Incorporation of healthy volunteers data on receptor occupancy into a phase II proofofconcept trial using a Bayesian dynamic borrowing design

Incorporation of phase I healthy volunteers data on receptor occupancy into a phase II proof-of-concept trial using a Bayesian dynamic borrowing design

Fulvio Di Stefano[1], Christelle Rodrigues[2], Stephanie Galtier[3], Sandrine Guilleminot[4], Veronique Robert[3], Mauro Gasparini[1], and Gaelle Saint-Hilary[1,3]

[1]Dipartimento di Scienze Matematiche (DISMA) "Giuseppe Luigi Lagrange", Politecnico di Torino, Torino, Italy
[2]Department of Quantitative Pharmacology, Institut de Recherches Internationales Servier, Suresnes, France
[3]Department of Clinical Statistics, Institut de Recherches Internationales Servier, Suresnes, France
[4]Department of Preclinical Statistics and Biomarkers, Institut de Recherches Internationales Servier, Suresnes, France

Abstract

Receptor occupancy in targeted tissues measures the proportion of receptors occupied by a drug at equilibrium and is sometimes used as a surrogate of drug efficacy to inform dose selection in clinical trials. We propose to incorporate data on receptor occupancy from a phase I study in healthy volunteers into a phase II proof-of-concept study in patients, with the objective of using all the available evidence to make informed decisions. A minimal physiologically based pharmacokinetic (mPBPK) modeling is used to model receptor occupancy in healthy volunteers and to predict it in the patients of a phase II proof-of-concept (PoC) study, taking into account the variability of the population parameters and the specific differences arising from the pathological condition compared to healthy volunteers. Then, given an estimated relationship between receptor occupancy and the clinical endpoint, an informative prior distribution is derived for the clinical endpoint in both the treatment and control arms of the phase II study. These distributions are incoporated into a Bayesian dynamic borrowing design to supplement concurrent phase II trial data. A simulation study in immuno-inflammation demonstrates that the proposed design increases the power of the study while maintaining a type I error at acceptable levels for realistic values of the clinical endpoint.

Keywords: Bayesian, borrowing, historical data, receptor occupancy, early phase, PBPK, phase II.

Abbreviations: PBPK, physiology based pharmacokinetic; mPBPK, minimal physiology based pharmacokinetic; RO, receptor occupancy; PoC, proof-of-concept; BDB, Bayesian dynamic borrowing; CrI, credibility interval.

1 Introduction

Extrapolating relevant information from prior studies has become a prominent topic in pharmaceutical research. The changing landscape of the clinical trials sometimes results in smaller sample sizes available at the testing phases [Bradley, 2012]. In certain instances, such as with targeted therapies and personalized approaches, conducting large studies may not be practical or feasible. Smaller sample sizes inevitably lead to reduced precision. Fortunately, there are innovative solutions that address this challenge, including borrowing from historical data.

This paper proposes an approach which incorporates data extrapolated from a minimal physiology based pharmacokinetic (mPBPK) model in healthy volunteers into a clinical trial design in patients. This approach is part of the broader framework of Model-based Drug Development, on which a rich literature is available in order to optimize trial designs and to make informed decisions [Lalonde et al., 2007].

Extrapolating healthy volunteer data to patients is a well-known topic in pharmacometrics' literature [Radanovic et al., 2022]. Previous work in this field has focused on extrapolation from adult to pediatric data ([Chan et al., 2021, Khosravan et al., 2021]). Other researchers have examined bridging models to identify differences between healthy volunteers and patients ([Jonsson et al., 2005, Willmann et al., 2021]). In the statistical literature, several publications have focused on extrapolation from pharmacokinetic models to clinical data [La Gamba et al., 2019] and increasing attention has been given to the incorporation of historical data [Natanegara et al., 2013, Ghadessi et al., 2020, Burger et al., 2021]. The main advantage of these methods is the ability to estimate the treatment effect with precision and increase of study power while limiting, or even reducing [Viele et al., 2014], type I error in cases of consistency between historical and concurrent data. However, type I error may be inflated in cases of prior data conflict. Bayesian dynamic borrowing (BDB) methods have been developed to address this issue [Neuenschwander et al., 2010, Schmidli et al., 2014] and have the advantage of down-weighting historical data in the analysis when the study data are not sufficiently similar [Lim et al., 2018, Viele et al., 2014, van Rosmalen et al., 2017, Smith et al., 2020]. These methods have also been studied in the context of platform trials [Meyer et al., 2022].

In the following, a motivating case study and the main steps of the proposed approach are presented in Section 2. Section 3 contains the methodology used, consisting of the mPBPK model, the extrapolation procedure to obtain informative priors and the BDB design. The proposed approach is applied to a case study in immuno-inflammation in Section 4, where comprehensive operating characteristics for different designs are presented in order to select the most appropriate one. Also, a fictive analysis is conducted as if study data were collected to anticipate the real final analysis. A discussion concludes in Section 5.

2 Motivating case study in immuno-inflammation

This work was motivated by a real case study in immuno-inflammation, where the use of receptor occupancy (RO) historical data from a phase I study in healthy volunteers was considered to inform the design of a phase II proof-of-concept (PoC) study in patients. The justification stems from the ethical and feasibility limitations imposed by the rarity of the disease, which restrict the sample size available for the phase II study. RO data in healthy volunteers were used, in combination with pharmacokinetic data, to construct an mPBPK model of the drug at the end of a phase I study. RO is defined as the proportion of receptors occupied by a particular drug over the total number of receptors. This measure helps to understand or confirm the mechanism of action of certain drugs [Christopoulos and El-Fakahany, 1999] and allows for quantification and characterization of the drug's binding profile to the target [Liang et al., 2016]. RO can be used as a pharmacodynamic biomarker for characterizing the exposure-response relationship of the drug in combination with a pharmacokinetic profile. It can be evaluated in phase I studies to predict the activity of the treatment and support dose selection in subsequent studies. For the presented application, the model in healthy volunteers is extended to patients accounting for two aspects: the variability of the population parameters and the peculiar differences due to the pathological condition.

Simultaneously, we consider a phase II PoC study, designed as randomised, double-blinded, two-arm efficacy and safety trial using a multiple dosing strategy over a 13-week period. The study will enroll 45 patients, with 30 patients in the treatment arm and 15 in the control arm. The clinical endpoint is continuous, with negative values indicating greater efficacy, as they represent an improvement in the pathological condition compared to baseline. A reduction of 3 points in the clinical endpoint is considered clinically relevant. It is assumed that the clinical endpoint follows a normal distribution with a standard deviation of 6, as reported in the literature. A linear relationship between RO and the clinical endpoint at the patient level has been estimated from a previous (internal and unpublished) study.

The study is designed using Bayesian Dynamic Borrowing (BDB) methods, as presented in the next Section, to incorporate prior evidence from phase I healthy volunteer's data on RO. The aim is to increase the power of the study while maintaining its sample size practical, and to make informed decisions at the end of the PoC study using all the available evidence from the current and past studies.

Details on the specific disease of the case study cannot be disclosed due to confidentiality reasons; however, the scientific investigation and simulations carried out are presented in a de-identified manner to provide useful insights for future applications. Although it was ultimately decided not to pursue this approach, the research conducted here provides a general framework that can be applied to any disease area where RO is expected to be related to drug efficacy, such as (but not limited to) monoclonal therapeutic antibodies in oncology [Junker et al., 2021, Jones et al., 2022], or in some neurological disorders [Srinivas et al., 2018]. In other applications, these methods would also permit to replace RO by a biomarker that is pathophysiologically close to the desired clinical endpoint, potentially resulting in a better correlation and/or lower variability, depending on the previous data or literature available in the context of the project.

3 Methods

The following section introduces the methodologies employed to model RO in healthy volunteers and patients, to incorporate historical data within a BDB design and to extrapolate from phase I to phase II data.

3.1 mPBPK model

Physiology based pharmacokinetic (PBPK) models are mathematical models that predict a drug's exposure and response for different dosage regimens in a target population. Compared to classical pharmacokinetic models, they are parameterised using known anatomical, physiological, physical and chemical mechanisms. They usually consist of a larger number of compartments with several differential equations describing the drug's behaviour in the various compartments. mPBPK models have been developed to overcome the implementation difficulties of such models by lumping tissues that exhibits similar kinetics into two categories according to their endothelial structure: tight and leaky compartments. A more detailed discussion on PBPK and mPBPK models is presented in [Jones and Rowland-Yeo, 2013] and [Cao and Jusko, 2014], respectively.

The following differential equation describes the behavior of the RO in the leaky compartment, which is responsible for producing the secretions of interest. The full mPBPK model used in this case study is detailed in the Supplementary Material and is represented generically as a function $f(\cdot)$.

$$\frac{\partial C_{leaky}}{\partial t} = f\left(C_{leaky}, C_P, V_{max}, K_m, V_{leaky}\right)$$
$$RO = \frac{C_{leaky}}{C_{leaky} + K_m}.$$

Table 1 shows the five parameters used in the model.

The distinctive characteristic of the illness requires that the maximum binding capacity in the binding site V_{max} in patients has to be multiplied by an individual factor λ, assumed to follow a log-normal

Parameter	**Description**
C_{leaky}	Concentration of the drug in the leaky compartment
C_P	Concentration of the drug in the plasma
V_{max}	Maximum binding capacity in the binding site
K_m	Concentration of the free (not bound) drug
V_{leaky}	Volume of the distribution of the drug in the leaky compartment

Table 1: Main parameters of the mPBPK model. The V_{max} and K_m are individual parameters estimated from previous studies in healthy volunteers. The V_{leaky} is a fixed volume for all individuals. More details on the model and the parameters can be found in Supplementary Material.

distribution estimated from external studies. This is assumed to be the only difference between the mPBPK model in healthy volunteers and patients.

3.2 Bayesian dynamic borrowing design

The upcoming phase II PoC study is planned as a BDB design, with the objective of incorporating historical data to obtain more evidence for decision making. Let θ_T and θ_C denote the true mean of the clinical endpoint in the treatment and control arms of the new study, respectively. Let π_T and π_C be the informative continuous prior distributions for θ_T and θ_C, respectively, summarizing the information on the clinical endpoint coming from the phase I study (see Section 3.3). To address potential prior-data conflict, these distributions are robustified using a mixture prior approach that combines them with two vague distributions π_T^V and π_C^V with the same mean but much larger variance. The robust distributions are derived as:

$$\pi_T^R = w\pi_T + (1-w)\pi_T^V$$
$$\pi_C^R = w\pi_C + (1-w)\pi_C^V$$

The distributions π_T^R and π_C^R will be used as prior distributions for θ_T and θ_C in the phase II PoC study.

The prior weight w represents the degree of belief, prior to observing the phase II data, that the information extrapolated from phase I is relevant for the phase II study. For simplicity, the same prior weight is used for both the treatment and control arms, assuming that the extrapolated data is equally relevant for both arms. However, if scientifically justified, different prior weights could be used for each arm. The choice of the prior weight(s) requires careful consideration: it is typically a compromise between *a priori* expectations about the relevance of extrapolated data and frequentist operating characteristics. A comprehensive presentation of the operating characteristics of the implemented methodology is provided in the following case study to aid in the selection of design parameters, including the prior weight on the extrapolated data. Furthermore, a tipping point sensitivity analysis [Best et al., 2021] is provided to demonstrate the process of evaluating the robustness of the results following the data collection phase.

3.3 Extrapolation

The extrapolation from phase I to phase II data is conducted in two steps: using data from healthy volunteers to predict RO in patients and using the predicted RO in patients to estimate the clinical endpoint. The methodology for both arms is analogous, so only the methodology for the treatment arm T is presented for simplicity.

Let N_T be the number of patients in the treatment arm and γ_i the logit of the true RO value for patient $i = 1, ..., N_T$: $\gamma_i = \text{logit}(\text{RO}_i) = \log(\text{RO}_i/(1-\text{RO}_i))$. Let denote by θ_i the true treatment effect on the clinical endpoint for the patient $i = 1, ..., N_T$; the true mean of the clinical endpoint in the treatment arm is then $\theta_T = \frac{1}{N_T}\sum_{i=1}^{N_T}\theta_i$. The relationship between the logit of the RO and the clinical endpoint of each patient is assumed to be previously estimated based on the literature or past studies. Although it should be true to a certain extent if the dose after phase I is intended to be chosen based on RO results, the estimation of this relationship may be lacking in practice and be replaced by clinical assumptions. Following [Saint-Hilary et al., 2018a], a regression model is used, where $\theta_i \sim N(a_i + b_i\gamma_i, \tau_i^2)$ for $i = 1, \ldots, N_T$, with $a_i \sim N(\mu_a, \sigma_a^2)$, $b_i \sim N(\mu_b, \sigma_b^2)$ and $\tau_i \sim HN(\frac{1}{\mu_\tau})$.

The prior distribution π_T for θ_T is approximated according to the following steps:

- Simulate from the mPBPK model K values of $\hat{\gamma}_{i,k}$ for $k = 1, ..., K$ and each patient $i = 1, ..., N_T$;
- Sample $a_{i,k}$, $b_{i,k}$ and $\tau_{i,k}$ values for each patient $i = 1, ..., N_T$ from the estimated distributions of a_i, b_i and τ_i;
- Sample $\hat{\theta}_{i,k}$ for $k = 1, ..., K$ and each patient $i = 1, ..., N_T$ from the normal distribution $N(a_{i,k} + b_{i,k}\hat{\gamma}_{i,k},\ \tau_{i,k}^2)$;
- Pick one value of $\hat{\theta}_{i,k}$ for each patient $i = 1, ..., N$, denoted $\hat{\theta}_{i,k^*}$, which represents the patient's outcome in a fictive trial k^*;
- Calculate the mean effect on the clinical endpoint over all patients in the trial $\hat{\theta}_{T,k^*} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{i,k^*}$;

- Repeat the two previous steps a large number of times to approximate the distributions of the mean clinical endpoint in the trial.

This procedure provides estimates of the mean effect on the clinical endpoint for the treatment arm, $\hat{\theta}_{T,k^*}$, which are used to approximate the prior distribution π_T. The same steps are used to obtain an approximation of the prior distribution π_C in the control arm.

4 Case study in immuno-inflammation

The methods proposed in this paper are now applied to a case study. The analyses of the mPBPK model have been performed using Simulx Version 2020R1 and the BDB design has been implemented in R Version 4.0.5 and package RBesT Version 1.6.1. As stated above, for confidentiality reasons, the data used in this section are simulated.

A dual criterion for success is considered in the study: a statistically significant reduction of the clinical endpoint in the treatment group compared to control at the one-sided 10% significance level (translated in terms of posterior probability as $P[(\theta_T - \theta_C) < 0] > 0.9$); and a fairly good assurance that the reduction of the clinical endpoint in the treatment group is 3 points or greater than control (expressed in terms of posterior probability as $P[(\theta_T - \theta_C) < -3] > 0.5$). These criteria were chosen to strike a balance between feasibility and risk, with the second criterion selected to limit the risk of making incorrect decisions [Neuenschwander et al., 2011]. As described in [Chuang-Stein and Kirby, 2017], the rationale for a Go decision under this approach is to have at least 50% confidence that the true treatment effect is greater than a targeted clinical value, while at the same time having sufficient precision to be confident that the true treatment effect is greater than 0. In a frequentist framework, equivalent to a Bayesian design assuming an implicit improper conjugate Normal prior with precision equal to zero, the use of these two criteria leads to an overall type I error of 5.2% and a power of 50% in case the true difference between treatment and control is only 3 points. This implies that the study, without borrowing, is designed to have a false positive rate of approximately 5% when the true difference between treatment and control is zero. Additionally, the study has a power of less than 14% for true differences between treatment and control of less than 1 point, while it has approximately 70% power when the true difference between treatment and control is 4 points.

Weight (w)	**Treatment Mean**	**Treatment Standard Deviation**	**Treatment ESS (ELIR)**	**Control Mean**	**Control Standard Deviation**	**Control ESS (ELIR)**	**Total ESS (ELIR)**
1	-3.786	1.148	27	-0.018	1.595	14	41
0.8	-3.786	2.873	18	-0.018	3.039	9	27
0.65	-3.786	3.668	13	-0.018	3.775	7	20
0.5	-3.786	4.320	9	-0.018	4.390	5	14
0	-3.786	6	1	-0.018	6	1	2

Table 2: Characteristics of different priors for different weights (w) on the informative component. A weight of 1 indicates the complete borrowing of information without robustification, while a weight of 0 indicates the use of the vague priors only. ESS is Effective Sample Size, using the ELIR method.

4.1 mPKPB model and BDB design

The distribution of the factor λ, which multiplies the maximum binding capacity in the binding site V_{max} in patients compared to healthy volunteers, is estimated based on external studies as $\log(\lambda) \sim N(1.297, 0.604^2)$. This distribution has a median of approximately 3.659, which means that V_{max} is generally much larger in patients with respect to healthy volunteers. Following the steps described in Section 3.3, $K = 1000$ clinical trials were simulated to obtain the prior distribution for each arm. This number was chosen because a further increase did not improve the overall precision of the results. More simulations may be needed in another context and the number of required iterations should be assessed in a case-by-case basis.

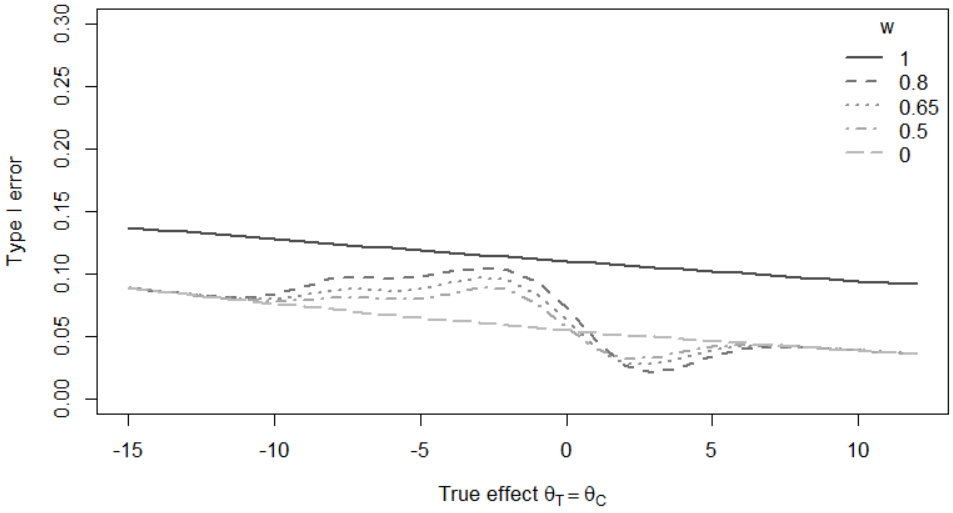

Figure 1: Type I error plotted against the effect in each arm, assumed equal, with 30 patients in the treatment arm, 15 patients in the control arm and different prior weights (w) on the informative component.

The replicates of ROs at the end of week 13 were simulated via Simulix. The histograms of the logits of ROs, $\hat{\gamma}_{i,k}$ for $i = 1, ..., N_T$ (or N_C for the control) and $k = 1, ..., K$ are presented in Supplementary Material. The relationship between the RO and the clinical endpoint for each patient was established in a previous (internal and unpublished) study, with $\theta_i \sim N(a_i + b_i\gamma, \tau_i^2)$, $a_i \sim N(-2.893, 0.029^2)$, $b_i \sim N(-0.181, 0.003^2)$ and $\tau_i \sim HN(\frac{1}{5.040})$ for $i = 1, ..., N_T$ (or N_C for the control).

The two informative prior distributions obtained are:

$$\pi_T \sim N(-3.786, 1.148^2)$$
$$\pi_C \sim N(-0.018, 1.595^2)$$

Vague priors π_T^V and π_C^V are used for the robustification. They have the same mean of the informative priors and a standard deviation of 6, the value of the sampling standard deviation of the clinical endpoint retrieved from external studies. Therefore, the vague prior components are equivalent to "one patient worth" of information in each arm. Table 2 contains the main characteristics of different mixtures of the treatment and control priors for several values of the prior weight on the informative component w: mean, standard deviation and effective sample size (ESS) according the ELIR method [Neuenschwander et al., 2020]. The informative priors have a total ESS = 41 patients: the treatment prior has an ESS of 27 patients and the control prior has an ESS of 14 patients, both slightly lower than the number of patients enrolled in their arm. The ESS decreases as the prior weight on the vague component increases, until reaching an ESS of 2 patients when the phase I information is entirely discarded ($w = 1$), as expected by construction.

4.2 Operating Characteristics

Operating characteristics are provided to evaluate the performance of the BDB design and help determine the appropriate weight on the informative component (w). Figure 1, Figure 2 and Table 3 show the type I error and power for different values of w, considering the setting of the phase II PoC study. Plausible ranges for the treatment and control effects were defined using the 0.1% percentile of π_T as the lower bound and the 99.9% percentile of π_C as the upper bound. These two percentiles define a range of plausible values from -7.3 to 4.9 accounting for prior knowledge.

Figure 1 displays the type I error for varying assumptions of the true effect on the clinical endpoint, assuming it is equal for the treatment and the control arms. The type I error rates remain below 10% when the weight on the informative component is zero ($w = 0$), but increase as the true effect decreases:

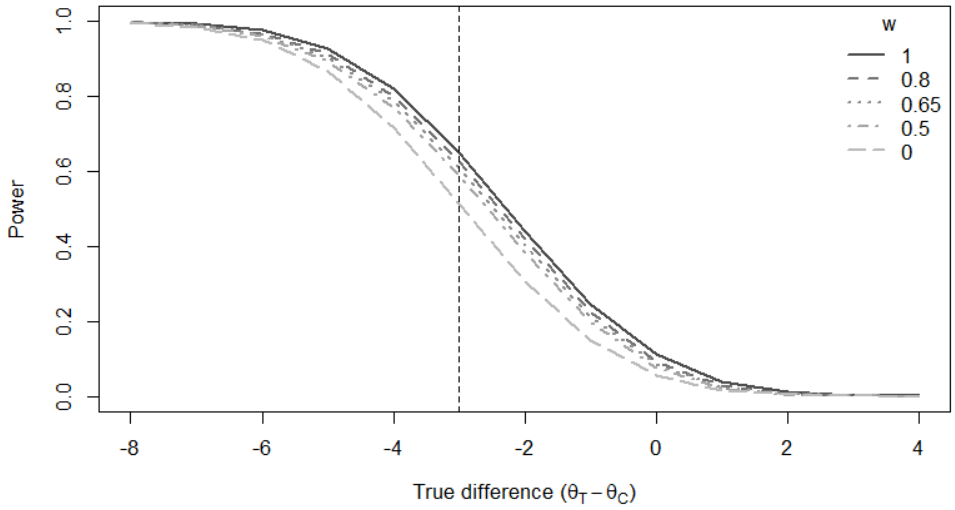

Figure 2: Power in the BDB design plotted against the difference between effects in each arms, while considering a control effect $\theta_C = -1$, with 30 patients in the treatment arm, 15 patients in the control arm and different prior weights (w) on the informative component. The lower the difference between treatment and control, the better the effect of the treatment. The vertical dotted line represents the minimum clinically relevant difference of -3.

Design	**Type I error when** $\theta_T = \theta_C = -1$	**Maximum type I error over the plausible range** (value at which occurs)	**Range of values where type I error is greater than 10%** (probability under informative priors)	**Power when** $\theta_T = -4$ $\theta_C = -1$
BDB with $w = 1$	11.2%	12.3% (-7.3)	[-7.3, 4.9] (99.8%)	64.9%
BDB with $w = 0.8$	9.3%	10.5% (-2.8)	[-4.4,-1.6] (10.7%)	62.8%
BDB with $w = 0.65$	8.4%	9.7% (-2.6)	-	61.0%
BDB with $w = 0.5$	7.6%	8.9% (-2.6)	-	59.1%
BDB with $w = 0$	5.7%	7% (-7.3)	-	51.2%
Without borrowing	5.2%	5.2% (all)	-	50%

Table 3: Summary of the operating characteristics for BDB designs with 30 patients in the treatment arm, 15 patients in the control arm, different prior weights (w) on the informative component and design without borrowing. The plausible range for the treatment and control effects is defined as [-7.3,4.9].

this is because the design is unbalanced. Indeed, the vague priors used for both arms contain the same amount of information (one subject's observation). The prior has therefore less impact on the treatment arm than on the control arm, because there are more patients in the treatment arm to counterbalance it. As a result, when the true effect is lower than the effect extrapolated from RO, the posterior estimate for the control arm decreases to a lesser extent than that of the treatment arm, leading to an increased posterior estimate of the difference between treatments and an increased type I error. On the other hand, when the previous information is fully borrowed ($w = 1$), there is a substantial increase in the type I error: it remains above 10% for all values and increases as the true effect decreases, again due to the unbalanced nature of the design. It should be noted that this situation, where $w = 1$ and the ESS of the borrowed data is almost equivalent to the sample size of the phase II study, is likely to be excessive and is presented for illustrative purposes only. Notwithstanding, for intermediate values of w the type I error increases are smaller and less likely. The type I error exceeds 10% across the plausible range only in the case of $w = 0.8$ when the true effect is between $[-4.4, -1.6]$, peaking at 10.5% when the true effect is -2.8. The probability that both the treatment and the control effects are in the range $[-4.4, -1.6]$ is only 10.7% under π_T and π_C. For true values lower than -4.4, the type I error remains between 10% and the type I error for the BDB design where the informative prior is entirely discarded ($w = 0$); while for true values higher than -1.6 the type I error decreases quickly, lying below the type I error with $w = 0$ for almost all positive values on the true effect.

The power is shown in Figure 2 for true values of the difference between the treatment and the control group while considering a control effect of -1. The vertical dashed line represents the second component of the dual criterion for success: a mean reduction in clinical endpoint in the treatment group of 3 points or greater than the control group. In this case, the design without borrowing has a power of 50%, by construction. On the other hand, the BDB designs have increased power with increasing weights: 59.1% ($w = 0.5$), 61% ($w = 0.65$), 62.8% ($w = 0.8$) and 64.9% ($w = 1$). It is worth noting that the BDB design with $w = 0$ already has a greater power (51.2%) compared to the design without borrowing because of the incorporation of information equivalent to one observation in each arm through the vague priors (see Section 3.1).

The operating characteristics of the design without borrowing and BDB designs are summarized in Table 3: type I error; maximum type I error; range of values where the type I error is greater than 10%; power. These results illustrate that the incorporation of extrapolated data from RO and the use of a BDB design permit to increase the power of the study while limiting the type I error inflation in almost all cases. The use of data extrapolated from RO in the phase I trial is considered equivalent to add 14 to 41 patients to the study, according to ELIR method, which is particularly useful in early phase PoC studies where the sample size is usually limited.

It can be observed from the figure in Supplementary Material that γ in the treatment arm exhibits two modes. This phenomenon occurs because some patients may experience a faster or slower decay in the RO depending on their individual characteristics. An additional analysis is performed to address this behaviour in the Supplementary Material, using an extension of the mixture prior. This extension includes an additional informative component to better capture such shape. The resulting model and operating characteristics are similar to those presented here.

Additional operating characteristics are also presented in the Supplementary Material. Heatmaps show that both the pointwise type I error and the maximum type I error tend to increase with increasing weight and decreasing sample size. However, the power tends to increase with higher weights and lower sample size. Separate heatmaps were produced using the first success criterion only (statistical significance) and the second success criterion only (clinical relevance). The results indicate that the second success criterion, which requires a clinically relevant difference between the treatment and control arms ($\mathrm{P}[(\theta_T - \theta_C) < -3] > 0.5$), mostly drives the overall operating characteristics of the designs.

Although the appropriate weight for the informative component of the prior should be discussed on a case-by-case basis depending on the project's stakes, let us assume for the time being that a weight $w = 0.8$ is considered appropriate. According to Table 3, this weight results in a type I error of around 10% and a power of 62.8%. With this weight, the long-run (i.e. repeated sampling) operating characteristics of the BDB design are presented, accounting for potential drift between the true treatment/control effects and the prior means. The results show that the average posterior weight on the prior distribution increases when there is greater similarity between the prior and the true effects, leading to a narrower posterior credibility interval (CrI) compared to a design without borrowing in cases of prior data consistency. However, as drift increases, less posterior weight is placed on the evidence informed by RO data, leading to decreased precision gains. Additional results, such as bias and illustrative results under selected data scenarios, are presented to provide a comprehensive description of the BDB designs.

Evidence Source	Treatment difference [80% CrI]	Treatment effect [80% CrI]	Control effect [80% CrI]
Phase I + phase II	-3.5 [-5.7,-1.3]	-3.9 [-5.1,-2.6]	-0.4 [-2.2,1.4]
Phase II only (without borrowing)	-3 [-6.6,0.6]	-4 [-6.1,-1.9]	-1 [-4,2]

Table 4: Summary of the primary analysis on the treatment difference, treatment and control response, considering fictive realistic data. The bottom row displays the fictive observed data from a design without borrowing, outlining failure to meet success criteria. The top row displays the results obtained by combining fictive observed data and informative components using a BDB design with weight $w = 0.8$, outlining the meeting of the success criteria.

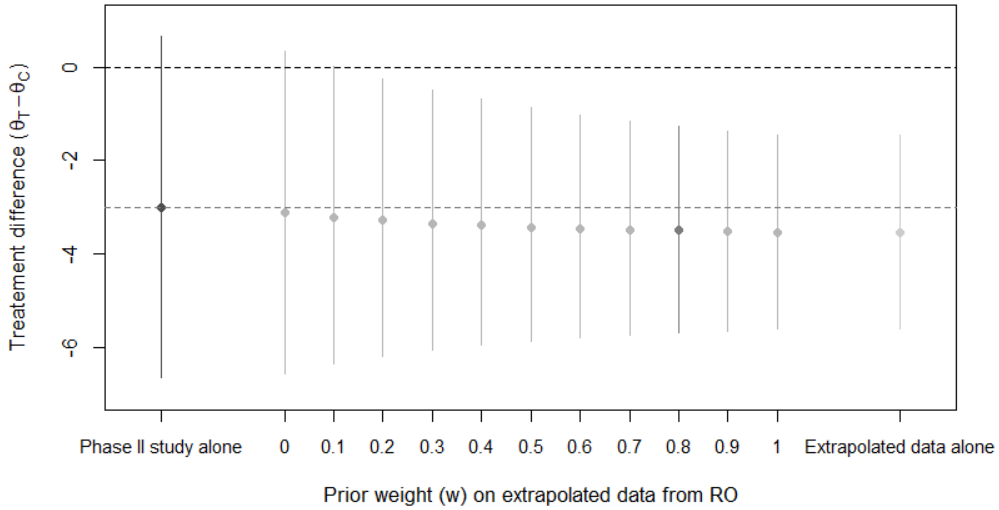

Figure 3: Sensitivity analysis performed on fictive realistic data showing posterior mean and 80% CrI for the estimated treatment difference versus prior weight. The two dashed lines represents the two thresholds for the success criteria: $(\mathrm{P}[(\theta_T - \theta_C) < 0] > 0.9)$ and $(\mathrm{P}[(\theta_T - \theta_C) < -3] > 0.5)$.

4.3 Fictive Analysis

In order to prepare for the final analysis of the trial, realistic fictive results are simulated and analysed in Table 4. The 80% CrI are presented to ensure consistency with the first success criterion, corresponding to a one-sided type I error of 10%. The observed treatment mean is -4 with a 80% CrI of [-6.1,-1.9] in the $N_T = 30$ patients and the observed control mean is -1 with a 80% CrI of [-4,2] in the $N_C = 15$ patients. The analysis of the design without borrowing is presented on the bottom line, with an observed treatment difference of -3 and an 80% CrI containing zero. Therefore, none of the success criteria are met. On the top line, the Bayesian analysis combines evidence of the current study data and the extrapolated data from RO. The posterior treatment difference is -3.5, slightly influenced by the informative priors, with better precision than the analysis of the design without borrowing. The 80% CrI does not contain zero and both success criteria are met.

Sensitivity analyses like the tipping point analysis [Best et al., 2021] presented in Figure 3 are essential to understand and assess the robustness of the results. By exploring the impact of different weights on the conclusions, it provides valuable insights into the design's sensitivity to the prior belief that the extrapolated data from RO provide relevant information about the effects in the new study. In this case, the analysis shows that the BDB design's conclusions are robust to changes in the weight, as the first success criterion is met only for weights higher than $w = 0.1$, illustrating the design's reliability.

A false positive scenario is presented in the Supplementary Material to illustrate that incorrect decisions can be avoided with an appropriate analysis and the use of sensitivity ones.

5 Discussion

This work presents a methodology to incorporate data collected in a phase I study on RO in healthy volunteers as prior information for a phase II PoC study in patients. The results and the comprehensive operating characteristics of the design lead to the conclusion that incorporating extrapolated data from RO using a BDB design permits an increase of power in the study, while limiting the type I error at acceptable levels. This confirm the findings of previous works that used historical borrowing [Lim et al., 2018, Viele et al., 2014, van Rosmalen et al., 2017, Smith et al., 2020]. They also permit to assess the impact of the choice of the prior weight on the results, allowing for a more informed choice rather than relying solely on external belief [Saint-Hilary et al., 2018a]. The fictive analysis conducted illustrates how the results could be presented and how their robustness can be evaluated when study data are collected. Moreover, they illustrate how additional evidence from RO data may reduce the risk of having phase II results in a "consider zone" [Frewer et al., 2016], where no clear decision could be made about the drug development's continuation in later phases. Additional analyses presented in the Supplementary Material confirm the usefulness of such analyses. The incorporation of RO data contributes valuable evidence to the study, potentially doubling the study ESS. However, considering the impact of using an informative prior with $w = 1$ on the operating characteristics, this is unlikely to be employed. In more realistic scenarios, incorporating RO data results in sample size increases ranging from one-third to two-thirds of the study sample size.

Although details on the disease that motivated this work cannot be disclosed due to confidentiality, the statistical methodology presented in this paper is widely applicable to other drug developments where the efficacy of the drug is expected to be related to some RO. The approach can be a valuable tool for optimizing the design and analysis of such trials. It is important to note that it is intended for use in early phase studies and should not replace the need for later-stage randomized trials to confirm the effects of the drug.

The power of 50% in our example may raise concerns for some readers, but it is inherent to the utilization of a dual-criterion design based on statistical significance and clinical relevance ([Roychoudhury et al., 2018, Saint-Hilary et al., 2018b]). In dual-criterion designs, the trial's success is determined not only by achieving statistical significance but also by exceeding a clinically meaningful threshold for the treatment effect estimate. The power calculated at this threshold value is approximately 50% because, if the true parameter equals the threshold, there is an equal probability that the effect estimate will fall on either side of it.

The proposed methodology has some limitations, mainly related to the extrapolations from healthy volunteers to patients and from RO to the clinical endpoint, which require external evidence. Indeed, while the BDB design has several advantages over traditional designs, there are also potential issues that should be taken into account. One such issue is that the prior evidence used in the design may not be representative of the current study population or may be based on flawed or incomplete information. This can lead to biased estimates, especially if the relationship between the RO and the clinical endpoint is wrongly estimated. In such cases, the BDB design may inadvertently propagate errors or biases to the current study, leading to unwanted or misleading results. However, mPBPK models are well-recognized in the literature [Kostewicz et al., 2014] and regulatory frameworks [Food and Drug Administration, 2018, Food and Drug Administration, 2020] and their predictive performance can be validated on independent datasets [Sager et al., 2015].

Moreover, we make the assumption that the relationship between RO and the clinical endpoint has been previously estimated based on the literature or past studies. While this assumption should be true to some extent when the dose is intended to be chosen based on RO results, the estimation of this relationship may be lacking in practice and replaced by clinical assumptions. We would not recommend to select doses based on RO without a reliable empirical estimation of the relationship between RO and the clinical endpoint, therefore we acknowledge that this assumption is the main limitation of our proposed methodology.

In general, before taking a drug into humans, a translational framework should have demonstrated a clear relationship between exposure of the target to the drug, desired pharmacodynamic (biomarker) effect and model efficacy [Bradley, 2012]. RO is a marker that can provide information on the direction and magnitude of treatment activity and therefore may give a sensitive view of whether or not the drug is having the desired effect. However, there may be other biomarkers that are pathophysiologically closer to the clinical endpoint and thus considered more proximate, such as proof of principle (POP) biomarkers or PoC biomarkers [Bradley, 2012]. Using these biomarkers in addition to RO could further improve confidence in the relationship to the clinical endpoint. Surrogates measuring the pharmacological effect

of the treatment [Rolan, 1997], or available pre-clinical information, are also valuable to assist with the dose selection. In the example presented in this article, the PoC study is performed on a well established clinical endpoint in the targeted indication and is enriched with information borrowed from a surrogate. If RO was found not to be a good predictor of the clinical response, the risk of erroneous borrowing from RO data is mitigated by the BDB design, which has the advantage of discarding the prior data in case of prior data conflict. Furthermore, the impact of the strength of the relationship between RO and the clinical endpoint on the design properties could be further assessed, as demonstrated on probabilities of success [Saint-Hilary et al., 2018a].

It should be noted that the number of simulated patients in the fictive trial does affect the precision of the prior distributions, and if it exceeds the number of enrolled patients N, the prior distributions may have an effective sample size larger than N. To address this issue, downweighting the prior could be considered by increasing the variance of all components by the same factor, as demonstrated for example in the recent study by [Richeldi et al., 2022]. This downweighting approach allows for the new data to have a sufficient weight in cases where the prior has a high ESS. It helps to ensure that the posterior distribution appropriately reflects the agreement or conflict between the prior beliefs and the current data. Therefore, in our proposed methodology, simulating a number of patients equal to N allows for the consideration of prior-data conflicts while allowing the current data to guide decisions. The downweighting approach can be employed if the prior has a high ESS, ensuring that the new data has sufficient influence on the posterior distribution.

The proposed methodology builds upon the approach proposed in [Saint-Hilary et al., 2018a] by extending it to the setting of using healthy volunteers RO data. In the present work the focus is on the use of historical data from a phase I study in healthy volunteers to inform the design and analysis of a phase II PoC study in patients. One natural extension of the methodology is to consider a seamless Phase I/II design within a Bayesian framework, where data from the phase I stage of the trial inform the analysis of the phase II stage. As another future extension of this work, a longitudinal analysis could be conducted to describe the behaviour of the RO over time and its relationship with the clinical endpoint. Also, sensitivity analyses on the factor λ relating the RO in healthy volunteers and patients, with different assumptions about its distribution, may be conducted to assess its impact on the BDB design's operating characteristics. Besides, in the case study, a fixed independent distribution for a_i, b_i and τ_i in all patients $i = 1, ..., N$ is considered. However, different distributions, depending on covariates corresponding to the characteristics of the patients, may be considered and the joint distribution could be estimated [Saint-Hilary et al., 2018a]. Furthermore, a linear relationship between the logit of the RO and the clinical endpoint is presented, but other settings could be suitable, such as linear improvement in the clinical endpoint for the logit of the RO above a minimum level, or piece-wise relationships.

In conclusion, the proposed methodology is expected to be valuable to support decision-making in early phases where the number of patients is limited. It could be extended to other contexts and other biomarkers or activity criteria, demonstrating the possibility of increasing the efficiency of PoC trials by using historical information in drug development.

acknowledgements

The authors would like to thank the editor, an associate editor and two anonymous reviewers for their insightful comments which helped in enhancing the quality of this paper. This research is supported by Institut de Recherches Internationales Servier.

conflict of interest

The authors declare no potential conflicts of interest.

References

[Best et al., 2021] Best, N., Price, R., Pouliquen, I., and Keene, O. (2021). Assessing efficacy in important subgroups in confirmatory trials: An example using bayesian dynamic borrowing. *Pharmaceutical Statistics*, 20(3):551–562. DOI: 10.1002/pst.2093.

[Bradley, 2012] Bradley, E. (2012). Incorporating biomarkers into clinical trial designs: points to consider. *Nature Biotechnology*, 30(7):596–599. DOI: 10.1038/nbt.2296.

[Burger et al., 2021] Burger, H. U., Gerlinger, C., Harbron, C., Koch, A., Posch, M., Rochon, J., and Schiel, A. (2021). The use of external controls: To what extent can it currently be recommended? *Pharmaceutical Statistics*. DOI: 10.1002/pst.2120.

[Cao and Jusko, 2014] Cao, Y. and Jusko, W. J. (2014). Incorporating target-mediated drug disposition in a minimal physiologically-based pharmacokinetic model for monoclonal antibodies. *Journal of Pharmacokinetics and Pharmacodynamics*, 41(4):375–387. DOI: 10.1007/s10928-014-9372-2.

[Chan et al., 2021] Chan, P. L., McFadyen, L., Quaye, A., Leister-Tebbe, H., Hendrick, V. M., Hammond, J., and Raber, S. (2021). The use of extrapolation based on modeling and simulation to support high-dose regimens of ceftaroline fosamil in pediatric patients with complicated skin and soft-tissue infections. *CPT: Pharmacometrics & Systems Pharmacology*, 10(6):551–563. DOI: 10.1002/psp4.12608.

[Christopoulos and El-Fakahany, 1999] Christopoulos, A. and El-Fakahany, E. E. (1999). Qualitative and quantitative assessment of relative agonist efficacy. *Biochemical Pharmacology*, 58(5):735–748. DOI: 10.1016/s0006-2952(99)00087-8.

[Chuang-Stein and Kirby, 2017] Chuang-Stein, C. and Kirby, S. (2017). *Quantitative Decisions in Drug Development*. Cham, Switzerland: Springer International Publishing AG. DOI: 10.1007/978-3-319-46076-5.

[Food and Drug Administration, 2018] Food and Drug Administration (2018). Physiologically based pharmacokinetic analyses — format and content guidance for industry.

[Food and Drug Administration, 2020] Food and Drug Administration (2020). The use of physiologically based pharmacokinetic analyses — biopharmaceutics applications for oral drug product development, manufacturing changes, and controls - draft guidance for industry.

[Frewer et al., 2016] Frewer, P., Mitchell, P., Watkins, C., and Matcham, J. (2016). Decision-making in early clinical drug development. *Pharmaceutical Statistics*, 15(3):255–263. DOI: 10.1002/pst.1746.

[Ghadessi et al., 2020] Ghadessi, M., Tang, R., Zhou, J., Liu, R., Wang, C., Toyoizumi, K., Mei, C., Zhang, L., Deng, C. Q., and Beckman, R. A. (2020). A roadmap to using historical controls in clinical trials – by drug information association adaptive design scientific working group (DIA-ADSWG). *Orphanet Journal of Rare Diseases*, 15(1). DOI: 10.1186/s13023-020-1332-x.

[Jones and Rowland-Yeo, 2013] Jones, H. and Rowland-Yeo, K. (2013). Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacometrics Syst Pharmacol*, 2(8):e63. DOI: 10.1038/psp.2013.41.

[Jones et al., 2022] Jones, N., Ngo, B., Fancke, B., Limson, J., and Davis, F. (2022). Abstract 4041: Use of receptor occupancy assays on cells from peripheral blood as a surrogate model of the tumor microenvironment. *Cancer Res*, 82(12 Supplement):832–843. DOI: 10.1158/1538-7445.AM2022-4041.

[Jonsson et al., 2005] Jonsson, E. N., Macintyre, F., James, I., Krams, M., and Marshall, S. (2005). Bridging the pharmacokinetics and pharmacodynamics of UK-279, 276 across healthy volunteers and stroke patients using a mechanistically based model for target-mediated disposition. *Pharmaceutical Research*, 22(8):1236–1246. DOI: 10.1007/s11095-005-5264-x.

[Junker et al., 2021] Junker, F., Gulati, P., Wessels, U., Seeber, S., Stubenrauch, K., Codarri-Deak, L., Markert, C., Klein, C., Teixeira, P. C., and Kao, H. (2021). A human receptor occupancy assay to measure anti-pd-1 binding in patients with prior anti-pd-1. *Cytometry A*, 99(8):832–843. DOI: 10.1002/cyto.a.24334.

[Khosravan et al., 2021] Khosravan, R., DuBois, S. G., Janeway, K., and Wang, E. (2021). Extrapolation of pharmacokinetics and pharmacodynamics of sunitinib in children with gastrointestinal stromal tumors. *Cancer Chemotherapy and Pharmacology*, 87(5):621–634. DOI: 10.1007/s00280-020-04221-x.

[Kostewicz et al., 2014] Kostewicz, E. S., Abrahamsson, B., Brewster, M., Brouwers, J., Butler, J., Carlert, S., Dickinson, P. A., Dressman, J., Holm, R., Klein, S., Mann, J., McAllister, M., Minekus, M., Muenster, U., Müllertz, A., Verwei, M., Vertzoni, M., Weitschies, W., and Augustijns, P. (2014). In vitro models for the prediction of in vivo performance of oral dosage forms. *European Journal of Pharmaceutical Sciences*, 57:342–366. DOI: 10.1016/j.ejps.2013.08.024.

[La Gamba et al., 2019] La Gamba, F., Jacobs, T., Geys, H., Jaki, T., Serroyen, J., Ursino, M., Russu, A., and Faes, C. (2019). Bayesian sequential integration within a preclinical pharmacokinetic and pharmacodynamic modeling framework: Lessons learned. *Pharmaceutical Statistics*. DOI: 10.1002/pst.1941.

[Lalonde et al., 2007] Lalonde, R. L., Kowalski, K. G., Hutmacher, M. M., Ewy, W., Nichols, D. J., Milligan, P. A., Corrigan, B. W., Lockwood, P. A., Marshall, S. A., Benincosa, L. J., Tensfeldt, T. G., Parivar, K., Amantea, M., Glue, P., Koide, H., and Miller, R. (2007). Model-based drug development. *Clinical Pharmacology & Therapeutics*, 82(1):21–32. DOI: 10.1038/sj.clpt.6100235.

[Liang et al., 2016] Liang, M., Schwickart, M., Schneider, A. K., Vainshtein, I., Nagro, C. D., Standifer, N., and Roskos, L. K. (2016). Receptor occupancy assessment by flow cytometry as a pharmacodynamic biomarker in biopharmaceutical development. *Cytometry Part B*, 90(B):117–127. DOI: 10.1002/cyto.b.21259.

[Lim et al., 2018] Lim, J., Walley, R., Yuan, J., Liu, J., Dabral, A., Best, N., Grieve, A., Hampson, L., Wolfram, J., Woodward, P., Yong, F., Zhang, X., and Bowen, E. (2018). Minimizing patient burden through the use of historical subject-level data in innovative confirmatory clinical trials: Review of methods and opportunities. *Therapeutic Innovation & Regulatory Science*, 52(5):546–559. DOI: 10.1177/2168479018778282.

[Meyer et al., 2022] Meyer, E. L., Mesenbrink, P., Dunger-Baldauf, C., Glimm, E., Li, Y., and and, F. K. (2022). Decision rules for identifying combination therapies in open-entry, randomized controlled platform trials. *Pharmaceutical Statistics*, 21(3):671–690.

[Natanegara et al., 2013] Natanegara, F., Neuenschwander, B., Seaman, J. W., Kinnersley, N., Heilmann, C. R., Ohlssen, D., and Rochester, G. (2013). The current state of bayesian methods in medical product development: survey results and recommendations from the DIA bayesian scientific working group. *Pharmaceutical Statistics*, 13(1):3–12. DOI: 10.1002/pst.1595.

[Neuenschwander et al., 2010] Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7(1):5–18. DOI: 10.1177/1740774509356002.

[Neuenschwander et al., 2011] Neuenschwander, B., Rouyrre, N., Hollaender, N., Zuber, E., and Branson, M. (2011). A proof of concept phase ii non-inferiority criterion. *Statistics in Medicine*, 30(13):1618–27. DOI: 10.1002/sim.3997.

[Neuenschwander et al., 2020] Neuenschwander, B., Weber, S., Schmidli, H., and O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. *Biometrics*, 76(2):578–587. DOI: 10.1111/biom.13252.

[Radanovic et al., 2022] Radanovic, I., Klarenbeek, N., Rissmann, R., Groeneveld, G. J., van Brummelen, E. M. J., Moerland, M., and Bosch, J. J. (2022). Integration of healthy volunteers in early phase clinical trials with immuno-oncological compounds. *Frontiers in Oncology*, 12. DOI: 10.3389/fonc.2022.954806.

[Richeldi et al., 2022] Richeldi, L., Azuma, A., Cottin, V., Hesslinger, C., Stowasser, S., Valenzuela, C., Wijsenbeek, M., Zoz, D., Voss, F., Maher, T., and "1305-0013 Trial Investigators" (2022). Trial of a preferential phosphodiesterase 4b inhibitor for idiopathic pulmonary fibrosis. *N Engl J Med*, 386(23):2178–2187. DOI: 10.1056/NEJMoa2201737.

[Rolan, 1997] Rolan, P. (1997). The contribution of clinical pharmacology surrogates and models to drug development-a critical appraisal. *British Journal of Clinical Pharmacology*, 44(3):219–225. DOI: 10.1046/j.1365-2125.1997.t01-1-00583.x.

[Roychoudhury et al., 2018] Roychoudhury, S., Scheuer, N., and Neuenschwander, B. (2018). Beyond p-values: A phase ii dual-criterion design with statistical significance and clinical relevance. *Clinical Trials*, 15(5):452–461. DOI: 10.1177/1740774518770661.

[Sager et al., 2015] Sager, J. E., Yu, J., Ragueneau-Majlessi, I., and Isoherranen, N. (2015). Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: A systematic review of published models, applications, and model verification. *Drug Metabolism and Disposition*, 43(11):1823–1837. DOI: 10.1124/dmd.115.065920.

[Saint-Hilary et al., 2018a] Saint-Hilary, G., Barboux, V., Pannaux, M., Gasparini, M., Robert, V., and Mastrantonio, G. (2018a). Predictive probability of success using surrogate endpoints. *Statistics in Medicine*, 38(10):1753–1774. DOI: 10.1002/sim.8060.

[Saint-Hilary et al., 2018b] Saint-Hilary, G., Robert, V., and Gasparini, M. (2018b). Decision-making in drug development using a composite definition of success. *Pharm Stat*, 17(5):555–569. DOI: 10.1002/pst.1870.

[Schmidli et al., 2014] Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032. DOI: 10.1111/biom.12242.

[Smith et al., 2020] Smith, C. L., Thomas, Z., Enas, N., Thorn, K., Lahn, M., Benhadji, K., and Cleverly, A. (2020). Leveraging historical data into oncology development programs: Two case studies of phase 2 bayesian augmented control trial designs. *Pharmaceutical Statistics*, 19(3):276–290. DOI: 10.1002/pst.1990.

[Srinivas et al., 2018] Srinivas, N., Maffuid, K., and Kashuba, A. (2018). Clinical pharmacokinetics and pharmacodynamics of drugs in the central nervous system. *Clin Pharmacokinet*, 57(9):1059–1074. DOI: 10.1007/s40262-018-0632-y.

[van Rosmalen et al., 2017] van Rosmalen, J., Dejardin, D., van Norden, Y., Löwenberg, B., and Lesaffre, E. (2017). Including historical data in the analysis of clinical trials: Is it worth the effort? *Statistical Methods in Medical Research*, 27(10):3167–3182. DOI: 10.1177/0962280217694506.

[Viele et al., 2014] Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., and Thompson, L. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13(1):41–54. DOI: 10.1002/pst.1589.

[Willmann et al., 2021] Willmann, S., Marostica, E., Snelder, N., Solms, A., Jensen, M., Lobmeyer, M., Lensing, A. W. A., Bethune, C., Morgan, E., Yu, R. Z., Wang, Y., Jung, S. W., Geary, R., and Bhanot, S. (2021). PK/PD modeling of FXI antisense oligonucleotides to bridge the dose-FXI activity relation from healthy volunteers to end-stage renal disease patients. *CPT: Pharmacometrics & Systems Pharmacology*, 10(8):890–901. DOI: 10.1002/psp4.12663.