439. TLDR: The Bayesian approach to finding the maximum-tolerated dose in phase I cancer trials is discussed and a comparison with the continual reassessment method (CRM) is performed with data from an actual trial and a simulation study. https://doi.org/10.1002/sim.3230. https://doi.org/10.1002/sim.3230. MR2432497

[10] Pallmann, P., Bedding, A. W., Choodari-Oskooei, B., Dimairo, M., Flight, L., Hampson, L. V., Holmes, J., Mander, A. P., Odondi, L., Sydes, M. R., Villar, S. S., Wason, J. M. S., Weir, C. J., Wheeler, G. M., Yap, C. and Jaki, T. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine* **16**(1) 29. TLDR: This tutorial paper provides guidance on key aspects of adaptive designs that are relevant to clinical triallists, and emphasises the general principles of transparency and reproducibility and suggest how best to put them into practice. https://doi.org/10.1186/s12916-018-1017-7. Accessed 2025-05-19. https://doi.org/10.1002/sim.2252. MR2225183

[11] Research, C. f. D. E. a. (2020). *Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry*. Publisher: FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry Accessed 2025-05-19.

[12] Rokita, J. L., Rathi, K. S., Cardenas, M. F., Upton, K. A., Jayaseelan, J., Cross, K. L., Pfeil, J., Egolf, L. E., Way, G. P., Farrel, A., Kendsersky, N. M., Patel, K., Gaonkar, K. S., Modi, A., Berko, E. R., Lopez, G., Vaksman, Z., Mayoh, C., Nance, J., McCoy, K., Haber, M., Evans, K., McCalmont, H., Bendak, K., Böhm, J. W., Marshall, G. M., Tyrrell, V., Kalletla, K., Braun, F. K., Qi, L., Du, Y., Zhang, H., Lindsay, H. B., Zhao, S., Shu, J., Baxter, P., Morton, C., Kurmashev, D., Zheng, S., Chen, Y., Bowen, J., Bryan, A. C., Leraas, K. M., Coppens, S. E., Doddapaneni, H., Momin, Z., Zhang, W., Sacks, G. I., Hart, L. S., Krytska, K., Mosse, Y. P., Gatto, G. J., Sanchez, Y., Greene, C. S., Diskin, S. J., Vaske, O. M., Haussler, D., Gastier-Foster, J. M., Kolb, E. A., Gorlick, R., Li, X. -N., Reynolds, C. P., Kurmasheva, R. T., Houghton, P. J., Smith, M. A., Lock, R. B., Raman, P., Wheeler, D. A. and Maris, J. M. (2019). Genomic Profiling of Childhood Tumor Patient-Derived Xenograft Models to Enable Rational Clinical Trial Design. *Cell Reports* **29**(6) 1675–16899. Publisher: Elsevier TLDR: Genomically characterize 261 PDX models from 37 unique pediatric cancers; demonstrate faithful recapitulation of histologies and subtypes; refine the understanding of relapsed disease; and use expression signatures to classify tumors for TP53 and NF1 pathway inactivation. https://doi.org/10.1016/j.celrep.2019.09.071. Accessed 2025-05-19.

[13] Townsend, E. C., Murakami, M. A., Christodoulou, A., Karakunnel, J., Le, T., Liu, P., Choe, S., Chong, Y., Vivanco, I., Damoiseaux, R. et al. (2016). The public repository of xenografts enables discovery and randomized phase II-like trials in mice. *Cancer Cell* **29**(4) 574–586.

[14] Walley, R., Sherington, J., Rastrick, J., Detrait, E., Hanon, E. and Watt, G. (2016). Using Bayesian analysis in repeated preclinical in vivo studies for a more effective use of animals. *Pharmaceutical Statistics* **15**(3) 277–285. TLDR: It is concluded that using Bayesian methods in stable repeated in vivo studies can result in a more effective use of animals, either by reducing the total number of animals used or by increasing the precision of key treatment differences. https://doi.org/10.1002/pst.1748.

[15] Welch, C., Forster, M., Ronaldson, S., Keding, A., Corbacho-Martín, B. and Tharmanathan, P. (2024). The performance of a Bayesian value-based sequential clinical trial design in the presence of an equivocal cost-effectiveness signal: evidence from the HERO trial. *BMC Medical Research Methodology* **24**(1) 155. TLDR: Evidence from the two retrospective applications of this Bayesian value-adaptive design suggests that, when the cost-effectiveness signal in a clinical trial is unambiguous, the Bayesian value-adaptive design can stop the trial before it reaches its maximum sample size, potentially saving research costs when compared with the alternative fixed sample size design. https://doi.org/10.1186/s12874-024-02248-9. Accessed 2025-05-19.

[16] Zhang, B., Magiera, L., Candido, J., Muraeva, O., Coates Ulrichsen, J., Eyles, J., Galvani, E. and Karp, N. A. (2024). Bayesian modeling for analyzing heterogeneous response in preclinical mouse tumor models. *Science Trans-*

lational Medicine **16**(771) 9004. TLDR: A statistical method called INSPECT (IN vivo reSPonsE Classification of Tumors) for analyzing heterogeneous responses through Bayesian modeling is developed and shown that INSPECT methodology is more accurate and sensitive than existing methods with respect to balancing false-negative and false-positive rates. https://doi.org/10.1126/scitranslmed.adi9004. Accessed 2025-05-09.

[17] Zhang, F., Wang, W., Long, Y., Liu, H., Cheng, J., Guo, L., Li, R., Meng, C., Yu, S., Zhao, Q., Lu, S., Wang, L., Wang, H. and Wen, D. (2018). Characterization of drug responses of mini patient-derived xenografts in mice for predicting cancer patient clinical therapeutic response. *Cancer Communications (London)* **38**(1) 60. https://doi.org/10.1186/s40880-018-0329-5.

Jiaxin Liu. Shanghai Bayes Data LLC., Shanghai, China.

Diandong Jiang. Shanghai LIDE Biotech Co. Ltd., Shanghai, China.

Xinmeng Zhang. Bayesoft Inc., Chicago, USA.

Hongkui Chen. Shanghai LIDE Biotech Co. Ltd., Shanghai, China.

Shizhu Zhao. Shanghai LIDE Biotech Co. Ltd., Shanghai, China.

Danyi Wen. Shanghai LIDE Biotech Co. Ltd., Shanghai, China.

Yuan Ji. Department of Public Health Sciences, The University of Chicago, Chicago, IL 60637, USA.
E-mail address: koaeraser@gmail.com