

Bayesian Borrowing Approaches in the Context of Clinical Trials for Rare Diseases - The example of slow-flow vascular malformations

By: Marina Saleeb

Under the supervision of DR/ Catherine Legrand

Table of Contents

1. Notations	4
2. Abstract	7
3. Introduction	9
3.1 Rare diseases.....	9
3.2 Slow-Flow Vascular Malformations.....	10
3.3 Bayesian Statistics and the concept of borrowing	10
3.3.1 The concept of borrowing	11
3.3.2 Dynamic borrowing	12
4. Literature Review of the statistical methods used for borrowing	14
4.1 Prior-based family	16
4.1.1 Power Priors	16
4.1.2 Commensurate prior	17
4.1.3 Meta-analytic predictive prior (MAP)	18
4.1.4 Robust Meta-analytic predictive priors (rMAP)	20
4.1.5 Self-Adaptive Mixture Priors (SAM).....	20
4.1.6 Elastic prior	23
4.2 Joint likelihood-based family	24
4.2.1 Shrinkage borrowing via standard Bayesian Hierarchical Modeling (BHM) and Normal-Normal Hierarchical Model (NNHM)	24
4.3 Practical implementation of the Bayesian borrowing methods for different types of endpoints.....	26
4.4 Implementation of every method in R.....	27
5. Materials and methods	30
5.1 Application to simulated data.....	30
5.1.1 Overview	30
5.1.2 Current Study Data Generation	30
5.1.3 Historical Studies generation	31
5.1.4 Final datasets (Four scenarios)	31
5.1.5 Bayesian Borrowing Methods	32
5.2 Real-World data (RWD)	34
5.2.1 Identification of the relevant studies.....	34
5.2.2 Data extraction	36

5.2.3 Quality-adjusted Similarity weights	40
5.2.4 Bayesian borrowing analysis and the inclusion of the pre-borrowing weights....	43
5.2.5 Sensitivity analysis	44
5.2.6 Point estimates and credible intervals calculation.....	45
5.2.7 Effective sample size (ESS)	46
6. Results	48
6.1 Application to simulated datasets	48
6.2 Real-World data	54
6.2.1 Binary endpoint.....	54
6.2.2 Continuous endpoint.....	59
7. Discussion	64
8. Strengths and limitations	67
9. Conclusion.....	68
10. References	69
11. Supporting Information	75
11.1 Supporting material I.....	75
11.2 Supporting material II.....	75
11.3 Supporting material III.....	75
11.4 Supporting material IV	75
11.5 Supporting material V	75
11.6 Supporting material VI	75
11.7 Supporting material VII.....	76
11.8 Supporting material VIII.....	76
11.9 Supporting material IX.....	76
12. Annex.....	77
12.1 Simulation	77
12.2 Bayesian simulated data	81
12.3 pre borrowing weights.....	118
12.4 RWD analysis (Binary endpoint)	126
12.5 RWD analysis (Continuous endpoint)	146

1. Notations

Category	Notation	Meaning / description
Prior-based family		
Data	D_c	Data from current study
	D_h	One historical dataset
	$D_{h_{1:j}}$	Historical dataset indexed by $j, j = 1, \dots, J$.
	J	Number of available historical datasets.
Population-level / true parameters	θ_c	True treatment effect parameter to be estimated in the current study (e.g., log-OR, log-HR, mean difference).
	θ_h	True effect from one historical dataset
	θ_{h_j}	True treatment-effect parameter for historical study $j, j = 1, \dots, J$.
Power prior	a_0	Power/discount factor controlling historical borrowing, $0 \leq a_0 \leq 1$.
	a_{0j}	Study- specific discount factor for the historical study j
Commensurate priors	τ	Commensurability precision controlling similarity between θ_c and historical θ_h
	τ_{h_j}	Study-specific commensurability precision controlling the similarity between θ_c and $\theta_{h_1}, \theta_{h_2}, \dots, \theta_{h_j}$
MAP/rMAP/SAM	μ_h	Overall mean effect across the historical studies

	σ_h^2	Between-study heterogeneity across the historical studies
rMAP/SAM	w	Mixture weight
SAM	Δ	Pre-defined congruence/equivalence margin
SAM/elastic	$\hat{\mu}_h$	Mean of the historical posterior distribution of θ_h
	$\hat{\sigma}_h^2$	Variance of the historical posterior distribution of θ_h
Elastic prior	T	Similarity/congruence statistic.
	$g(\cdot)$	Monotone function mapping similarity statistic T to a weight in (0,1).
	a, b	Calibration/tuning parameters linking T and $g(T)$
Joint likelihood-based family		
Population parameters	θ_j	True treatment effect in study j (here all the studies (current & historical) are given the notation: $j, j = 1, \dots, J$), **In this family, the notations c and h for current and historical data will not be used anymore**
	μ	Population-level mean treatment effect across studies
	σ^2	Between-study variance (heterogeneity) of true effects
Observed data	D_j	Observed data from study j (current & historical)

	y_j	Observed treatment-effect estimate from study j (e.g. log-OR, log-HR, mean difference)
	σ_j^2	Within-study variance of y_j (assumed to be known)
NNHM	w_j	Shrinkage weight across study j (current & historical)

** I used AI to help me improve my writing style.

2. Abstract

Background:

Clinical trials of rare diseases are often limited by small sample sizes, which restrict the precision and reliability of treatment effect estimates. Slow-flow vascular malformations, as rare and clinically heterogeneous disorders, pose significant challenges for clinical trials because of limited patient populations, variable disease presentation, and diverse treatment responses. In this context, Bayesian borrowing approaches provide a principled framework for incorporating external or historical data with current study data. These methods include both prior-based approaches, which explicitly model prior information, and likelihood-based hierarchical models, which borrow information through partial pooling across studies.

Methods:

An exploratory simulation study was conducted based on two realizations of simulated datasets, representing a large ($n = 100$) and a small ($n = 20$) current studies, with 10 historical studies exhibiting varying degrees of similarity. These scenarios were designed to assess borrowing behavior under different levels of similarity. In addition, a real-world analysis was performed using 19 historical studies extracted from the literature to reinforce the estimation of treatment effects from a study evaluating the effect of sirolimus in patients with slow-flow vascular malformations (our current study), focusing on two endpoints: pain response after 12 months and D-dimer decrease after 12 months. A two-stage framework was applied, combining pre-borrowing weights, based on clinical, demographic, and methodological similarity with the current study, with four Bayesian borrowing methods: robust meta-analytic-predictive prior (rMAP), Self-adaptive mixture prior (SAM), elastic prior, and normal-normal hierarchical model (NNHM).

Results:

Simulation results suggest that all methods adapt the level of borrowing according to the degree of similarity between historical and current data, with increased effective sample size (ESS) and narrower intervals under high similarity levels, and reduced borrowing when the similarity decreases. In the real-world analysis, minimal borrowing was observed for the pain response endpoint, with posterior estimates remaining close to the current study estimate of 76.2%. For the D-dimer endpoint, all borrowing methods show an upward shift in the posterior mean relative to the current-only estimate. Moderate borrowing methods (rMAP and SAM) yield posterior means around 1.19–1.20, with credible intervals approximately 0.28–0.29 to 2.09–2.10, corresponding to moderate precision gains (relative ESS ≈ 14 –16%). The Elastic models remain closer to the current-only estimate, with posterior means around 1.15 with credible intervals from 0.23–2.07 and smaller borrowing gains (relative ESS ≈ 10 –11%). In contrast, NNHM produces the largest shift, with posterior means around 1.55 and credible intervals approximately 0.60 to 2.44, reflecting the strongest borrowing from historical data and the highest precision gains (relative ESS ≈ 19 –24%).

Conclusion:

Bayesian dynamic borrowing methods provide a flexible framework for integrating historical data while adapting to the level of agreement between current and historical studies. The proposed two-stage approach provides a structured way to account for clinical relevance and methodological quality before integrating external data. Although borrowing led to modest gains in precision and small shifts in estimates, it did not substantially alter clinical conclusions. The recommendations for future work include performing large scale simulations with varying current study sample sizes and different levels of agreement between historical and current data.

Keywords: Rare diseases, Bayesian borrowing, Dynamic borrowing, Effective sample size, Hierarchical models, Historical data

3. Introduction

3.1 Rare diseases

Rare diseases, sometimes called orphan diseases, are defined as conditions affecting fewer than one person in every 2,000 individuals in any WHO region (1). Despite the term "Rare Diseases", their worldwide prevalence is estimated at 4-7.6% and this number is equivalent to 329-624 million people collectively (2). In Belgium alone, from 600,000 to 800,000 persons suffer from a rare disease and this estimate is conservative as reported by Sciensano (2). According to a 2024 editorial in The Lancet (1), nearly 80% of rare diseases have a genetic origin, 70% of which manifest in early childhood, and approximately 95% currently lack an approved treatment. The challenges surrounding rare diseases extend beyond clinical complexity, their diagnosis and management are difficult because clinicians often encounter conditions they have never seen before, and patients face substantial health, economic, and psychosocial burdens. Many individuals living with rare diseases experience poor quality of life, social isolation, and higher levels of psychological distress due to the chronic nature of their conditions, the limited therapeutic options, and the persistent discrimination (3).

The absence of effective treatments for the vast majority of rare diseases (4) is closely linked to the obstacles encountered throughout the drug development and, more specifically, in conducting methodologically robust clinical trials. Research in rare diseases is hindered by small and heterogeneous populations, which often leads to underpowered studies and challenges in identifying reliable clinical endpoints (5). Additional difficulties are related to the variations in phenotypes and genotypes within the same disease which leads to high heterogeneity among the population. Furthermore, the practical difficulties associated with recruiting a control group as these diseases are serious conditions and patients often refuse to be in a control group. Consequently, many clinical trials of rare diseases adopt single-arm designs (6) since traditional randomized controlled trials (RCTs) are not feasible.

Therefore, the Food and Drug Administration (FDA) has recognized these challenges in its guidance on common issues in rare disease drug development (7) where it is stated that, if patient availability is limited, alternative development strategies have to be adopted in both the design and the analysis phases of the trial. For example, externally-controlled clinical trials, and adaptive clinical trial designs are explicitly stated and accepted by the FDA (8) (9).

Besides, the FDA has also endorsed the use of Bayesian methods in clinical development because they provide a natural framework for integrating external evidence with the current data, making them particularly suitable for rare disease settings. In fact, the FDA now accepts the use of external or synthetic control arms and Bayesian borrowing approaches that incorporate information from historical trials, registries, or real-world evidence to improve the precision and efficiency of underpowered trials. However, these approaches require rigorous justification of key assumptions, including the comparability of external

and current patient populations, consistency in outcome definitions and measurement, similarity in study design and standard of care, and the absence of substantial bias that could compromise the validity of the borrowed information (8).

Likewise, the European Medicines Agency (EMA) promotes methodological flexibility in rare diseases and recognizes that producing standard trial evidence may be impossible in small populations. This is clearly mentioned in the EMA's Guideline on clinical trials in small populations, which sets out regulatory principles for the design, conduct, analysis, and interpretation of trials when recruitment is limited. Additionally, the EMA supports the use of adaptive study designs and sophisticated modelling in drug development, especially when these approaches are scientifically reasonable (10).

3.2 Slow-Flow Vascular Malformations

Slow-flow vascular malformations are complex congenital disorders arising from developmental anomalies of veins, lymphatic vessels, or capillaries (11). Lesions may be present at birth and grow proportionally or unpredictably throughout life. They often cause chronic pain, swelling, bleeding, disfigurement, and functional impairment, depending on their size and location (12). Clinical presentation varies substantially between patients, reflecting the broad anatomical and biological heterogeneity that complicates timely diagnosis and selection of optimal treatment strategies. Quality of life is often markedly reduced due to physical disability, psychological distress, and social limitations imposed by the chronic and disfiguring nature of the disease (13).

Conventional therapeutic options including sclerotherapy, surgery, or laser are only partially effective and often associated with recurrence or complications. Sirolimus, a mechanistic target of rapamycin (mTOR) inhibitor, has emerged as one of the most promising systemic treatments for complicated vascular malformations by targeting the dysregulated phosphoinositide 3-kinase/protein kinase B/mechanistic target of rapamycin (PI3K/AKT/mTOR) signaling pathway implicated in their pathophysiology (14). However, despite encouraging clinical outcomes, evidence supporting such use comes largely from small, heterogeneous observational studies, case series, and single-arm early-phase trials. The field still lacks large, randomized, or long-term controlled studies. Therefore, rigorous statistical methods, such as Bayesian borrowing, are critical to generate robust inference from limited data.

3.3 Bayesian Statistics and the concept of borrowing

In this landscape, Bayesian statistics have become increasingly valuable. The Bayesian framework offers a formal mechanism for updating our beliefs about a parameter using observed data. It begins with prior knowledge, which is mathematically represented by a prior distribution, and incorporates new information through the likelihood function, yielding an updated estimate known as the posterior distribution (15). This relationship is expressed through Bayes' theorem:

$$\pi(\theta | D) = \frac{\pi(D | \theta)\pi_0(\theta)}{\pi(D)}$$

where

- θ is the parameter of interest,
- $\pi_0(\theta)$ is the prior (our belief before observing the data),
- $\pi(D | \theta)$ is the likelihood (compatibility of the data with the parameter),
- $\pi(D)$ is the marginal likelihood (normalizing constant), and
- $\pi(\theta | D)$ is the posterior (the updated belief).

A key strength of Bayesian methods in rare disease research is their ability to borrow information from external or historical sources.

In most cases, Bayes' theorem is written in the following proportional form:

$$\pi(\theta | D) \propto \pi(D | \theta) \pi_0(\theta),$$

or equivalently,

$$\pi(\theta | D) \propto L(\theta | D) \pi_0(\theta),$$

where $\propto$ denotes proportionality. This notation indicates that the posterior distribution is proportional to the product of the likelihood and the prior, up to a normalizing constant that does not depend on θ . This constant is typically omitted for notational simplicity. The two expressions are mathematically identical, since the likelihood function $L(\theta | D)$ is defined as the probability (or density) of the observed data given θ , viewed as a function of θ for fixed data.

3.3.1 The concept of borrowing

The term "borrowing" (especially in Bayesian analysis) indicates acquiring information from external sources like past studies, registries, historical controls, or real-world evidence to strengthen the estimation of the treatment effect of the ongoing trial, especially if the ongoing trial has a small number of participants.

Borrowing helps in increasing the precision of the treatment effect estimate, tightening credible intervals, increasing power and decision-making confidence when current data is weak (16) (17). The idea of borrowing can be illustrated with a very basic example: Consider a rare disease trial with a current sample of $n = 20$ patients and an observed response proportion of 60%. Because of the small sample size, the estimate is associated with substantial uncertainty and high variability. Suppose that relevant historical evidence is available, such as a registry or a prior study including $n = 100$ patients with a response rate of 55%. If the populations, endpoints, and study context are reasonably comparable, the historical information can be incorporated to strengthen the inference by combining evidence from both the current and past data sources. Importantly, borrowing does not imply a complete pooling or simple merging of the datasets. Instead, dynamic borrowing typically operates through partial pooling, whereby the historical information is down-weighted according to its compatibility with the current study.

3.3.2 Dynamic borrowing

Dynamic borrowing (18) is a term that describes the methods where the similarity of the historical evidence to the current study data determines the amount of information taken from the past studies. If the two sources are consistent, maximum borrowing takes place; if they are inconsistent, borrowing is minimized to a small level. This principle gives rise to the application of such flexible borrowing techniques such as robust meta-analytic predictive (rMAP) priors, commensurate priors, and associated robust prior formulations.

The extent of dynamic borrowing is often summarized using the Effective Sample Size (ESS) (19), which reflects the amount of information contributed by the historical data after discounting. Conceptually, the ESS can be interpreted as the number of **“additional patients”** that the historical data effectively contribute to the current analysis after the model determines how much that information should be trusted. In other words, the ESS increases when historical data are similar to the current study and decreases when they are different. In the previous example, if there is notable dissimilarity between the historical and current studies, the model will automatically discount the historical evidence, so that the 100 historical observations may only contribute information equivalent to, for instance, only 5 effective patients rather than their full sample size.

Although borrowing information from historical data can improve estimation accuracy and increase statistical efficiency in small-sample settings such as rare disease trials, inappropriate borrowing may also introduce substantial bias. This problem arises when the historical data are not sufficiently compatible with the current study population (20). Differences in patient characteristics, disease severity, or endpoint definitions between the historical and current studies may lead to systematic discrepancies between the two data sources. In such situations, excessive borrowing can shift the posterior estimates toward values that do not accurately reflect the current population, potentially resulting in misleading inference. Bias could also arise due to the presence of hidden or unmeasured confounding factors in the historical data. For example, changes over time in diagnostic criteria, supportive care, or patient selection procedures may affect outcomes in ways that are not explicitly captured in the model. Since these latent differences cannot always be directly quantified, naive borrowing strategies that treat historical and current data as fully exchangeable may produce biased posterior estimates (21).

For this reason, the use of dynamic borrowing methods is generally preferred since they protect against prior-data conflict by discounting historical information when incompatibility is detected between the historical and current data, thereby reducing the influence of potentially misleading external evidence.

Building on this framework, the present work focuses on slow-flow vascular malformations and the evaluation of sirolimus in the context of the Vascular Anomaly–Sirolimus–Europe (VASE) trial, a single-arm phase III clinical study involving both paediatric and adult patients (22). The goal of this work is to summarize and examine Bayesian borrowing methodologies to evaluate the effect of sirolimus in the context of the VASE trial. Specifically, this work has two objectives:

1. Exploring and comparing different Bayesian borrowing methods across different levels of similarity and different endpoints, and
2. Borrowing treatment-effect estimates from previous historical studies to enhance the precision of effect estimation in our current trial.

4. Literature Review of the statistical methods used for borrowing

Several Bayesian approaches have been applied in clinical trials for rare diseases to address challenges such as small sample sizes and the absence of a control group. We can divide the Bayesian methods used in the context of borrowing into **2 main families** based on the mechanism through which borrowing occurs.

Firstly, **the prior-based family** which can be presented with this formula:

$$\pi(\theta_c | D_c, D_h) \propto L(\theta_c | D_c) [f(L(\theta_c | D_h) \pi_0(\theta_c))]$$

In this formulation, θ_c denotes the treatment-effect parameter for the current study, such as log-odds ratio, log-hazard ratio, or mean difference. The term $L(\theta_c | D_c)$ is the likelihood function of the current study data, this term remains unchanged across all prior-based borrowing methods. Historical data D_h do not enter the model through an additional or joint likelihood term. Instead, information from historical studies is incorporated through the construction of an informative prior distribution for θ_c . This is represented by the effective prior $f(L(\theta_c | D_h)) \pi_0(\theta_c)$ which combines a baseline prior $\pi_0(\theta_c)$ with information derived from the historical likelihood $L(\theta_c | D_h)$. The function $f(\cdot)$ denotes a method-specific borrowing mechanism that determines how historical evidence influences the prior. Different choices of $f(\cdot)$ correspond to different prior-based borrowing approaches, such as power discounting in the Power Prior, predictive distributions in the Meta-Analytic Predictive (MAP) Prior, or robust mixture formulations in robust MAP and SAM Priors. While the functional form of $f(\cdot)$ varies across methods, its role is always to modify the prior of the current study rather than its likelihood. Consequently, historical evidence influences inference indirectly through the prior distribution for θ_c , rather than through a joint likelihood involving both current and historical data. This ensures that the likelihood contribution from the current study remains intact, while borrowing is controlled entirely through prior specification. Owing to this structure, prior-based borrowing methods are particularly well suited for strengthening or stabilizing estimation in small-sample or single-arm settings, where information from the current study alone may be limited.

Secondly, **the joint likelihood-based family** can be represented by the following formulation:

$$\pi(\theta_{1:J}, \mu, \sigma^2 \mid D_{1:J}) \propto \prod_{j=1}^J L(\theta_j \mid D_j) \prod_{j=1}^J \pi(\theta_j \mid \mu, \sigma^2) \pi_0(\mu, \sigma^2).$$

In this formulation, each study $j = 1, \dots, J$, whether historical or current, contributes to the analysis through its own likelihood term $L(\theta_j \mid D_j)$. Historical data therefore enter the model directly through the joint likelihood across all studies. Borrowing of information across studies is achieved by linking the study-specific parameters θ_j through shared hyperparameters, namely a population-level mean μ and a between-study variance σ^2 . This hierarchical structure induces partial pooling (shrinkage) of the study-specific effects θ_j toward the common mean μ , with the degree of borrowing governed by the magnitude of σ^2 . Smaller values of σ^2 correspond to stronger borrowing and greater shrinkage across studies, while larger values allow for increased between-study heterogeneity and therefore, weaker borrowing and lower shrinkage. Across all the different approaches belonging to this family, the joint likelihood component (the green part of the equation) remains fixed, as all studies are modelled simultaneously. What changes across methods is the hierarchical prior specification (the grey part of the equation) that links the study-specific parameters θ_j . Different modelling choices for this hierarchical prior give rise to different joint likelihood-based borrowing approaches, such as the Normal–Normal Hierarchical Model and its extensions, including bias-corrected formulations.

There is a second classification of the Bayesian borrowing approaches into **supervised** and **unsupervised** approaches (23), depending on how the degree of borrowing is determined. The unsupervised methods, including the power prior, the commensurate priors, the Meta-Analytic Predictive (MAP) prior, and the Normal–Normal Hierarchical Model (NNHM), determine the extent of borrowing implicitly through the statistical model and the observed data. In these approaches, borrowing is typically governed by parameters such as between-study heterogeneity or model-based weights, without explicit incorporation of clinical judgment. However, in supervised methods, such as the Self-Adaptive Mixture (SAM) prior and the elastic prior, the amount of borrowing is explicitly regulated by user-defined rules or expert judgements that incorporate clinical expertise, for example through discrepancy measures, or discounting functions. These methods allow investigators to control how similarity between historical and current data is assessed and how rapidly borrowing should decrease in the presence of disagreement. These will be explained in details in the following sections.

4.1 Prior-based family

4.1.1 Power Priors

The power prior (24) (25) is one of the simplest Bayesian methods for incorporating information from historical data into the current analysis. It introduces a discount parameter $a_0 \in [0,1]$, which specifies the proportion of information borrowed from the historical data when combining it with the current study as follow:

$$\pi(\theta_c \mid D_c, D_h, a_0) \propto L(\theta_c \mid D_c) (L(\theta_c \mid D_h))^{a_0} \pi_0(\theta_c)$$

In this formulation, the prior for θ_c is constructed as the product of a non-informative prior $\pi_0(\theta_c)$ and the likelihood of the historical data $L(\theta_c \mid D_h)$ raised to the power a_0 . Through this construction, an informative prior is obtained that controls the degree of borrowing from historical evidence. When $a_0 = 1$, the historical data are fully incorporated, effectively treating them as fully compatible with the current study. Conversely, when $a_0 = 0$, the historical data are completely discounted and do not contribute to inference. Values of a_0 between 0 and 1 correspond to partial borrowing, allowing for the control the degree of influence of the historical data on the posterior distribution.

The power prior was initially developed to strengthen the effect of a single historical study D_h (as shown in the above formula), however it can be extended to include several studies.

When multiple historical studies $D_{h1}, \dots, D_{hJ}$ are available, the likelihood contributions from all historical studies can be incorporated in the model following this equation:

$$\pi(\theta_c \mid D_c, D_{h1:J}, a_0) \propto L(\theta_c \mid D_c) \prod_{j=1}^J [L(\theta_c \mid D_{h_j})]^{a_0} \pi_0(\theta_c)$$

where a single global discount parameter a_0 is applied to all historical studies.

Alternatively, study-specific discount parameters a_{0j} may be introduced to allow different degrees of borrowing from each historical study, reflecting varying levels of similarity between each historical study and the current one.

The main challenge here is how to choose an appropriate value for a_0 . Fixing it subjectively can introduce over-borrowing or under-borrowing bias to the final result (24), while estimating it from the data increases flexibility but adds some computational complexity. To address this issue, we can use the **Normalized Power Prior** (26)

4.1.1.1 Normalized Power Priors

In the normalized power prior, the discount parameter a_0 is treated as a random variable rather than being fixed in advance. A prior distribution is assigned to a_0 , commonly a Beta distribution, such as $a_0 \sim \text{Beta}(1,1)$, which represents a neutral and non-informative prior on the degree of borrowing (27)

The model in this case can be described with the following equation:

$$\pi(\theta_c, a_0 \mid D_c, D_h) \propto L(\theta_c \mid D_c) [L(\theta_c \mid D_h)]^{a_0} \pi_0(\theta_c) \pi_0(a_0)$$

According to the formula shown above, a_0 is no longer fixed but is learned from the data through the posterior distribution. When the historical data are highly compatible with the current study, the model will choose higher values for a_0 to upweight historical data since we have an agreement and this will lead to stronger borrowing. Conversely, when there is disagreement between the historical and current data, the posterior of a_0 is driven toward smaller values, corresponding to weaker borrowing. This mechanism enables automatic and data-driven calibration of the borrowing strength, allowing the model to adaptively downweighs historical information when incompatibility is present.

When several historical studies are available, the normalized power prior can similarly be extended by assigning either a single prior to a global discount parameter a_0 or separate priors to study-specific discount parameter a_{0j} , depending on whether borrowing strength is assumed to be common across historical studies or allowed to vary by historical study.

4.1.2 Commensurate prior

Commensurate priors (20) extend the Power Prior framework by allowing the degree of borrowing to depend explicitly on the similarity between the current and historical studies, rather than through fixed or estimated discount weights applied to historical likelihoods. Instead of a discount parameter a_0 , borrowing is governed by a commensurability parameter τ , which controls how strongly the current study treatment-effect parameter θ_c is shrunk toward historical evidence

When there is **only one historical study**, inference under the commensurate prior is based on the following posterior distribution

$$\pi(\theta_c, \theta_h, \tau \mid D_c, D_h) \propto L(\theta_c \mid D_c) L(\theta_h \mid D_h) \pi(\theta_c \mid \theta_h, \tau) \pi_0(\theta_h) \pi_0(\tau),$$

where $L(\theta_c \mid D_c)$ and $L(\theta_h \mid D_h)$ denote the likelihoods for the current and historical data, respectively. The conditional prior $\pi(\theta_c \mid \theta_h, \tau)$ induces commensurability between the current and historical effects, while $\pi_0(\theta_h)$ and $\pi_0(\tau)$ represent baseline priors for the historical parameter and the commensurability parameter. Importantly, the historical parameter θ_h must be estimated because similarity is defined through the relationship between θ_c and θ_h ; consequently, both the likelihood and prior for θ_h appear in the above equation.

The parameter τ directly controls borrowing as small values of τ lead to a strong borrowing and the contrary is true. This can be seen from the prior $\pi(\theta_c \mid \theta_h, \tau)$, which is typically specified as a normal distribution $\mathcal{N}(\theta_h, \tau^2)$. The parameter τ controls the spread of this distribution and thus the degree of similarity between θ_c and θ_h . When τ is small, the distribution is tightly concentrated around θ_h , forcing θ_c to be close to θ_h and resulting in

strong borrowing. Conversely, when τ is large, the distribution is more diffuse, allowing greater differences between θ_c and θ_h , which corresponds to weaker borrowing

In other words, we can say that τ represents the standard deviation of the unexplained difference between the current and the historical study that directly controls the amount of borrowing. A weak informative prior (Half-normal(0,1)) is usually specified for τ .

In the case we have multiple historical studies $D_{h1}, \dots, D_{hJ}$, the posterior will be modified like this:

$$\pi(\theta_c, \theta_{h1:J}, \tau \mid D_c, D_{h1:J}) \propto L(\theta_c \mid D_c) \prod_{j=1}^J L(\theta_{h_j} \mid D_{h_j}) \prod_{j=1}^J \pi(\theta_c \mid \theta_{h_j}, \tau) \prod_{j=1}^J \pi_0(\theta_{h_j}) \pi_0(\tau)$$

The current study could be linked to all historical studies through a single **global similarity parameter τ** if we assume a common borrowing level across all the historical studies, or we can use a **study-specific commensurability** approach by assigning a separate parameter τ_{h_j} to each historical study j , allowing different degrees of borrowing from different historical sources. While this offers greater flexibility, it substantially increases the number of parameters that must be specified and estimated.

More generally, as the number of historical studies increases, commensurate priors become increasingly demanding to implement. With J historical studies, one must specify priors for J historical parameters θ_{h_j} and either J commensurability parameters τ_{h_j} (in the study-specific case) or at least one additional hyperparameter τ (in the global case), in addition to the current study parameter θ_c . This growing number of priors increases modelling effort, computational burden, making model construction time-consuming. Similar challenges arise in the Power Prior framework, where each historical study requires its own discount parameter a_{0j} . For this reason, hierarchical alternatives such as the Meta-Analytic Predictive (MAP) prior are often preferred in settings with many historical studies, as they pool evidence through a random-effects structure and capture between-study heterogeneity using fewer hyperparameters.

4.1.3 Meta-analytic predictive prior (MAP)

Meta-Analytic Predictive prior (MAP) (28) is a Bayesian approach designed to combine evidence from **several** historical studies when analyzing a new one. Unlike the Power and Commensurate Priors, which are difficult to apply with several external studies (29), the MAP prior can naturally handle multiple historical studies in a structured and coherent way by performing Bayesian random-effects meta-analysis of the historical studies. Inference for the current study's true parameter θ_c under the MAP framework is based on the following posterior distribution:

$$\pi(\theta_c \mid D_c, D_{h1:J}) \propto L(\theta_c \mid D_c) \pi_{\text{MAP}}(\theta_c \mid D_{h1:J}) \quad (1)$$

where $\pi_{\text{MAP}}(\theta_c \mid D_{h_{1:j}})$ is a prior distribution constructed using the historical datasets $D_{h1}, D_{h2}, \dots, D_{hj}$ through this equation:

$$\pi_{\text{MAP}}(\theta_c \mid D_{h_{1:j}}) = \int \mathcal{N}(\theta_c \mid \mu_h, \sigma_h^2) \pi(\mu_h, \sigma_h \mid D_{h_{1:j}}) d\mu_h d\sigma_h \quad (2)$$

Where $\mathcal{N}(\theta_c \mid \mu_h, \sigma_h^2)$ represents the normal probability density function of θ_c conditional on the mean effect across the historical studies μ_h and the heterogeneity between the historical studies σ_h^2 . (For binary and Poisson data, the model is formulated using a normal approximation on the logit and log scales, respectively).

In other words, we can summarize the steps for the MAP approach as the following:

- 1- We first summarize the historical studies by performing a Bayesian random-effects meta-analysis to obtain a posterior distribution for the hyperparameters (μ_h, σ_h^2)

$$\pi(\mu_h, \sigma_h^2 \mid D_{h_{1:j}}) \propto L(\mu_h, \sigma_h^2 \mid D_{h_{1:j}}) \pi(\mu_h, \sigma_h^2)$$

- 2- We then predict the treatment effect of the current study θ_c by integrating over that posterior distribution of μ_h and σ_h^2 following **equation (2)**
- 3- The resulting predictive distribution is used as a prior for θ_c as in **(1)**

We can notice from the above equations that we have to assign only two (hyper)priors to the population-level parameters μ_h and σ_h regardless of the number of the historical studies. This contrasts with the Power Prior and Commensurate Prior approaches, which, in the presence of multiple historical datasets, require additional study-specific discount or commensurability parameters and corresponding hyperpriors, which makes the model very heavy and time-consuming to execute.

Before applying the MAP approach, it is important to carefully evaluate whether the historical studies are sufficiently comparable to the current study in terms of participant characteristics and study conditions, since these methods rely on an underlying assumption of exchangeability between past and current data. In other words, the model assumes that the current and historical studies are similar, and that any observed differences are random and can be captured through the between-study heterogeneity parameter. Consequently, the MAP approach will generally continue to borrow information from historical data even when some degree of dissimilarity exists. For this reason, robust MAP methods ([next section](#)) are often preferred, as they reduce the risk of inappropriate borrowing in the presence of prior-data conflict

Another key consideration concerns the estimation of the between-study heterogeneity parameter σ_h^2 . When only a small number of historical studies (e.g., two or three) are available, the estimation of σ_h^2 may be unstable, leading to over-borrowing or under-borrowing issues ([28](#)). In other words, MAP assumes that the best prediction for a new study comes from the variability observed in past studies. With too few historical studies, the model can mistakenly conclude that “all studies are very similar”, or “All studies are

different” due to the absence of enough information to reliably estimate heterogeneity σ_h^2 which can lead to either over-borrowing or under-borrowing.

4.1.4 Robust Meta-analytic predictive priors (rMAP)

One of the main limitations of the MAP approach is that it borrows information from the historical studies used in the meta-analysis regardless of whether their treatment effects are similar to the current study effect.

The rMAP prior is an extension of the MAP prior designed to protect the analysis against situations in which the treatment effects in the historical data are not consistent with the effect in the current study (30). The rMAP approach combines two priors: the MAP prior derived from historical data and a vague prior representing the absence of reliable historical information (31), as follows:

$$\pi_{\text{rMAP}}(\theta_c | D_{h_{1:j}}) = w \pi_{\text{MAP}}(\theta_c | D_{h_{1:j}}) + (1 - w) \pi_V(\theta_c)$$

where $\pi_V(\theta_c)$ denotes a vague prior for the current study parameter θ_c . This vague component allows the analysis to revert to a “no borrowing” scenario when the historical data are in conflict with the current data. When $w = 0$, historical data are completely ignored and inference relies solely on the vague prior, however, when $w = 1$, the MAP prior is used alone, corresponding to full reliance on the historical data. For values w falling between 0 and 1, borrowing from historical data is partial and controlled by the weight w . The weight w is typically specified during the design phase and is not estimated from the current data (28) (30). In practice, conservative values such as $w = 0.1$, 0.2 , or 0.3 are commonly used, with the final choice guided by expert judgment and supported by sensitivity analyses evaluating the operating characteristics of the design.

4.1.5 Self-Adaptive Mixture Priors (SAM)

A key limitation of the rMAP prior is the subjective selection of the mixture weight w , which controls the amount of borrowing from historical data. As a result, the degree of borrowing reflects a subjective or design-based level of trust in the historical data and is not full-data-adaptive. To address this, the SAM prior was proposed (23) (32)

The SAM prior is defined as a mixture between an informative prior derived from historical data and a vague prior representing the absence of reliable historical information as follows:

$$\pi(\theta_c | D_c, D_h) \propto L(\theta_c | D_c) \pi_{\text{SAM}}(\theta_c | D_h)$$

$$\pi_{\text{SAM}}(\theta_c | D_h) = w \pi_{\text{inf}}(\theta_c | D_h) + (1 - w) \pi_V(\theta_c)$$

By looking at this equation, we can notice that it looks like the equation used in the rMAP framework, despite this, the calculation of the mixture weight w differs between both approaches.

We know that w is fixed in rMAP, however, it is **fully data-driven in SAM**, and is determined through a likelihood ratio (R) that quantifies the agreement between the current and historical information as follows:

$$R = \frac{L(\theta_c = \hat{\theta}_h \mid D_c)}{\max \{L(\theta_c = \hat{\theta}_h + \Delta \mid D_c), L(\theta_c = \hat{\theta}_h - \Delta \mid D_c)\}}$$

Where Δ denotes a clinically meaningful deviation from the historical value selected based on an expert opinion, $L(\theta_c = \hat{\theta}_h \mid D_c)$ is the likelihood under the agreement between the current and historical data and $L(\theta_c = \hat{\theta}_h + \Delta \mid D_c)$, $L(\theta_c = \hat{\theta}_h - \Delta \mid D_c)$ are the likelihoods under the disagreement.

The likelihood ratio compares the likelihood of the current data under exact agreement with the historical estimate to the likelihood under the most favourable clinically relevant deviation. The mixture weight w is then defined as

$$w = \frac{R}{1 + R}$$

w takes values between 0 and 1, when $w = 0$, this indicates no borrowing from historical and the contrary is true.

Now after I gave a general overview about SAM, I will explain step by step how SAM could be constructed in practice in the case we have a binary end point as it is explained in [\(23\)](#)

****A concrete example ****

Step 1: Defining the vague prior

For a binary endpoint, the vague prior could be specified as a Beta distribution, $\pi_V(\theta_c) = \text{Beta}(a, b)$, typically with $a = b = 1$.

Step 2: Summarizing the historical data and defining the informative prior

The historical studies are summarized using a meta-analytic predictive (MAP) prior, as explained in the previous section, rather than by complete pooling, in order to account for potential heterogeneity across studies.

The informative prior is therefore defined as:

$$\pi_{\text{inf}}(\theta_c \mid D_h) = \pi_{\text{MAP}}(\theta_c \mid D_{h_{1:j}}).$$

Step 3: Defining the likelihood of the current data

the likelihood of the current data D_c , consisting of x responses out of n subjects, is given by the binomial model

$$L(\theta_c \mid D_c) \propto \theta_c^x (1 - \theta_c)^{n-x}.$$

Step 4: updating both the vague prior and the informative prior after seeing D_c

Starting from the general formula of the posterior distribution

$$\pi(\theta_c | D_c, D_h) \propto L(\theta_c | D_c) \pi_{\text{SAM}}(\theta_c | D_h),$$

where

$$\pi_{\text{SAM}}(\theta_c | D_h) = w \pi_{\text{Inf}}(\theta_c | D_h) + (1 - w) \pi_V(\theta_c)$$

we obtain: $\pi(\theta_c | D_c, D_h) \propto w L(\theta_c | D_c) \pi_{\text{MAP}}(\theta_c | D_h) + (1 - w) L(\theta_c | D_c) \pi_V(\theta_c)$

After the incorporation of $L(\theta_c | D_c)$, each term will be updated as follows:

- An **informative component**, obtained by updating the MAP prior with the current data:

$$\pi_{\text{MAP}}(\theta_c | D_c, D_h) \propto L(\theta_c | D_c) \pi_{\text{MAP}}(\theta_c | D_{h_{1:j}})$$

- Vague component:

$$\text{Beta}(a + x, b + n - x)$$

At this stage, w represents the **prior weight** assigned to the informative component.

Step 5: Calculation of data-driven mixture weight

The mixture weight w is determined through a likelihood ratio that quantifies the agreement between the current data and the historical estimate,

$$R = \frac{L(\theta_c = \hat{\theta}_h | D_c)}{\max\{L(\theta_c = \hat{\theta}_h + \Delta | D_c), L(\theta_c = \hat{\theta}_h - \Delta | D_c)\}}$$

$$w = \frac{R}{1 + R}.$$

Step 6: Updating w to w^* and concluding the final posterior distribution

The update from w to w^* arises naturally by applying Bayes' theorem, as the prior mixture weights are reweighted according to the marginal likelihood of each component. Applying Bayes' theorem at the mixture level, the updated weight becomes:

$$w^* = \frac{w z_1}{w z_1 + (1 - w) z_0}.$$

$$z_0 = \frac{B(a + x, b + n - x)}{B(a, b)}$$

$$z_1 = \int L(\theta_c | D_c) \pi_{\text{MAP}}(\theta_c | D_{h_{1:j}}) d\theta_c.$$

Where $B(\cdot, \cdot)$ is a beta function, z_0 and z_1 are the marginal likelihoods of the current data under the vague and the MAP prior respectively.

Substituting these results into the mixture representation yields the final posterior distribution:

$$\pi(\theta_c | D_c, D_h) = w^* \pi_{\text{MAP}}(\theta_c | D_c, D_h) + (1 - w^*) \text{Beta}(a + x, b + n - x).$$

Like I mentioned before, SAM prior is a supervised borrowing method since Δ is selected based on expert opinion, and the likelihood ratio test will differ based on the type of the endpoint under investigation. In conclusion, there is no uniform methodology in the calculation of SAM prior, the methods differ by the type of end points and the type of the distribution selected for the vague prior.

4.1.6 Elastic prior

The Elastic Prior (33) can be formulated within the general Bayesian framework for historical borrowing, in which inference on the current study parameter θ_c is based on the following posterior distribution:

$$\pi(\theta_c | D_c, D_h) \propto L(\theta_c | D_c) \pi_{\text{Elastic}}(\theta_c | D_h) \quad (3)$$

where $\pi_{\text{Elastic}}(\theta_c | D_h)$ represents an adjusted prior derived from the historical data (33). The construction of this elastic prior can be explained in three steps:

Step 1: The historical data are summarized using a fully Bayesian MAP approach. The resulting MAP prior is approximated and its mean $\hat{\mu}_h$ and variance $\hat{\sigma}_h^2$ are computed, providing a moment-based summary of the historical information.

Step 2: The degree of agreement between the historical and current data is quantified using a discrepancy statistic T following this formula

$$T = \frac{(\hat{\theta}_c - \hat{\mu}_h)^2}{\text{Var}(\hat{\theta}_c) + \hat{\sigma}_h^2}$$

where $\hat{\theta}_c$ denotes the observed effect estimate from the current study, $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ are the historical mean and variance from **step 1**. We can see that smaller values of T indicate stronger agreement and larger values indicate disagreement.

This discrepancy measure must then be transformed into a discount function $g(T) \in (0,1)$, typically specified using a logistic form,

$$g(T) = \frac{1}{1 + \exp \{a + b \log (T)\}}$$

with tuning parameters a and b controlling how rapidly borrowing decreases as disagreement increases.

Step 3: the historical posterior obtained in **Step 1** is modified by inflating its variance by a factor $g(T)^{-1}$, yielding the elastic-adjusted prior

$$\pi_{\text{Elastic}}(\theta_c | D_h) \sim \mathcal{N}(\hat{\mu}_h, \hat{\sigma}_h^2 g(T)^{-1})$$

Where $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ denote the mean and variance of the historical posterior, respectively and we obtained them from **step 1**.

Under the assumption that the historical and current studies target the same underlying treatment effect, the posterior distribution obtained for θ_h in **Step 1** is reinterpreted as a prior distribution for the current-study parameter θ_c . When the historical and current data are in close agreement, $g(T) \approx 1$, the variance inflation is minimal and borrowing is strong. Conversely, when substantial disagreement is detected, $g(T) \approx 0$, the variance becomes large and the adjusted prior approaches a vague prior, effectively eliminating borrowing.

The discrepancy statistic T is inversely related to the combined variance of the current and historical estimates, as it represents a standardized measure of the difference between $\hat{\theta}_c$ and $\hat{\mu}_h$. When the variance is large, T becomes small, leading to $g(T) \approx 1$, indicating minimal discounting. Importantly, in such cases the elastic prior remains diffuse due to the large historical variance $\hat{\sigma}_h^2$, rather than the effect of the discount function. This can be seen in this formula: $\pi_{\text{Elastic}}(\theta_c | D_h) \sim \mathcal{N}(\hat{\mu}_h, \hat{\sigma}_h^2 g(T)^{-1})$

The final posterior distribution for θ_c is then obtained by combining this elastic-adjusted prior with the likelihood from the current study, as shown in **equation (3)**.

As I mentioned above, the Elastic prior is a supervised borrowing method, as the form of the discrepancy statistic T and the parameters (a, b) are specified by the analyst. Different formulations of T have been proposed in the literature (33); in this work, I adopted a simple and practical form based on standardized differences between current and historical estimates.

4.2 Joint likelihood-based family

4.2.1 Shrinkage borrowing via standard Bayesian Hierarchical Modeling (BHM) and Normal-Normal Hierarchical Model (NNHM)

To understand this approach, we will begin with the general formulation of the joint likelihood-based families:

$$\pi(\theta_{1:j}, \mu, \sigma^2 \mid D_{1:j}) \propto \prod_{j=1}^J L(\theta_j \mid D_j) \prod_{j=1}^J \pi(\theta_j \mid \mu, \sigma^2) \pi_0(\mu, \sigma^2).$$

where θ_j denotes the true effect in study j , including both the current and historical studies. In this formulation, all studies are modeled jointly through their likelihood contributions, while borrowing of information is induced through the hierarchical prior linking the study-specific effects.

The NNHM (34) is the simplest special case of this general framework, consisting of two hierarchical levels that account for within-study variability and between-study heterogeneity.

First level of the hierarchy:

Each study provides an observed effect estimate y_j , which is assumed to be a noisy measurement of its own true effect θ_j , with known within-study variance σ_j^2 (typically taken as the squared standard errors reported by each study)

$$y_j \mid \theta_j \sim \mathcal{N}(\theta_j, \sigma_j^2)$$

Second level of the hierarchy:

In addition, the true study-specific effects θ_j are allowed to vary across studies. This between-study variation is modeled by assuming that the study effects are centered around a common mean μ , with between-study variance σ^2 :

$$\theta_j \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$$

Here, μ represents the overall average effect across studies, while σ^2 quantifies heterogeneity between studies. Smaller values of σ^2 indicate that studies are more similar and therefore lead to stronger borrowing of information, whereas larger values allow for greater heterogeneity and weaker borrowing. Priors are assigned to μ and σ^2 , typically a weakly informative normal prior for μ and a Half-Normal or Half-Cauchy prior for σ^2 .

Combining the within-study model and the between-study model through Bayes' theorem yields posterior estimates of the study-specific effects that reflect partial pooling across studies. In particular, the posterior mean of each θ_j can be expressed as a weighted average of the observed study estimate and the overall mean:

$$\mathbb{E}(\theta_j \mid y_j, \mu, \sigma^2) = w_j y_j + (1 - w_j) \mu$$

where the weight w_j denotes the study-specific shrinkage weight, defined as a function of the within-study variance σ_j^2 and the between-study heterogeneity σ^2 and it is given by these formulas:

$$w_j = \frac{1/\sigma_j^2}{1/\sigma_j^2 + 1/\sigma^2} = \frac{\sigma^2}{\sigma^2 + \sigma_j^2}$$

These weights make the shrinkage mechanism of the NNHM explicit. When between-study heterogeneity σ^2 is small, study effects are strongly shrunk toward the common mean μ , resulting in strong borrowing across studies. When σ^2 is large, shrinkage is weak and each study's estimate is driven primarily by its own data. Similarly, studies with large within-study variance σ_j^2 (i.e. imprecise estimates) are shrunk more strongly toward the overall mean, while more precise studies retain greater influence.

The NNHM can also be extended to incorporate bias adjustments, leading to the bias-corrected meta-analysis model proposed by Verde (35). In this extension, additional hierarchical components are introduced to explicitly model potential systematic biases in the historical studies before borrowing information for the current trial.

4.3 Practical implementation of the Bayesian borrowing methods for different types of endpoints

The application of simple power priors or commensurate priors becomes less straightforward when multiple historical studies are available. Therefore, in this section, more flexible approaches are considered, including meta-analytic predictive (MAP) priors, mixture priors (rMAP and SAM), elastic priors, and the Normal–Normal Hierarchical Model (NNHM).

For prior-based approaches, the implementation follows a two-stage strategy. In the first stage, the historical studies are analysed to construct an informative prior distribution for the parameter of interest in the current study. This is achieved using a MAP framework, where the study-specific effects from the historical data are assumed to arise from a common distribution characterized by an overall mean and a between-study heterogeneity parameter. The model is specified on an appropriate scale for each endpoint: the logit scale for binary outcomes, the log scale for event rates, and the identity scale for continuous outcomes. **Weakly informative Normal priors are assigned to the overall mean parameter, while Half-Normal priors are specified for the between-study heterogeneity where σ_j .** These priors are selected to cover clinically plausible values while avoiding unrealistically large variability that could lead to unstable borrowing.

In the second stage, the resulting MAP prior is used as the informative prior for the current study analysis. For the binary endpoint, the MAP prior is back-transformed to the probability scale and approximated as a mixture of Beta distributions. This means that the prior is represented as a weighted combination of several Beta distributions rather than a single distribution. Each component reflects a possible range and shape of the response probability informed by the historical data, while the weights represent their relative

plausibility. This mixture representation provides flexibility and allows the prior to capture skewness or heterogeneity in the historical evidence. For the event-rate endpoint, the MAP prior is constructed on the log scale and approximated using a mixture of Normal distributions; posterior summaries are then back-transformed using the exponential function to obtain event rates per person-year. For continuous outcomes, the MAP prior is constructed directly on the original scale and approximated using a mixture of Normal distributions, without the need for back-transformation.

In robust borrowing approaches such as rMAP and SAM, the informative MAP prior is combined with a vague prior component representing minimal prior information. This mixture structure provides protection against prior–data conflict: when the current data are consistent with the historical data, the informative component dominates and borrowing is substantial, whereas in the presence of disagreement, the vague component receives more weight, thereby reducing the influence of historical information.

In contrast, the Normal–Normal Hierarchical Model (NNHM) adopts a joint likelihood-based approach in which current and historical studies are analysed simultaneously within a single hierarchical model. Each study provides an estimate of the effect with an associated within-study variance, and the true study-specific effects are assumed to follow a normal distribution with mean μ and between-study variance σ_j^2 . **Priors are assigned directly to these higher-level parameters: weakly informative Normal priors for the overall mean μ and Half-Normal (or alternatively Half-Cauchy) priors for the heterogeneity parameter σ_j .** For binary and event-rate endpoints, the model is specified on the logit and log scales, respectively, and posterior estimates are back-transformed to the natural scale for interpretation. For continuous outcomes, the model is defined on the original scale and no back-transformation is required. In this framework, borrowing arises naturally through the hierarchical structure, with the degree of information sharing across studies governed by the magnitude of the heterogeneity parameter τ .

4.4 Implementation of every method in R

In this section, I will mention the R packages that could be used in the implementation of every method (**Table1**)

Table1. Different R packages and R functions used in the implementation of every borrowing method:

Method	R Package	CRAN Link	Key Functions	Supported Outcomes / Features
Power Prior	BayesPPD	https://CRAN.R-project.org/package=BayesPPD	two.grp.fixed.a0(), two.grp.random.a0(), power.two.grp.fixed.a0()	Normal, binary, Poisson, exponential; fixed or random power

				parameter; multiple historical datasets
	BayesPPDS urv	https://CRAN.R-project.org/package=BayesPPDSurv	phm.fixed.a0(), phm.random.a0()	Proportional hazards models with piecewise constant hazard
	bayesDP	https://CRAN.R-project.org/package=bayesDP	bdpnormal(), bdpbinomial(), bdpsurvival()	Discount function-based power prior
	hdbayes	https://CRAN.R-project.org/package=hdbayes	glm.pp()	GLMs with Stan backend
Normaliz ed Power Prior	hdbayes	https://CRAN.R-project.org/package=hdbayes	glm.npp(), glm.npp.lognc()	GLMs; log normalizing constant estimation
	BayesPPD	https://CRAN.R-project.org/package=BayesPPD	two.grp.random.a0()	Random a_0 via normalized power prior framework
Commen surate Prior	hdbayes	https://CRAN.R-project.org/package=hdbayes	glm.commensurate()	GLMs
	psborrow2	https://CRAN.R-project.org/package=psborrow2	borrowing_hierarchical_ commensurate()	Simulation framework; propensity score integration
	psBayesbor row	https://CRAN.R-project.org/package=psBayesborrow	commensurate.bin(), commensurate.t2e(), commensurate.cont()	Binary, time-to- event, continuous; propensity score matching
MAP Prior /rMAP	RBesT	https://CRAN.R-project.org/package=RBesT	gMAP(), automixfit(), postmix()	Binary, normal, Poisson; mixture approximation of MAP prior
	hdbayes	https://CRAN.R-project.org/package=hdbayes	glm.bhm()	GLMs; Bayesian hierarchical model inducing MAP prior

SAM Prior	SAMprior	https://CRAN.R-project.org/package=SAMprior	SAM_weight(), SAM_prior(), get_OC()	Binary, continuous.
Elastic Prior	—	—	Custom implementation required	No dedicated CRAN package available
NNHM	hdbayes	https://CRAN.R-project.org/package=hdbayes	glm.bhm()	GLMs; CmdStan backend
	historicalborrow	https://CRAN.R-project.org/package=historicalborrow	hb_mcmc_hierarchical() , hb_mcmc_pool()	Hierarchical and mixture models; benchmark models (independent, pooled)
	psborrow2	https://CRAN.R-project.org/package=psborrow2	borrowing_hierarchical_ commensurate()	Hierarchical commensurate prior (variant of NNHM)

5. Materials and methods

This work was conducted along two complementary directions: an application to simulated data and an application to real-world data.

5.1 Application to simulated data

5.1.1 Overview

Four borrowing methods (rMAP prior, SAM prior, elastic prior, and NNHM) were applied to two realizations of simulated data to compare and discuss the behavior of these methods across different scenarios. The simulation framework was designed to explore method performance across varying degrees of similarity between historical and current studies.

5.1.2 Current Study Data Generation

A single “current study” was simulated and held constant across all four scenarios to enable direct comparison of borrowing methods under identical current data conditions. The current study comprised:

- **Sample size:** $n = 100$ patients in one realization and $n=20$ in a second realization
- **Follow-up:** 200 person-years total or 40 person-years (2 years per patient)

Three types of endpoints were considered for generating the current study (**Table2**)

Table2. Simulation of the three different types of endpoints for the current study

Endpoint	True Parameter	Data Generation Mechanism
Binary (Response Proportion)	$\theta = 0.6$	<i>Responders $\sim$ Binomial (n, θ)</i>
Events per person-years	$\lambda = 0.25 \text{ events/person-years}$	<i>Events $\sim$ poisson ($\lambda, \text{person-years}$)</i>
Continuous (Mean)	$\mu = 25, \sigma = 12.0$	$Y \sim \text{Normal}(\mu, \sigma^2)$

The random seed was fixed (seed = 2024) to ensure reproducibility, yielding identical current study observations across all scenarios.

5.1.3 Historical Studies generation

Ten “historical studies” were simulated with **fixed sample sizes** across all scenarios: $n = \{45, 87, 132, 56, 98, 71, 115, 39, 144, 63\}$. For event per person time outcomes, follow-up time was set to $2 \times \text{sample size (person-years)}$ per historical study.

5.1.4 Final datasets (Four scenarios)

The four datasets differed only in the degree of similarity between historical and current study true parameters for every endpoint following this description:

Scenario 1: High Similarity All 10 historical studies had true parameters within $\pm 3\%$ (proportions), ± 0.02 (event per person year), or ± 2 units (continuous mean) of the current study values.

Scenario 2: Low Similarity

All 10 historical studies have true parameters deviating by 15–25% (proportions), 0.12–0.20 (event per person years), or 12–18 units (continuous mean) from current study values, with direction randomly assigned.

Scenario 3: Medium Similarity All 10 historical studies have true parameters deviating by 8–12% (proportions), 0.05–0.10 (event per person years), or 5–10 units (continuous mean) from current study values.

Scenario 4: Mixed Similarity Historical studies comprise: 3 high similarity, 3 low similarity, and 4 medium similarity studies.

For each historical study, the code first chooses a **“true” value** based on the similarity level. Then it simulates observed data from that true value.

The **standard errors** are calculated separately for each outcome as follows:

For the **response proportion**, $SE = \sqrt{\frac{p(1-p)}{n}}$ where p is the observed response proportion and n is the study sample size.

For the **event per person-years**, $SE = \sqrt{\frac{\text{observed event rate}}{\text{person-years}}}$, where person-years are calculated as: $n \times 2$

For the **continuous outcome**, $SE = \frac{SD}{\sqrt{n}}$, where SD is the observed standard deviation in that simulated historical study. To calculate the SD , each historical study’s mean is first shifted (perturbed) to reflect its level of similarity to the current study, and individual data are then generated from a normal distribution with that mean and a fixed standard deviation of 12.

The code used to simulate both current and historical studies are available in appendix: see [Simulation](#) and the final datasets are available in [Supporting material I](#).

5.1.5 Bayesian Borrowing Methods

Bayesian borrowing methods were implemented to incorporate historical data D_{h_j} , ($j = 1, \dots, 10$) into the estimation of the effect of the current study θ_c (probability event for binary endpoint, event rate for count data, and mean for continuous data). All analyses were conducted separately for binary, event-rate, and continuous endpoints (**Table 3**).

All the Bayesian borrowing methods were applied according to the description mentioned in the section: [Literature Review of the statistical methods used for borrowing](#).

Table 3. Model specification and prior choices in the prior-based borrowing methods.

		rMAP	SAM	Elastic
Summary of historical data	Prior choice for the mean	For all endpoints, a weakly informative Normal prior was specified for the overall mean parameter, $\theta_h \sim \mathcal{N}(0, 2^2)$, defined on the appropriate model scale. For the binary endpoint, the model was formulated on the logit scale, and the resulting MAP prior was then transformed and approximated as a mixture of Beta distributions on the probability scale. Similarly, for the event-rate endpoint, the model was specified on the log scale.		
	Prior choice for the between-study heterogeneity	The between-study heterogeneity was assigned endpoint-specific Half-Normal priors on the standard deviation, with scale parameters of 1 for binary outcomes, 0.5 for event-rate outcomes (on the log scale), and 10 for continuous outcomes.		
	Use of historical information	The historical studies were summarized using a Bayesian hierarchical meta-analytic predictive (MAP) framework implemented in the RBest package in R. In the mixture-based approaches (rMAP and SAM), the full MAP prior distribution was retained and used directly (as a mixture), whereas in the Elastic approach the MAP prior was summarized using its mean and variance, and replaced by a single normal distribution.		

Vague prior for the mixture	The vague priors specified for the mixture priors were specific for every endpoint as follows: Beta(1,1) for the binary endpoint Normal(log(0.25), 4) for the event per person time at the log scale Normal(25, 400) for the Normal endpoint		Not used explicitly (borrowing controlled via variance inflation)
Borrowing mechanism	Fixed mixture weight	Data-adaptive mixture weight	Variance inflation via discount function $g(T) = \frac{1}{1 + \exp\{a + b \log(T)\}},$ where $T = \frac{(\hat{\theta}_c - \hat{\mu}_h)^2}{\text{Var}(\hat{\theta}_c) + \hat{\sigma}_h^2}$
Weight / tuning parameters (these values are chosen arbitrary in this realization of simulated data)	0.3 for all the endpoints	Data-driven w (via SAMprior, using threshold $\Delta = 0.10$ for the binary endpoint, $\Delta = 0.30$ for the event rate, and $\Delta = 3$ for the continuous endpoint)	Fixed $a = 0$, $b = 2$ for all the endpoints

Normal–Normal Hierarchical Model (NNHM)

In the joint likelihood-based framework, current and historical studies were analyzed simultaneously. For each study $j = 1, \dots, J$, observed data D_j provided an estimate y_j with known within-study variance σ_j^2 , linked to a true study-specific effect θ_j . These effects were assumed to arise from a common population distribution with mean μ and between-study variance σ^2 .

A prior distribution must be specified for the second level of the hierarchy, namely for the overall mean μ and the between-study heterogeneity (σ^2). In our analyses, these choices were made separately for each endpoint. Specifically, for the **response proportion (logit scale)** with $\mu \sim \text{Normal}(0, 3^2)$ and $\sigma \sim \text{Half-Normal}(0, 1)$; for the **event per person time (log**

scale with $\mu \sim \text{Normal}(\log(0.25), 2^2)$ and $\sigma \sim \text{Half-Normal}(0, 1)$; and for the **continuous mean**, $\mu \sim \text{Normal}(25, 20^2)$ and $\sigma \sim \text{Half-Normal}(0, 10)$. These priors are weakly informative but reflect plausible ranges for each outcome on the transformed scale. The choice of prior for σ governs the extent of allowed heterogeneity and therefore directly influences the degree of borrowing across studies, while the prior for μ represents prior belief about the overall average effect.

The posterior inference was performed using Gibbs sampling, with a Metropolis–Hastings step for updating the between-study heterogeneity parameter, implemented in R. The code is available in the appendix: [Bayesian simulated data](#).

5.2 Real-World data (RWD)

Our current study of interest (22) is a multicenter single-arm trial evaluating the safety and effectiveness of Sirolimus for slow-flow vascular malformation in adults and children. The total sample size of the current study is 132, and this is a special case of a clinical trial of a rare disease as the sample size is quite high.

There is a wide range of age groups among the participants with a female dominance. The participants suffer from different types of vascular malformations existing in several anatomical locations. Some of the participants undergone surgical procedures prior to Sirolimus administration, and some were not. Some of the patients were symptomatic and some were not. The second problem is the absence of a control group due to the ethical concerns of recruiting patients without giving them the treatment.

Several clinical endpoints were assessed in the current study; however, the primary endpoint (binary endpoint) considered in our analysis is **the improvement in pain at 12 months. This outcome was reported in 76% of patients (77/101)**, as only 101 participants have reported they suffered from pain at baseline.

The secondary endpoint (continuous endpoint) evaluated in this analysis is **the mean reduction in D-dimer levels from baseline after 12 months of Sirolimus treatment (n = 64)** (calculated as the pre-Sirolimus levels – post-Sirolimus levels / pre-Sirolimus levels)

Our goal now is to search for some historical studies and extract the data that best match our study of interest to reinforce its our estimates.

5.2.1 Identification of the relevant studies

An advanced search was conducted in PubMed with the following search term: (sirolimus OR rapamycin OR mTOR) AND ("vascular anomalies" OR "vascular malformations" OR "vascular abnormalities" OR "vascular tumors" OR "lymphatic malformations" OR "lymphatic anomalies" OR "venous malformations" OR "capillary malformations" OR "arteriovenous malformations" OR "combined vascular malformations").

The publication year and study design were not restricted so that numerous observational studies including case series as well as single-arm trials were included in the search. The

primary aim was to select studies with both pre- and post-sirolimus data similar to my current study of interest.

606 studies were captured in the initial search; after applying filters to select only those that included humans, there were 463 articles. Titles and abstracts screening narrowed the selection down to 31 studies, and 18 of them were selected as relevant for inclusion **(Figure 1)**.

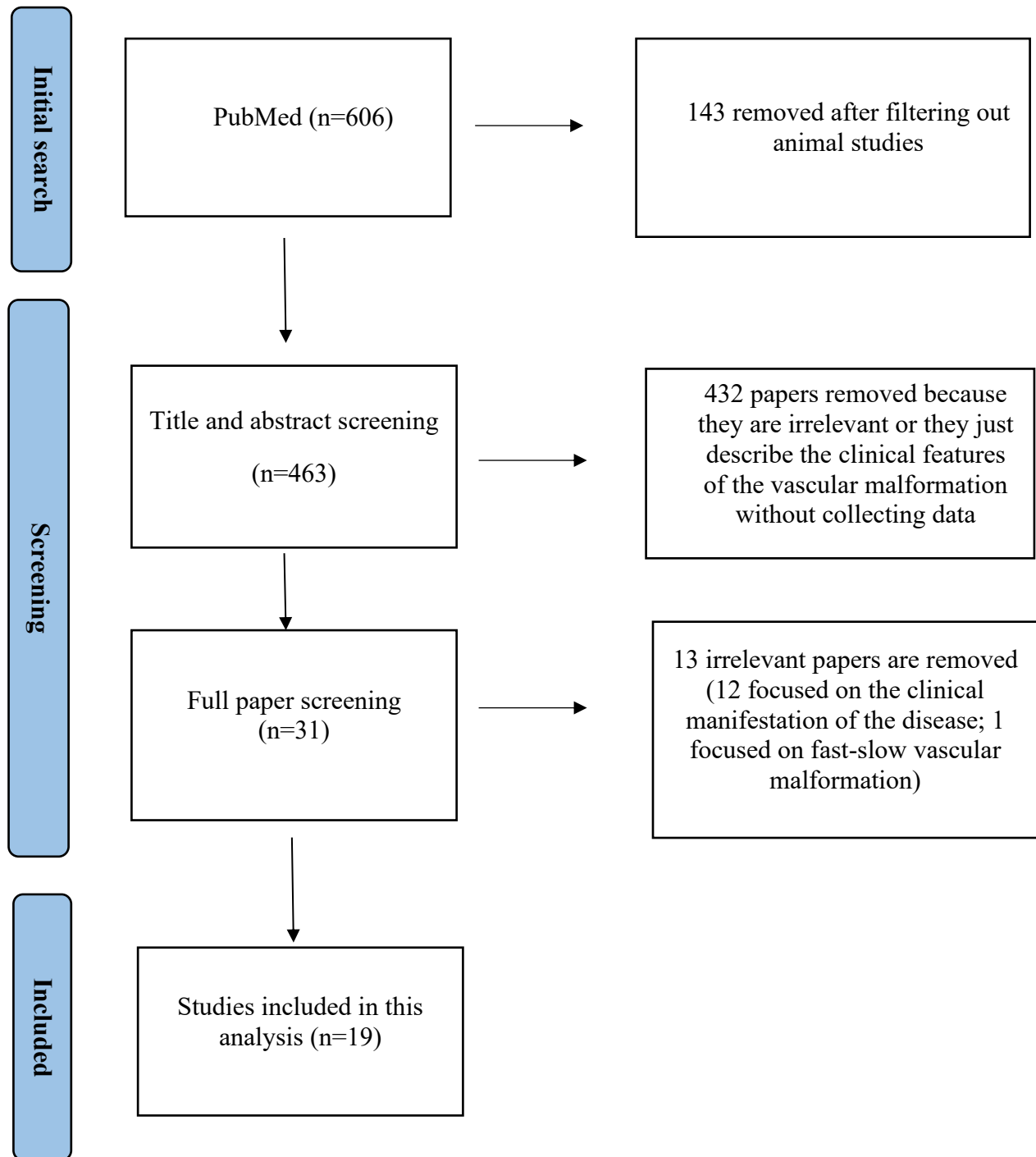


Figure 1. PRISMA diagram to describe the final inclusion of 19 studies in the analysis

5.2.2 Data extraction

Among the 19 studies and our current study, we extracted the total number of participants, median age, male proportion, the dose of sirolimus, procedures and medications prior to

Sirolimus administration, type and anatomical location of the vascular malformation (all of these demographic and clinical features were used later on to calculate similarity weights) **(Table 3)** and the entire data extraction table is available in [Supporting material II](#).

Regarding responses, data are collected regarding the change in pain, response type (complete versus partial response and no response), and the change in the D-dimer levels are collected.

Table 3. Baseline characters of the studies included in the analysis

Studies	Study design	Baseline characteristics				Type of vascular malformation			
		N	Age median in years (range)	Males (n)	Median duration of Sirolimus treatment in months (range)	Venous (n)	Capillary (n)	Lymphatic (n)	Combined (n)
Seront et al, 2023 (our study of interest) (22)	Trial	132	30 (0-73)	46	7.5 (3-12)	76	7	27	22
Fernández Oliveira et al, 2023 (36)	Observational retrospective	7	4.2 (3 months-17 years)	6	12.5 (1-96)	2		3	2
Mack et al, 2019 (37)	Observational retrospective	15	13 (1-25)	5	21 (9-42)	3			12
Tian et al, 2019 (38)	Case series	56	2 (2 months-16 years)	32	12 (1-24)			56	
Harbers et al, 2021 (39)	Case series	12	16 (1-50 years)	8	6 initial (3-6), follow-up (7-36)	2		5	5
R Gabeff et al, 2019 (40)	Observational retrospective	10	13 (2-47)	6	24.5 (3-72)	10			
Engel et al, 2023 (41)	Prospective and retrospective	29	8 (0.33–42.36)	9	7.5 (3-12)				29
Rebecca et al, 2025 (42)	Observational retrospective	55	41 (23-72)	26	4.2 IQR: 424.5 days	35	2	4	14
Hammer, 2018 (43)	Trial	19	16 (3-64)	7	28 (1-48)	7		8	4

Ozeki et al, 2024 (44)	Trial	13	6 (8 months-71)	6	13	1		4	8
Adams et al, 2016 (45)	Trial	60	8.1 (21 d-28.5 y)	25	6 (3-12)			22	38
Yi Ji e al, 2021 (46)	Trial	126	4.8 (0.1-14)	64	6 years in total, assessment after 6, 12, 24 months	19		20	76
Pang et al, 2021 (47)	Prospective Case series	7	43 (23-60)	3	6 in total	5		2	
Yesil et al, 2015 (48)	Case series	6	8.667 (3.66- 13)	3	13 (9-18)	1			5
Kim et al, 2023 (49)	Case series	20	5.05 (1 month-19 y)	14	25.2 (7–51.6)	4		7	9
Wang et al, 2025 (50)	Retrospective case control	36	24	15	6.34			36	
Triana et al, 2017 (51)	Retrospective review	41	12.8 (0.16–47)	17	8.5 (1-51)	8		19	14
Sandbank et al, 2019 (52)	Multicenter study and a review	19	mean: 6.2 years, range: 0.5–322 months	13	11 (1.5-60)	1		6	12
Weigand et al, 2018 (53)	Systematic review	55	5 (0-64)	18	10 (1.5-78)			28	27
Weigand et al, 2022 (54)	Systematic review	89	1 (0-17)	16	12 (2-48)			89	

5.2.3 Quality-adjusted Similarity weights

A pre-borrowing framework is introduced to quantify the degree of **similarity** between each historical study and the current study. This step is motivated by the presence of both demographical and clinical heterogeneity of the participants of historical studies, as patients differ in age (paediatric vs adult populations), type of vascular malformation (venous, lymphatic, capillary, combined), and prior treatments and procedures (sclerotherapy, surgery, laser, steroids) and this is shown in the main data extraction table available in [Supporting material II](#). As a result, assuming full similarity across studies would be unrealistic, and study-specific weights are required to reflect both comparability and methodological reliability.

To address this, a **study-specific similarity weight** (q_j) is constructed by combining three components: (1) methodological quality assessed through risk of bias, (2) study design, and (3) demographic and clinical similarity.

First, **risk-of-bias (ROB) weights** (w_j^{ROB}) are assigned based on the NIH quality assessment tool for before–after studies without a control group (55). Each study is evaluated using a standardized set of 12 criteria and classified into one of three categories: Good, Fair, or Poor. These categories are mapped to numerical weights (1.0, 0.7, and 0.3, respectively), reflecting decreasing internal validity. This approach ensures that studies with higher risk of bias contribute less to the borrowing process (**Table 4**).

The detailed ROB assessment for all studies is provided in [Supporting material III](#).

Second, **study design weights** (w_j^{Design}) are derived based on the hierarchy of evidence as described in the GRADE framework (56). Although GRADE does not provide explicit numerical weights, it emphasizes that randomized and prospective designs provide stronger evidence than observational studies and case series. Based on these principles, study designs are assigned weights reflecting their relative methodological strength, with higher values assigned to prospective and multicentre studies and lower values to less robust designs like case series (**Table 4**).

Third, **The demographic and clinical characteristics** are combined into a single similarity measure (S_j) through the use of the Gower distance (57), which aggregates information across multiple variables of different types. For each historical study j , the distance D_j from the current study is calculated as the average of variable-specific distances:

$$D_j = \frac{1}{K} \sum_{k=1}^K d_{jk} \quad \text{where:}$$

- K is the number of variables included in the similarity assessment
- $d_{jk} \in [0,1]$ represents the distance between study j and the current study for variable k

Each variable contributes equally to the overall distance, after being scaled to ensure comparability. For continuous variables (e.g., age or disease duration), the distance is computed as the absolute difference between the two studies divided by the observed range of that variable. For categorical variables (e.g., anatomical location), the distance is defined as 0 if the categories are identical and 1 if they differ. Proportions (e.g., distribution of malformation types) are treated as continuous variables and compared using absolute differences.

The overall distance D_j is therefore the average of these standardized differences across all variables, ensuring that no single variable dominates the measure. This distance is then transformed into a similarity measure: $S_j = 1 - D_j$, so that higher values of S_j indicate greater similarity between the historical study and the current study (**Table 4**).

The variables included in this calculation are age, gender, disease duration, proportion of malformation type, anatomical location, and procedures undergone before Sirolimus administration. All of these variables are available in the main data extraction sheet in the section: [Supporting material II](#)

Only the available demographic and clinical data were used, relying exclusively on the information reported in each study rather than imputing missing values. In the similarity calculation, variables that were not reported for a given study were simply excluded, so that similarity was based only on characteristics shared with the current study. To account for differences in the amount of available information across studies, a completeness adjustment was applied**, assigning slightly lower weights to studies with more missing data. This approach avoids introducing unsupported assumptions about unreported clinical features and ensures that the similarity assessment remains transparent and grounded in observed data.

The three components are combined into a single **similarity weight** q_j , defined as:

$$q_j = S_j \times \sqrt{w_j^{ROB} \times w_j^{Design}}$$

This formulation prioritizes demographic and clinical similarity as the primary driver of similarity, while incorporating methodological quality as a moderating factor. The use of the square root (i.e., a geometric mean-type adjustment (58)) ensures that lower-quality studies are down-weighted without being completely excluded.

The resulting weights q_j are used in the Bayesian borrowing analyses to modulate the contribution of each historical study, allowing borrowing to reflect both similarity to the target population and the credibility of the evidence.

**The completeness metric is calculated by dividing the number of variables observed in both the current study and historical study j over the total number of variable used in the similarity assessment and then this metric is multiplied by S_j to downweigh the historical study with large number of missing data

Table 4. The individual component scores and the resulting final pre-borrowing weights.

Study	ROB	Study Design	Demographic and clinical similarity weight (S_j)	ROB Weight (w_j^{ROB})	Design Weight (w_j^{Design})	Final weight (q_j)
Fernandez Oliveira et al, 2023	Poor	Observational retrospective	0.367	0.3	0.6	0.156
Mack et al, 2019	Fair	Retrospective review	0.211	0.7	0.5	0.125
Tian et al, 2019	Fair	Case series	0.256	0.7	0.5	0.152
Harbers et al, 2021	Poor	Case series	0.287	0.3	0.5	0.111
R Gabeff et al, 2019	Poor	Observational retrospective	0.313	0.3	0.6	0.133
Engel et al, 2022	Fair	Observational Prospective and retrospective	0.207	0.7	0.6	0.134
Rebecca et al, 2025	Fair	Observational retrospective	0.345	0.7	0.6	0.224
Hammer, 2018	Good	Trial	0.374	1	0.85	0.345
Ozeki et al, 2024	Good	Trial	0.145	1	0.85	0.133
Adams et al, 2016	Good	Trial	0.138	1	0.85	0.127
Yi Ji e al, 2021	Good	Trial	0.128	1	0.85	0.118
Pang et al, 2021	Fair	Prospective Case series	0.301	0.7	0.85	0.232
Yesil et al, 2015	Poor	Case series	0.341	0.3	0.5	0.132
Kim et al, 2023	Fair	Case series	0.315	0.7	0.5	0.186
Wang et al, 2025	Good	Retrospective case control	0.337	1	0.7	0.282
Triana et al, 2017	Poor	Retrospective review	0.364	0.3	0.5	0.141
Sandbank et al, 2019	Poor	Multicentre study and a review	0.193	0.3	0.75	0.092

Weigand et al, 2018	Fair	Systematic review	0.27	0.7	0.8	0.202
Weigand et al, 2022	Fair	Systematic review	0.248	0.7	0.8	0.186

The R code used to calculate these weights is available in the appendix: [pre borrowing weights](#) and the final excel sheet with all the detailed weights (with the completeness metric and the downweighting of the studies with several missing data) is available in [Supporting material IV](#)

5.2.4 Bayesian borrowing analysis and the inclusion of the pre-borrowing weights

Binary Endpoint (Pain response)

For the binary endpoint, pre-borrowing weights (q_j) were incorporated on the probability scale by converting each historical study into quality-adjusted effective counts:

$$n_j^{\text{eff}} = N_j \times q_j$$

$$r_j^{\text{eff}} = R_j \times q_j$$

where N_j denotes the original sample size of study j , R_j the number of responders, and q_j the corresponding similarity-quality weight. The quantities n_j^{eff} and r_j^{eff} represent the effective sample size and effective number of responses after incorporating similarity weights.

A MAP prior was then fitted using the gMAP function of the R package RBest. This quality-adjusted MAP prior served as the informative component for constructing both the robust MAP (rMAP) and the SAM prior. Additionally, elastic priors were derived based on the mean and standard deviation estimates of the MAP distribution.

For the Normal-Normal Hierarchical Model (NNHM), the pre-borrowing weights were incorporated at the second hierarchical level rather than through effective counts, according to:

$$\theta_j \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2/q_j).$$

Continuous Endpoint (d-dimer decrease)

For the continuous endpoint, similarity weights were incorporated via a quality-adjusted effective sample size. For each historical study j , the effective sample size was defined as:

$$n_j^{\text{eff}} = N_j \times q_j$$

The within-study standard error was then recalculated as:

$$SE_j^{\text{eff}} = \frac{SD_j}{\sqrt{n_j^{\text{eff}}}},$$

where SD_j is the reported standard deviation of study j .

A Gaussian MAP prior was fitted using the *RBesT* package in R. This MAP prior was subsequently used as the informative component in both the rMAP and SAM approaches. Elastic priors were constructed using the mean and standard deviation of the MAP distribution.

For the NNHM, similarity weights were incorporated at the second hierarchical level in the same manner as for the binary endpoint.

5.2.5 Sensitivity analysis

The main objective of the sensitivity analysis is to assess the stability and robustness of each borrowing method across a wide range of parameter values. Ideally, the level of borrowing should remain relatively stable when these parameters are varied. Large fluctuations in results indicate that a method is highly sensitive to specific parameter choices and may therefore lack robustness. To evaluate this, the analysis was repeated across 50 different parameter settings for each method, with parameter grids deliberately constructed to span a broad spectrum of scenarios, ranging from low similarity (limited borrowing) to high similarity (strong borrowing).

Binary endpoint (pain response)

For the binary outcome, the sensitivity grids are defined on both the **probability and logit scales**, reflecting the underlying formulation of each method. The rMAP robustness weight $w \in [0.02, 1.00]$ controls the mixture between informative (historical) and vague priors on the probability scale; exploring this full range allows evaluation from minimal to complete borrowing. The SAM similarity tolerance $\Delta \in [0.01, 0.50]$ is specified on the probability scale, as agreement is assessed in terms of response proportions; this range captures scenarios from very strict similarity to moderate clinical differences. The Elastic prior parameters (a, b) are varied over a 50-point grid with $a \in [-2, 3]$ and $b \in [0.5, 4]$, operating on the logit scale through a discrepancy-based discounting function; this grid spans weak to strong levels of discounting. Finally, the NNHM heterogeneity prior scale $\sigma \in [0.1, 5.0]$ is defined on the logit scale, representing between-study variability in log-odds; the chosen range covers near-homogeneity to substantial heterogeneity.

Continuous endpoint (D-dimer decrease)

For the continuous outcome, the sensitivity grids are defined primarily on the **mean difference scale** for all the methods. The rMAP weight $w \in [0.02, 1.00]$ again governs the degree of borrowing through the prior mixture. The SAM tolerance $\Delta \in [0.1, 3.0]$ is defined on the mean difference scale, reflecting the use of a normal likelihood to assess agreement between current and historical data; this range spans small to clinically meaningful

differences in D-dimer reduction. The Elastic prior grid (a, b) , with $a \in [-2, 3]$ and $b \in [0.5, 4]$, to enable exploration from aggressive borrowing to low borrowing. The NNHM heterogeneity prior $\sigma \in [0.1, 3.0]$ is defined on the mean scale, representing between-study standard deviation and covering scenarios from high exchangeability to substantial heterogeneity.

5.2.6 Point estimates and credible intervals calculation

Binary endpoint (pain response)

Regarding the effects of the historical and current data, 95% confidence intervals were calculated using standard methods for binomial proportions, implemented via the 'prop.test()' function in R.

For the rMAP and the SAM priors, the posterior distributions were represented as mixtures derived from the MAP prior combined with the binomial likelihood. Point estimates were obtained as the posterior mean of the mixture distribution. Credible intervals were computed using the exact posterior mixture quantiles.

For the Elastic prior, the model was constructed on the logit scale. The MAP prior was sampled on the probability scale and transformed to the logit scale, from which a normal approximation was obtained via moment matching. After applying the elastic variance inflation, the posterior distribution on the logit scale was normal. Posterior samples were then generated from this normal distribution and transformed back to the probability scale. Point estimates were calculated as the mean of the transformed posterior samples, and credible intervals were computed as the empirical 2.5th and 97.5th percentiles of these samples.

For the NNHM, posterior inference was obtained via MCMC sampling using a normal likelihood for the logit-transformed study-level proportions and their corresponding standard errors, representing a normal approximation to the binomial sampling distribution. Posterior samples of the current-study parameter were obtained after burn-in and transformed back to the probability scale using the inverse-logit function. Point estimates were calculated as the mean of the transformed posterior samples, and 95% credible intervals were computed as the empirical 2.5th and 97.5th percentiles of these samples.

Continuous endpoint (D-dimer decrease)

Regarding the effects of the historical and current data for the continuous endpoint, 95% confidence intervals were calculated in the conventional way using standard t-based methods for the mean, based on the sample mean, standard deviation, and sample size.

For the rMAP and the SAM priors, the posterior distributions were represented as mixtures of normal components obtained via the RBesT framework. Point estimates were obtained as the posterior mean of the mixture distribution. In these cases, 95% credible intervals were computed using the exact quantiles of the posterior mixture distributions.

For the Elastic prior, the MAP prior was first approximated by a moment-matched normal distribution with mean $\hat{\mu}_h$ and variance $\hat{\sigma}_h^2$. After applying the elastic variance inflation factor $g(T)^{-1}$, the posterior distribution remained normal due to conjugacy with the normal likelihood. Point estimates were obtained as the posterior mean of the normal distribution, and 95% credible intervals were calculated analytically using the corresponding normal quantiles.

For the NNHM, posterior inference was obtained via Markov Chain Monte Carlo (MCMC) sampling using a custom Gibbs sampler with a Metropolis–Hastings update for the between-study heterogeneity parameter. The posterior distribution of the current-study parameter was summarized using retained samples after burn-in. Point estimates were computed as the posterior mean of these samples, and 95% credible intervals were obtained as the empirical 2.5th and 97.5th percentiles.

5.2.7 Effective sample size (ESS)

The concept of effective sample size (ESS) is used to quantify the amount of information in a statistical model by expressing it as an equivalent number of observations. In the Bayesian literature, this is often defined for a prior distribution alone, before observing the current data. For example, in a binary-beta conjugate setting, a Beta (a, b) prior can be interpreted as contributing **a prior successes** and **b prior failures**, so it behaves as if it represents $a + b$ observations in total. This is because, when combined with current data, the posterior becomes Beta ($a + x, b + n - x$) where x is the observed success in the current study, meaning the prior adds directly to the observed counts. Therefore, the ESS of a Beta prior is $a + b$, as it reflects the strength of the prior in terms of an equivalent sample size (59).

However, in this analysis, the borrowing methods used (rMAP, SAM, Elastic priors, and NNHM) combine historical and current data in a data-adaptive way, so the contribution of the historical information is not fixed. Instead, it depends on the level of agreement between historical and current evidence. As a result, a single fixed prior ESS cannot be defined for these methods (59). In other words, there is no universal closed-form expression that can be applied for all the Bayesian borrowing methods and for different types of endpoints.

For this reason, ESS is defined here at the **posterior level**, to measure how much additional information is gained from borrowing after combining the prior with the current data. This is done by comparing the posterior variance with and without borrowing. Since statistical information is inversely related to variance (via Fisher information), a reduction in variance corresponds to an increase in information.

The additional ESS is defined as:

$$\text{ESS} = \max \left(0, n_c \left(\frac{\text{Var}_{\text{no borrow}}}{\text{Var}_{\text{borrow}}} - 1 \right) \right)$$

where n_c is the current study sample size, $\text{Var}_{\text{no borrow}}$ is the posterior variance obtained using only current data, and $\text{Var}_{\text{borrow}}$ is the posterior variance under the borrowing model. This definition expresses the gain in precision due to borrowing as an equivalent number of additional observations from the current study.

6. Results

6.1 Application to simulated datasets

First case: Current study (n= 100)

When the historical studies are very similar to the current study, most Bayesian borrowing methods use a large portion of the historical information. In simple terms, the models consider the historical data reliable and borrow strongly from it. As a result, the estimates become more precise and the intervals are narrower compared to using the current study alone. This also leads to a higher effective sample size (ESS), because the historical data is consistent with the current evidence and supports the same effect (**Figure 2A**).

In the moderate similarity scenario, the historical data is somewhat aligned with the current study but not perfectly. The methods still borrow information, but more cautiously. This means the final estimate is slightly influenced by the historical data. The intervals are slightly narrower than the current-only analysis, and the ESS increases moderately, reflecting balanced and careful borrowing (**Figure 2B**).

When there are clear differences between the historical studies and the current study, the borrowing methods automatically reduce how much they rely on the historical data. In simple words, the models detect disagreement and shift more weight toward the current study. As a result, the ESS decreases and the estimates are driven mainly by the current data. The intervals remain close to the uncertainty of the current study, showing cautious borrowing under heterogeneity (**Figure 2C**).

When the historical data is clearly inconsistent with the current study, the dynamic borrowing approaches strongly down-weight the historical evidence, sometimes almost ignoring it. In practice, this means the final estimate is mainly based on the current trial, and the ESS from historical data becomes very small. This highlights a key advantage of Bayesian dynamic borrowing in rare disease trials: it protects the analysis from misleading conclusions when the historical studies are not compatible with the current population (**Figure 2D**).

N-current=100

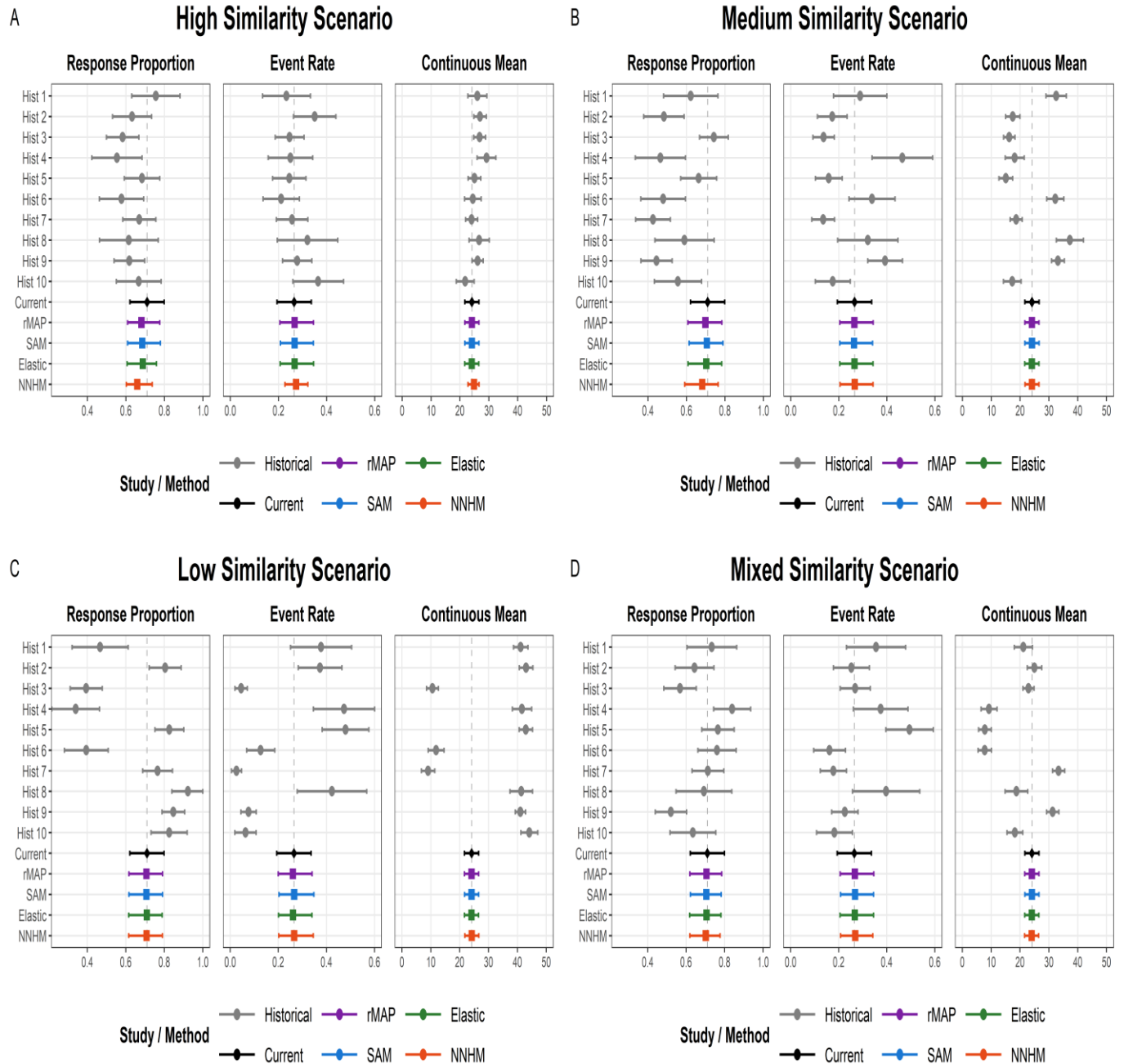


Figure 2. Forest plots showing point estimates and 95% CI based on 4 Bayesian borrowing methods across different scenarios and different endpoints (when n-current=100)

Based on this limited simulation-based investigation, the borrowing methods show a **consistent pattern of behavior**. When similarity is high, all methods borrow more and increase precision; when similarity decreases, they reduce borrowing and rely more on the

current study. This consistency across endpoints indicates that the dynamic borrowing mechanism is stable and can be applied to different outcome type, which is particularly important in rare disease trials where multiple types of outcomes are often analyzed.

Now, let's concentrate our attention to only one endpoint type (the proportion) and compare between the different methods in more details. The simulation results for the binary endpoint indicate that all dynamic borrowing methods behave adaptively across similarity scenarios when compared with the no-borrowing baseline (CI width = 0.177).

In the high similarity scenario, notable precision gains are observed. NNHM achieves the narrowest credible interval (CI width=0.135), followed by Elastic (CI width=0.153), indicating substantial borrowing from historical data. In contrast, rMAP (CI width=0.167) and SAM (CI width=0.170) remain more conservative, with interval widths closer to the no-borrowing scenario

Under low and medium similarity, CI widths across all methods (approximately 0.174–0.177) closely match the no-borrowing level, demonstrating appropriate weakness of borrowing in the presence of heterogeneity.

In the mixed similarity scenario, moderate borrowing is evident. NNHM (CI width=0.157) and Elastic (CI width=0.163) again yield narrower intervals, reflecting partial pooling across compatible studies, whereas rMAP (CI width=0.166) and SAM (CI width=0.160) show more cautious borrowing (**Table 5**)

Table 5: Precision and Borrowing Performance of Bayesian Methods for the Binary Endpoint Across Similarity Scenarios

Method	Scenario	Posterior Mean	CI Lower	CI Upper	CI width
rMAP	High Similarity	0.681	0.609	0.776	0.167
rMAP	Low Similarity	0.707	0.616	0.791	0.175
rMAP	Medium Similarity	0.698	0.607	0.784	0.177
rMAP	Mixed Similarity	0.706	0.619	0.785	0.166
SAM	High Similarity	0.685	0.609	0.779	0.170
SAM	Low Similarity	0.707	0.616	0.791	0.175
SAM	Medium Similarity	0.706	0.614	0.79	0.176
SAM	Mixed Similarity	0.705	0.622	0.782	0.160
Elastic	High Similarity	0.688	0.606	0.759	0.153
Elastic	Low Similarity	0.709	0.615	0.789	0.174
Elastic	Medium Similarity	0.703	0.608	0.783	0.175
Elastic	Mixed Similarity	0.706	0.618	0.781	0.163
NNHM	High Similarity	0.659	0.602	0.737	0.135
NNHM	Low Similarity	0.708	0.615	0.79	0.175
NNHM	Medium Similarity	0.682	0.591	0.765	0.174
NNHM	Mixed Similarity	0.702	0.619	0.776	0.157

No borrowing	0.706	0.614	0.790	0.177
--------------	-------	-------	-------	-------

Overall, the ranking of borrowing strength based on interval width is approximately: the NNHM (strongest borrowing) followed by the Elastic prior (moderate borrowing) followed by the rMAP and SAM (most cautious). These findings seem to indicate that hierarchical pooling methods enable greater information sharing when appropriate, while robust mixture approaches maintain stronger protection against potential incompatibility with historical data (**Figure 3**)

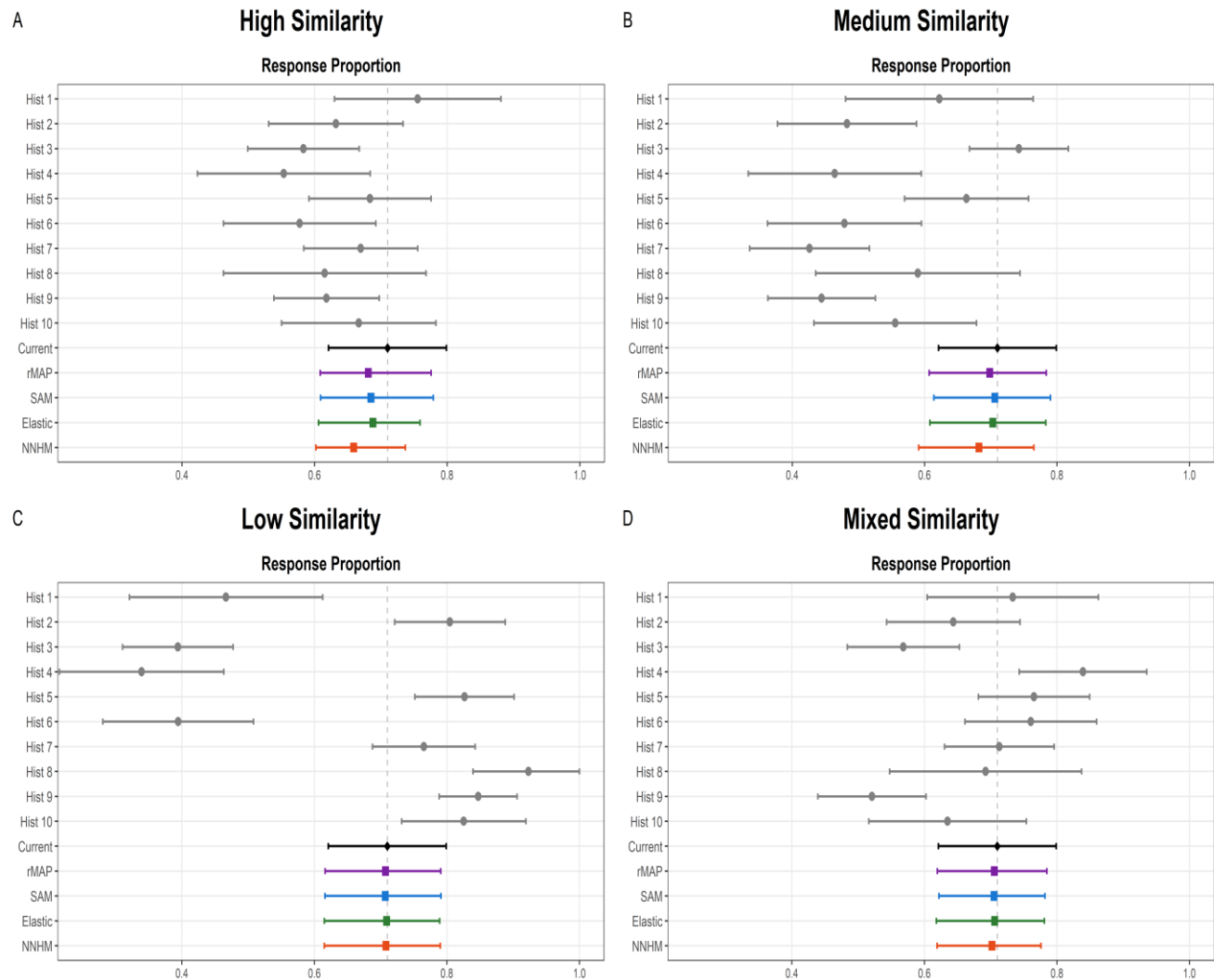


Figure 3. Forest plots showing different borrowing methods across different similarity scenarios for a proportion endpoint

The excel sheets containing all posterior mean estimates and the 95% crl for all the scenarios when the sample size of the current study =100 is available in [Supporting material V](#)

Second case: Current study (n= 20)

For a current study with a small sample size ($n = 20$), representing a typical rare disease setting, **Figure 4** shows that the degree of borrowing decreases as the similarity between the historical data and the current study declines. When the historical studies are highly similar, substantial borrowing occurs, reflected by a shift of the estimates toward the historical data and an increase in precision across the different endpoints. As similarity decreases (medium and low scenarios), the methods progressively downweigh the historical information, resulting in estimates that move closer to the current study results. In the mixed similarity scenario, partial borrowing is observed, while still maintaining robustness against potentially conflicting historical data.

Figure 4 also highlights that differences between the Bayesian borrowing methods are more pronounced in this small-sample setting, particularly in the medium and low similarity scenarios. In contrast, for a current study with larger sample size ($n = 100$), the posterior estimates were closer to the current study estimate, and differences between methods become minimal. The reduction in credible interval width is also less pronounced, indicating a weaker influence of the historical data.

The excel sheets containing all posterior mean estimates and the 95% crl for all the scenarios (when the sample size of the current study =20) is available in [Supporting material VI](#)

N-current=20

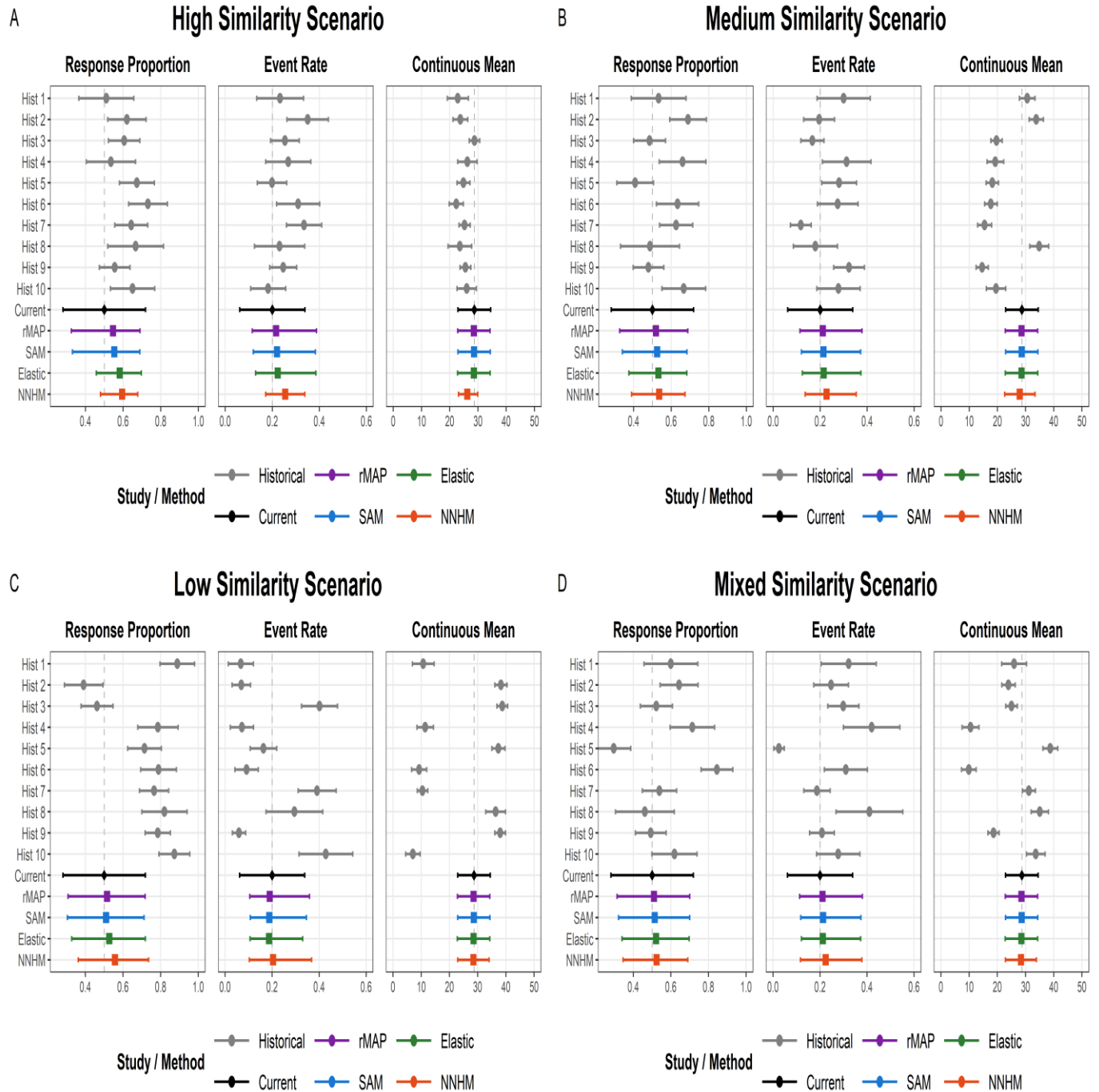


Figure 4. Forest plots showing point estimates and 95% CI based on 4 Bayesian borrowing methods across different scenarios and different endpoints (when n-current=20)

6.2 Real-World data

In this section, the primary aim is to strengthen the analysis of both the pain proportion (a binary endpoint) and the reduction in D-dimer levels (a continuous endpoint) by incorporating historical data from 19 studies identified in the literature.

6.2.1 Binary endpoint

For the binary endpoint, all dynamic borrowing methods yielded posterior means that were very close to the current-only estimate (76.2%). Overall, the historical data are not fully consistent with the observed effect in the current study. Some studies report similar effects, such as those by Yi Ji, Triana, and Sandbank, while others show different results, including studies by Hammer, Tian, Weigand (2018), and Weigand (2022). This pattern suggests a mixed similarity scenario, where some historical studies align with the current data and others differ substantially. As a result, the overall borrowing is limited (**Figure 5**).

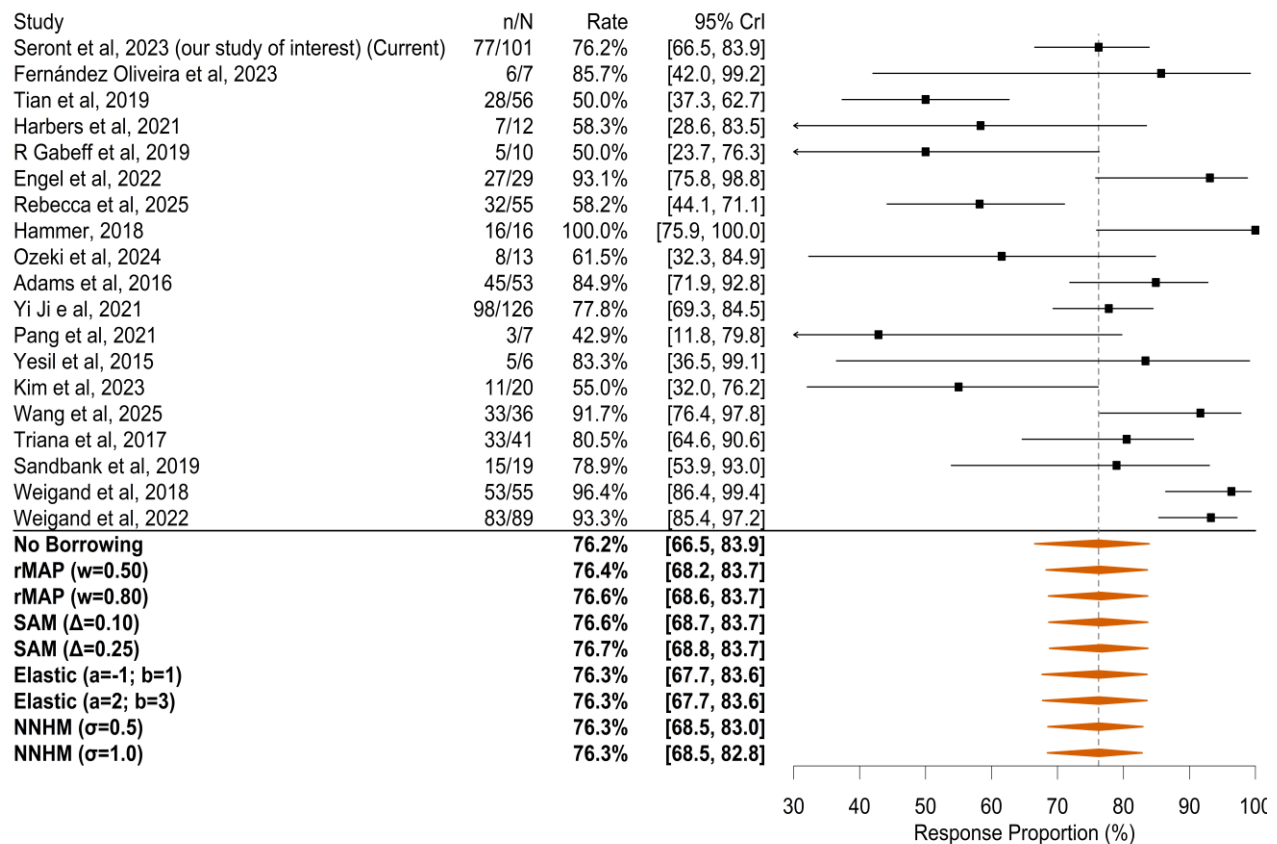


Figure 5. Forest Plot of Current and Historical Studies with Posterior Response Rates under Bayesian Borrowing Methods

rMAP ($w = 0.5-0.8$)

In this setting, rMAP produces credible interval widths of 15.1–15.4 percentage points (pp), modestly narrower than the no-borrowing width of 17.4 pp, indicating a moderate influence of historical data on the final estimate. The borrowed information is limited but noticeable, with an ESS of 13–20 ($\approx 12.8-20.0\%$), reflecting cautious borrowing. Increasing the mixture weight from $w = 0.50$ to $w = 0.80$ leads to a clear gain in precision and ESS, although the improvement remains controlled. This suggests that rMAP balances borrowing and robustness, allowing more information sharing when compatibility with historical data is supported while still protecting against over-borrowing (**Table 6**).

SAM prior

SAM performs similarly to rMAP but shows slightly stronger borrowing. Credible interval widths are 15.0 pp for $\Delta = 0.10$ and 14.9 pp for $\Delta = 0.25$, representing further narrowing compared to rMAP. The borrowed ESS is 20–21, corresponding to $\approx 19.6-20.5\%$ relative ESS. These results indicate a consistent precision gain over no borrowing and slightly greater borrowing than rMAP at $w = 0.50$, while being broadly comparable to rMAP at $w = 0.80$. The minimal differences between Δ values suggest that SAM is stable with respect to its tuning parameter in this setting (**Table 6**).

Elastic prior

Elastic prior yields more conservative borrowing compared to rMAP and SAM. The credible interval widths are 15.8–15.9 pp, only slightly narrower than the no-borrowing interval. The borrowed ESS is 7 in both configurations, corresponding to $\approx 6.5-6.6\%$ relative ESS. This indicates limited information sharing from historical data. Compared with other methods, Elastic shows smaller gains in precision. The near-identical results across parameter settings also indicate low sensitivity to tuning (**Table 6**).

NNHM ($\sigma = 0.5$ and 1.0)

NNHM shows the strongest borrowing among all methods. It achieves the narrowest credible intervals, 14.4 pp for both $\sigma = 0.5$ and $\sigma = 1.0$, together with the largest borrowed ESS values of 28–30, corresponding to $\approx 28.0-29.6\%$ relative ESS. This reflects substantial information sharing from the historical data. Compared with mixture-based approaches, NNHM is clearly more aggressive in borrowing, leading to the greatest precision gains. The similar results across the two σ values suggest that borrowing remains strong even under increased heterogeneity assumptions. (**Table 6**).

Table 6. Posterior Estimates, Credible Interval Width, and Effective Sample Size borrowed under Bayesian Borrowing Methods for the Binary endpoint

Method	Posterior Mean	95% CrI	CrI Width	ESS Borrowed	Relative ESS (%)
No Borrowing	76.20%	[66.5%, 83.9%]	17.4 pp	0	0.00%
rMAP (w=0.50)	76.40%	[68.2%, 83.7%]	15.4 pp	13	12.80%
rMAP (w=0.80)	76.60%	[68.6%, 83.7%]	15.1 pp	20	20.00%
SAM ($\Delta = 0.10$)	76.60%	[68.7%, 83.7%]	15.0 pp	20	19.60%
SAM ($\Delta = 0.25$)	76.70%	[68.8%, 83.7%]	14.9 pp	21	20.50%
Elastic (a=-1, b=1)	76.30%	[67.7%, 83.6%]	15.9 pp	7	6.60%
Elastic (a=2, b=3)	76.30%	[67.7%, 83.6%]	15.8 pp	7	6.50%
NNHM ($\sigma = 0.5$)	76.30%	[68.5%, 83.0%]	14.4 pp	28	28.00%
NNHM ($\sigma = 1.0$)	76.30%	[68.5%, 82.8%]	14.4 pp	30	29.60%

Figure 6 shows that across all parameter settings and sensitivity scenarios, the posterior response proportions remain tightly centered around the observed current-study effect, indicated by the dashed vertical line. This suggests that the borrowing methods do not substantially shift the estimates toward the historical data. Rather, the historical information appears to be discounted because of limited compatibility with the current study, so the current study remains the main driver of posterior inference.

rMAP (top block) shows that posterior means are nearly identical across all 50 scenarios, with relatively consistent and moderately wide credible intervals. This indicates stable and robust borrowing, with limited sensitivity to changes in the robustness weight. The consistency of CI widths suggests controlled borrowing that avoids over-reliance on historical data.

SAM (second block) also centres very tightly around the current study estimate, with slightly narrower and more uniform credible intervals compared to rMAP. The near-overlap of intervals across scenarios indicates that SAM provides stable borrowing with minimal sensitivity to the Δ parameter in this setting.

Elastic (third block) exhibits a small dispersion of posterior means around the dashed line, slightly greater than rMAP and SAM. This reflects sensitivity to the tuning parameters (a, b), which modulate the degree of borrowing. However, the overall spread remains limited, and the estimates stay close to the current study effect, indicating gradual and controlled borrowing.

NNHM (last block) shows the most pronounced reduction in credible interval widths across scenarios, in comparison to the other methods. While the posterior means remain stable and centred near the current estimate, the narrower intervals reflect stronger borrowing driven by hierarchical pooling. In this example, NNHM leverages historical

information more aggressively, leading to greater precision gains while maintaining consistency in point estimates.

The excel sheet containing the posterior means and 95% credible intervals for all the 200 scenarios (50 scenarios per method) are available in [Supporting material VII](#)

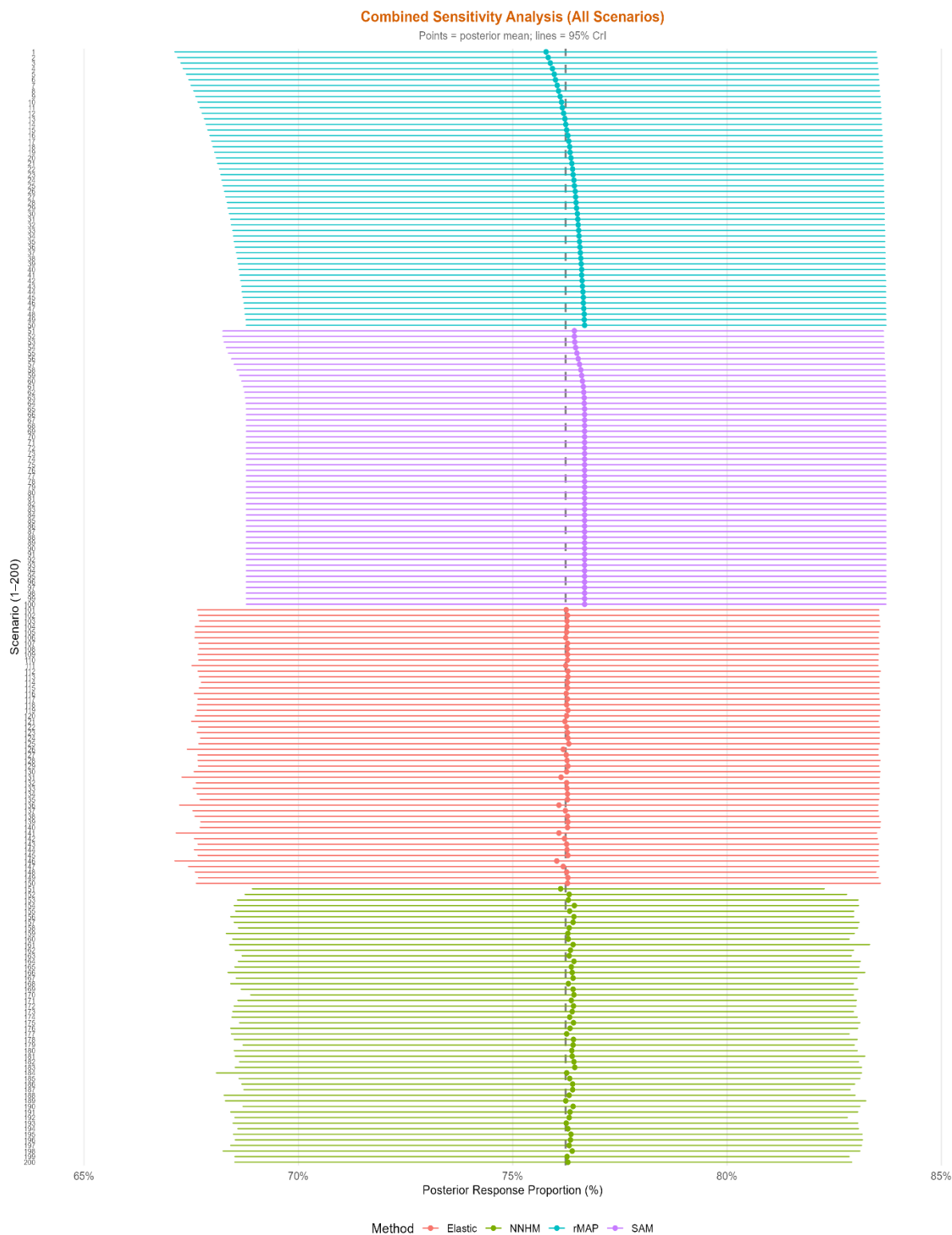


Figure 6. Combined sensitivity analysis across all 200 scenarios (50 per method) for the binary endpoint

The codes used to produce these results are available in appendix : [RWD analysis \(Binary endpoint\)](#) the dataset used in analyzing both the binary endpoint is available in [Supporting material IV](#)

6.2.2 Continuous endpoint

The current study alone gives a mean D-dimer decrease of 1.072 with a 95% CI of 0.082 to 2.062, indicating substantial uncertainty. All three historical studies report larger mean decreases, ranging from 2.340 to 3.240, so incorporating them leads to an upward shift in the posterior mean across all borrowing methods.

Figure 7 shows that all borrowing approaches increase the posterior estimate compared to the no-borrowing analysis, although the magnitude of this shift varies by method.

The ESS results indicate moderate borrowing for rMAP and SAM, with relative ESS of approximately 1 to 16%, while the Elastic prior shows more limited borrowing, around 10 to 11%. The largest borrowing is observed under NNHM, with relative ESS of 19.0% for $\sigma = 0.5$ and 24.0% for $\sigma = 1.0$. This is consistent with the binary endpoint, where NNHM also produced the largest ESS (**Figure 7; Table 7**)

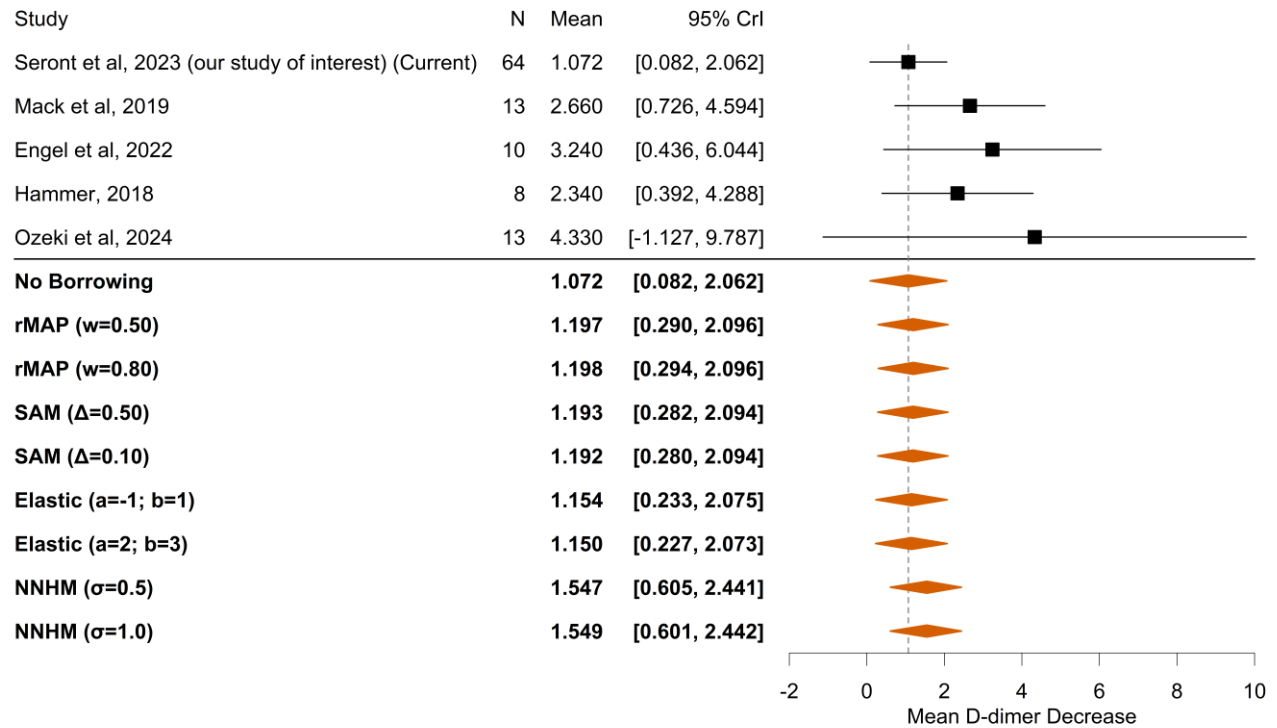


Figure 7. Forest Plot of Observed and Posterior Mean D-dimer Decrease Across Bayesian Borrowing Methods

It is noticeable that both the rMAP and SAM provide controlled borrowing, with posterior means moving only modestly from the no-borrowing estimate of 1.072 to approximately 1.19–1.20. Elastic is slightly more conservative, with posterior means around 1.15. NNHM borrows more strongly from the historical data, shifting the posterior mean further upward to approximately 1.55, while also providing the largest ESS gain (**Table 7**).

Table 7. Posterior Mean D-dimer Decrease, Credible Intervals, and Borrowing Metrics by Method

Method	Posterior Mean [95% CrI]	CrI Width	ESS Borrowed	Relative ESS (%)
No Borrowing	1.072 [0.082, 2.062]	1.981	0	0.00%
rMAP (w=0.50)	1.197 [0.290, 2.096]	1.805	10	15.10%
rMAP (w=0.80)	1.198 [0.294, 2.096]	1.803	10	15.70%
SAM ($\Delta = 0.50$)	1.193 [0.282, 2.094]	1.813	9	14.50%
SAM ($\Delta = 1.00$)	1.192 [0.280, 2.094]	1.814	9	13.90%
Elastic (a=-1, b=1)	1.154 [0.233, 2.075]	1.842	7	11.10%
Elastic (a=2, b=3)	1.150 [0.227, 2.073]	1.847	7	10.50%
NNHM ($\sigma = 0.5$)	1.547 [0.605, 2.441]	1.836	12	19.00%
NNHM ($\sigma = 1.0$)	1.549 [0.601, 2.442]	1.841	15	24.00%

The sensitivity analysis demonstrated that the posterior mean was shifted to the right for all the 200 scenarios and under the different tuning parameters (**Figure 8**).

rMAP (Top block) shows very stable posterior means across all scenarios, tightly clustered around ~ 1.18 – 1.20 . The credible intervals are relatively consistent in width, with minimal variation across scenarios. This indicates robust and controlled borrowing, with little sensitivity to the choice of robustness weight.

SAM (second block) behaves similarly to rMAP, with posterior means also tightly clustered around ~ 1.19 . The intervals are slightly more uniform and show minimal dispersion across scenarios. This suggests stable borrowing with low sensitivity to the tuning parameter (Δ).

Elastic (third block) shows moderate dispersion in posterior means, with values spreading slightly more than rMAP and SAM (roughly ~ 1.10 – 1.25). This reflects greater sensitivity to the tuning parameters (a , b), which control the degree of shrinkage. However, the estimates still remain relatively close to the current-study effect (Middle vertical line), indicating gradual and controlled borrowing, without large shifts toward historical data.

Finally, the **NNHM** (last block) provides the largest variability in both posterior means and credible intervals. Posterior means span a much wider range (approximately ~ 1.3 up to ~ 1.8), and the width of the credible intervals varies a lot across different scenarios and this demonstrates that the NNHM in this example is sensitive to any change in σ levels.

The excel sheet containing the posterior means and 95% credible intervals for all the 200 scenarios (50 scenarios per method) are available in [Supporting material VIII](#)

The codes used to produce these results are available in the appendix: [RWD analysis \(Continuous endpoint\)](#) and the dataset used in analyzing the continuous endpoint is available in [Supporting material IV](#)

All the R scripts with the datasets used to perform the analysis are available in [Supporting Material IX](#)

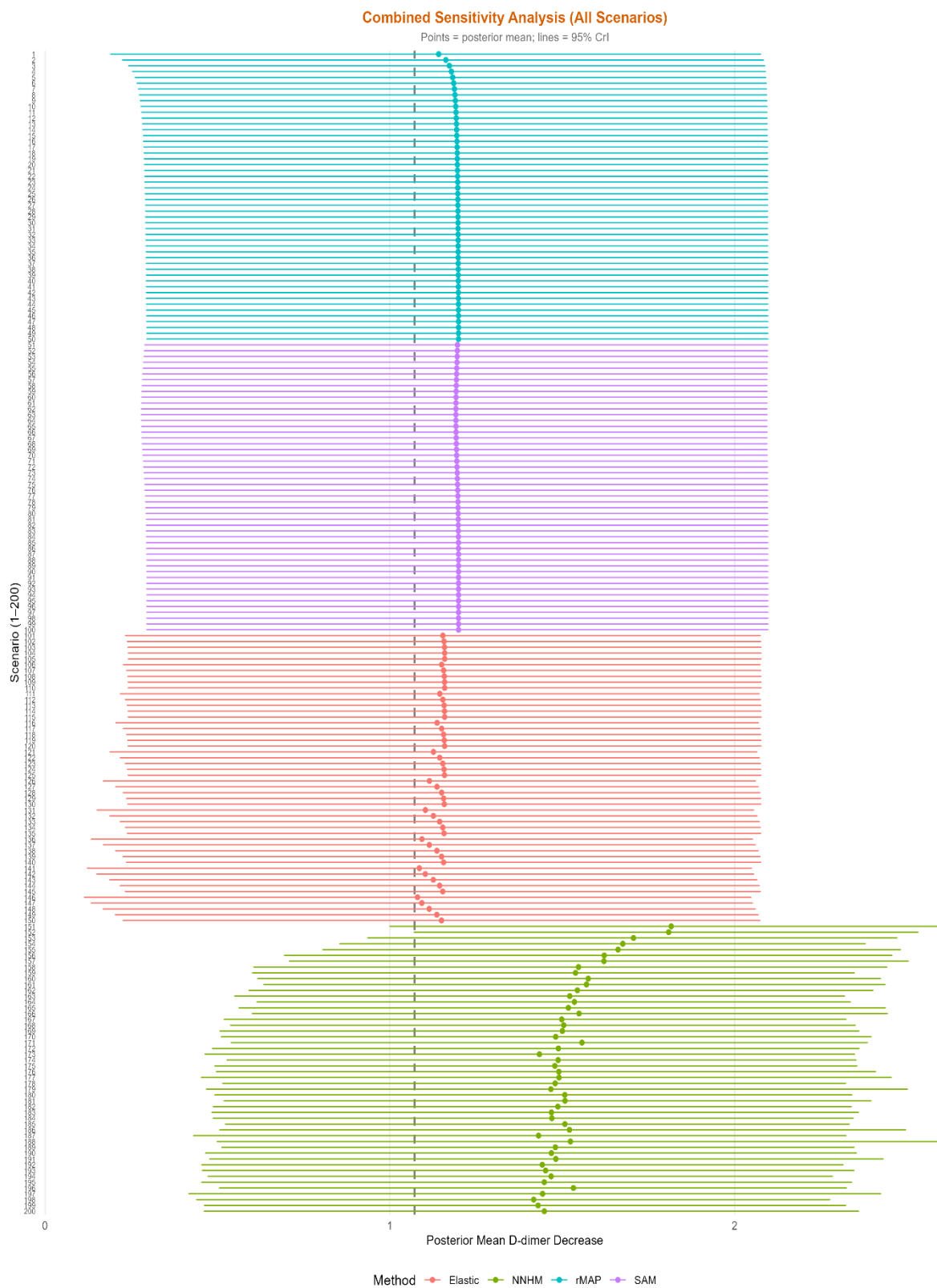


Figure 8. Combined sensitivity analysis across all 200 scenarios (50 per method) for the continuous endpoint

From a **clinical perspective**, the results across both endpoints indicate that incorporating historical data did not change the overall conclusions regarding Sirolimus effectiveness.

For the **pain response endpoint**, the observed response rate in the current study is approximately 76%, and all borrowing methods yield very similar estimates, with only minimal differences in both the posterior means and credible intervals. Although historical studies show a wide range of response rates, these differences limit their influence on the combined analysis. As a result, the inclusion of external data does not meaningfully alter the estimated treatment effect. Clinically, this suggests that the evidence for pain reduction remains robust and is primarily driven by the current study, with historical data providing only minor refinement without changing the overall interpretation.

For the **D-dimer decrease endpoint**, the current study indicates a more modest effect compared to the historical studies, which generally report larger reductions. Incorporating historical data results in an upward shift in the estimated treatment effect across all methods, with this shift being more pronounced under NNHM, reflecting stronger borrowing. However, despite this increase, the estimates remain within a clinically comparable range, and credible intervals continue to overlap with the no-borrowing analysis. This suggests that, although the historical evidence indicates a potentially stronger biological effect, the incorporation of borrowed information does not meaningfully change the overall clinical interpretation of treatment efficacy.

Importantly, the differences observed between borrowing methods do not translate into clinically meaningful differences in conclusions. Even in methods that allow stronger borrowing, such as hierarchical modelling, the resulting estimates remain within a similar range and do not suggest a different clinical decision. Overall, these findings indicate that borrowing historical data can improve precision or slightly adjust estimates, but in this setting, it did not lead to a change in the clinical interpretation of treatment benefit

7. Discussion

Simulated datasets

The primary objective of this part is exploratory, aiming to investigate the behaviour of different Bayesian borrowing techniques in a setting with multiple historical studies with varying degrees of similarity to the current study across different endpoints. Two illustrative scenarios are considered: a relatively large current study ($n = 100$) and a smaller current study ($n = 20$), the latter reflecting a typical rare disease context where borrowing external information could be relevant.

Overall, the results suggest that Bayesian dynamic borrowing methods adjust the extent of borrowing according to the level of agreement between historical and current data, which is a fundamental feature of dynamic borrowing approach (20) (60). However, these findings should be interpreted as illustrative rather than definitive, since there is no replications.

When the historical studies are highly similar to the current study, all methods exhibit substantial borrowing, leading to increased precision and narrower credible intervals compared to the current-only analysis. This is accompanied by a higher effective sample size (ESS), reflecting the contribution of consistent external information. Such behaviour is well established in Bayesian hierarchical and meta-analytic frameworks, where similarity across studies supports stronger pooling of information (61) (62).

Differences between the borrowing approaches become more apparent when discrepancies exist between the historical and current data. The NNHM generally facilitates more continuous borrowing through its exchangeability assumption, which considers study-specific effects as arising from a common underlying distribution and therefore supports information sharing across studies (34). In contrast, the rMAP, SAM, and Elastic prior approaches are specifically designed to mitigate the influence of historical information when prior-data conflict is detected, often resulting in more conservative borrowing under conditions of disagreement (31) (23) (33).

The application of the Bayesian borrowing methods on the 2 simulated realizations also shows the influence of the sample size of the current study on borrowing behaviour. For smaller current sample sizes ($n = 20$), borrowing is more pronounced and differences between methods are more visible, particularly in scenarios with partial or low similarity. In contrast, when the current sample size is larger ($n = 100$), the posterior estimates are primarily driven by the current data, and the gains in precision from borrowing become more limited. This reflects a fundamental principle of Bayesian inference, the relative influence of prior information decreases as the amount of observed data increases (61).

More broadly, these findings illustrate that the posterior distribution represents a balance between prior information and the likelihood. When the current data are limited, borrowing plays a larger role and can substantially influence the estimates. As the sample size of current data increases, the likelihood dominates, reducing the impact of the prior and

leading to posterior estimated near to the current study's estimate. This trade-off is a central feature of Bayesian evidence synthesis and hierarchical modelling (60) (62).

Real-World analysis

In this part, we evaluated the impact of incorporating historical data to estimate the effect of Sirolimus for both pain response after 12 months of treatment (binary endpoint) and the D-dimer decrease after 12 months of treatment (continuous endpoint) using several Bayesian borrowing approaches. Overall, the findings demonstrate that the extent of borrowing is strongly driven by the level of agreement between the current and historical data, as well as by the structure of the borrowing model.

For the **pain response endpoint**, all methods yield posterior estimates very close to the current-only analysis, indicating minimal borrowing from historical data. This can be explained by the substantial heterogeneity observed across the historical studies, with response rates varying widely. Some studies report very high effects (around 90% to 100%), whereas others show much lower response rates (approximately 30-40%), indicating a lack of consistency in the direction and magnitude of the effect. Such variability is common in rare diseases, where individual studies often have small sample sizes and differences in endpoint definitions and reporting.

In such settings, robust borrowing approaches are designed to downweigh external information to avoid bias arising from prior-data conflict. This behaviour is consistent with the framework of robust meta-analytic-predictive priors introduced by Neuenschwander (63) and further developed by Schmidli (31) where borrowing is adaptively reduced when historical data are not similar to the current study. From a methodological perspective, this confirms that these approaches effectively protect against inappropriate borrowing. From a clinical perspective, it reinforces that the conclusions for pain response are robust and primarily driven by the current study.

In contrast, for the **D-dimer decrease endpoint**, all methods resulted in an upward shift of the posterior mean relative to the no-borrowing analysis. This is expected, as all historical studies consistently reported larger treatment effects than the current study. Even though the historical studies are small, they all consistently suggest a larger treatment effect. As a result, they collectively influenced the location of the posterior estimate, shifting the mean upward. However, because of their limited size with respect to the sample size of the current study, they do not substantially improve precision, which explains why the effective sample size remains low (61) (64). This behavior is consistent with the properties of random-effects models, where agreement in effect direction facilitates stronger information sharing across studies as suggested by Higgins (65). The explanation behind this is that hierarchical models tend to pool estimates toward a common mean when the historical studies show effects in the same direction, even when sample sizes are small (65) (34). However, the magnitude of this shift differed across methods, reflecting their underlying borrowing mechanisms. Methods such as rMAP and SAM exhibited moderate borrowing, leading to modest increases in the estimated effect and slight gains in precision. Elastic models were more conservative, particularly under certain

parameterizations, resulting in estimates closer to the current study. In contrast, the hierarchical model (NNHM) produced the largest upward shift.

Overall, the results illustrate that no single borrowing method is universally optimal. Robust approaches such as rMAP and SAM provide protection against prior-data conflict and are well suited for settings with heterogeneous or uncertain historical evidence. Elastic approaches offer additional flexibility through parameter tuning but tend to remain conservative. Hierarchical models enable stronger information borrowing but are more sensitive to assumptions about between-study variability. These findings are consistent with the general principles of Bayesian evidence synthesis described by Spiegelhalter (62), where the balance between bias and efficiency depends on the compatibility of the data sources. Importantly, despite methodological differences, all approaches led to similar overall conclusions, particularly for the pain response endpoint, where borrowing had minimal impact. For the D-dimer endpoint, borrowing led to modest increases in the estimated effect, but these changes did not substantially alter the clinical interpretation.

These findings are particularly relevant in the context of rare diseases, where small sample sizes often limit the ability to draw precise conclusions from a single study. In such settings, the use of historical data represents an important opportunity to improve the precision and support decision-making.

8. Strengths and limitations

This work has several important strengths. First, it presents a two-stage framework combining pre-borrowing weighting with model-based Bayesian borrowing, which represents a novel and practical approach to incorporating historical data. By explicitly accounting for clinical, demographic, and methodological differences between studies before applying Bayesian borrowing. Second, it includes a comparison of four different borrowing methods covering both prior-based approaches and hierarchical likelihood-based models. This allows an investigation of how different methodological assumptions influence borrowing behaviour, providing valuable insight into their relative performance. Finally, the combination of simulation scenarios and real-world application strengthens the relevance of the findings and demonstrates the practical implications of these methods in a clinical context.

These strengths should be considered alongside several limitations. The proposed pre-borrowing weighting framework, while conceptually appealing, involves a degree of subjectivity, particularly in the specification of risk-of-bias and study design weights. Although these choices are grounded in established frameworks, they may influence the resulting borrowing and introduce variability across analysts.

The simulation component is also limited in scope. The results are only based on a single simulated realization per scenario, and do not include a comprehensive evaluation of operating characteristics such as type I error control. Furthermore, the comparison of borrowing methods was restricted to a limited set of configurations, and additional simulations would be required to fully characterize their performance under different conditions of heterogeneity and sample size. These limitations suggest that the simulation findings should be interpreted as illustrative rather than definitive.

An additional limitation is related to the definition of study endpoints used in the analysis of real-world data, which was not fully consistent across the included historical studies. In rare disease settings, endpoints are often patient-reported or defined differently across studies, reflecting variations in clinical practice and study design. This heterogeneity in endpoint definition may introduce additional variability and affect the comparability of results when pooling data through Bayesian borrowing approaches.

Future work should focus on more extensive simulation studies to further evaluate the performance of these methods under a wider range of scenarios. In particular, it would be valuable to explore settings where all the historical data agree on the direction of the effect and other settings where they disagree, as well as scenarios with varying sample sizes for the current study. Such analyses would provide a more comprehensive understanding of the robustness and reliability of borrowing approaches and help strengthen the evidence base for their application.

9. Conclusion

In conclusion, this work shows that Bayesian borrowing methods can be effectively used to incorporate historical data while adapting to the level of agreement between historical studies and the current study. The proposed two-stage framework, combining pre-borrowing weighting with model-based approaches, provides a structured way to account for clinical relevance and methodological quality when integrating external data.

After the application of the borrowing methods to 2 simulated datasets, borrowing seemed to increase when historical studies were highly similar, leading to increased precision and higher effective sample size, whereas it decreased as similarity declined. The simulations also highlighted the impact of the current sample size, with stronger borrowing and clearer differences between methods observed in small-sampled current study, and more limited influence of historical data when the current sample size was larger.

In the real-world analysis, borrowing led to modest improvements in precision and small shifts in estimates, without substantially altering the overall conclusions.

These findings are particularly relevant in the context of rare diseases, where limited sample sizes make the use of external information essential. In such settings, Bayesian dynamic borrowing offers a flexible and robust approach to improve efficiency while maintaining protection against inappropriate information sharing.

10. References

1. Health TLG. The landscape for rare diseases in 2024. The Lancet Global Health. 2024 Mar 1;12(3):e341. doi:10.1016/S2214-109X(24)00056-1 PubMed PMID: 38365397.
2. Sciensano. Rare diseases. Sciensano. Available from: <https://www.sciensano.be/en/health-topics/rare-diseases>
3. Wattal A, Martz L. Rare Diseases and their Unique Challenges. J Stud Res. 2024 Nov 30;13(4). doi:10.47611/jsrhs.v13i4.8088
4. WHO. Rare diseases: a global health priority for equity and inclusion. WHO; 2025. Available from: https://apps.who.int/gb/ebwha/pdf_files/EB156/B156_CONF2-en.pdf
5. Hilgers R. Design and analysis of clinical trials for small rare disease populations. J Rare Dis Res Treat. 2016 Nov 1;1(3):53–60. doi:10.29245/2572-9411/2016/3.1054
6. Wang M, Ma H, Shi Y, Ni H, Qin C, Ji C. Single-arm clinical trials: design, ethics, principles. BMJ Support Palliat Care. 2024 Dec 25;15(1):46–54. doi:10.1136/spcare-2024-004984 PubMed PMID: 38834238; PubMed Central PMCID: PMC11874317.
7. FDA. Rare Diseases: Common Issues in Drug Development Guidance for Industry. FDA; 2019. Available from: <https://www.fda.gov/media/119739/download>
8. FDA. Innovative Designs for Clinical Trials of Cellular and Gene Therapy Products in Small Populations. FDA; 2025. Available from: <https://www.fda.gov/media/188892/download>
9. FDA. Complex Innovative Trial Design Meeting Program. FDA; 2025. Available from: <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program>
10. EMA. Guideline on Clinical Trials in Small Populations. EMA; 2005. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-trials-small-populations_en.pdf
11. ISSVA. Classification for vascular anomalies. International Society For the Study of Vascular Anomalies; 2025. Available from: <https://www.issva.org/classification>
12. Domp Martin A, Vikkula M, Boon LM. Venous Malformation: update on etiopathogenesis, diagnosis & management. Phlebology. 2010;25(5):224–35. doi:10.1258/phleb.2009.009041 PubMed PMID: 20870869; PubMed Central PMCID: PMC3132084.
13. Lokhorst MM, Horbach SER, Waner M, O TM, van der Vleuten CJM, Mokkink LB, et al. Responsiveness of quality-of-life measures in patients with peripheral vascular

malformations: the OVAMA project. *Br J Dermatol*. 2020 Jun;182(6):1395–403. doi:10.1111/bjd.18619 PubMed PMID: 31628861; PubMed Central PMCID: PMC7318211.

14. Hammill AM, Wentzel M, Gupta A, Nelson S, Lucky A, Elluru R, et al. Sirolimus for the treatment of complicated vascular anomalies in children. *Pediatric Blood & Cancer*. 2011 Dec;57(6):1018–24. doi:10.1002/pbc.23124
15. Bernardo JM. Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1979 Jan 1;41(2):113–28. doi:10.1111/j.2517-6161.1979.tb01066.x
16. Su L, Chen X, Zhang J, Yan F. Comparative Study of Bayesian Information Borrowing Methods in Oncology Clinical Trials. *JCO Precis Oncol*. 2022 Mar 9;6:e2100394. doi:10.1200/PO.21.00394 PubMed PMID: 35263169; PubMed Central PMCID: PMC8926037.
17. Lesaffre E, Qi H, Banbeta A, Van Rosmalen J. A review of dynamic borrowing methods with applications in pharmaceutical research. *Braz J Probab Stat*. 2024 Mar 1;38(1). doi:10.1214/24-BJPS598
18. Edwards D, Best N, Crawford J, Zi L, Shelton C, Fowler A. Using Bayesian Dynamic Borrowing to Maximize the Use of Existing Data: A Case-Study. *Ther Innov Regul Sci*. 2024;58(1):1–10. doi:10.1007/s43441-023-00585-3 PubMed PMID: 37910271; PubMed Central PMCID: PMC10764450.
19. Zhang H, Anderson KM, Zimmer Z, Golm G, Sapre A, Ibrahim JG. Prior Effective Sample Size When Borrowing on the Treatment Effect Scale. *Statistics in Medicine*. 2025;44(18–19):e70235. doi:10.1002/sim.70235
20. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. *Biometrics*. 2011 Sep 1;67(3):1047–56. doi:10.1111/j.1541-0420.2011.01564.x
21. Ohigashi T, Maruo K, Sozu T, Gosho M. Nonparametric Bayesian approach for dynamic borrowing of historical control data. *arXiv*. 2024 Nov 18. doi:https://doi.org/10.48550/arXiv.2411.11675
22. Seront E, Van Damme A, Legrand C, Bisdorff-Bresson A, Orcel P, Funck-Brentano T, et al. Preliminary results of the European multicentric phase III trial regarding sirolimus in slow-flow vascular malformations. *JCI Insight*. 2023 Nov 8;8(21):e173095. doi:10.1172/jci.insight.173095
23. Yang P, Zhao Y, Nie L, Vallejo J, Yuan Y. SAM: Self-Adapting Mixture Prior to Dynamically Borrow Information from Historical Data in Clinical Trials. *Biometrics*. 2023 Dec 1;79(4):2857–68. doi:10.1111/biom.13927

24. Ibrahim JG, Chen MH, Gwon Y, Chen F. The Power Prior: Theory and Applications. *Stat Med*. 2015 Dec 10;34(28):3724–49. doi:10.1002/sim.6728 PubMed PMID: 26346180; PubMed Central PMCID: PMC4626399.
25. Chen MH, Ibrahim JG, Shao QM. Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*. 2000 Mar 1;84(1):121–37. doi:10.1016/S0378-3758(99)00140-8
26. Carvalho LM, Ibrahim JG. On the normalized power prior. *Statistics in Medicine*. 2021 Oct 30;40(24):5251–75. doi:10.1002/sim.9124
27. Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*. 2006 Feb;17(1):95–106. doi:10.1002/env.752
28. Weber S, Li Y, Iii JWS, Kakizume T, Schmidli H. Applying Meta-Analytic-Predictive Priors with the *R* Bayesian Evidence Synthesis Tools. *J Stat Soft*. 2021;100(19). doi:10.18637/jss.v100.i19
29. Van Rosmalen J, Dejardin D, Van Norden Y, Löwenberg B, Lesaffre E. Including historical data in the analysis of clinical trials: Is it worth the effort? *Stat Methods Med Res*. 2018 Oct;27(10):3167–82. doi:10.1177/0962280217694506
30. Zhang H, Shen Y, Chiang AY, Li J. An Empirical Bayes Robust Meta-Analytical-Predictive Prior to Adaptively Leverage External Data. *arXiv: arXiv*; 2021. Available from: <http://arxiv.org/abs/2109.10237> doi:10.48550/arXiv.2109.10237
31. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information. *Biometrics*. 2021 Jul 12;70(4):1023–32. doi:10.1111/biom.12242
32. Chiaruttini MV, Lorenzoni G, Gregori D. Bayesian dynamic borrowing in group-sequential design for medical device studies. *BMC Med Res Methodol*. 2025 Mar 20;25(1):78. doi:10.1186/s12874-025-02520-6
33. Jiang L, Nie L, Yuan Y. Elastic priors to dynamically borrow information from historical data in clinical trials. *Biometrics*. 2023 Mar;79(1):49–60. doi:10.1111/biom.13551
34. Röver C, Friede T. Dynamically borrowing strength from another study through shrinkage estimation. *Stat Methods Med Res*. 2020 Jan;29(1):293–308. doi:10.1177/0962280219833079
35. Verde PE. A bias-corrected meta-analysis model for combining, studies of different types and quality. *Biometrical J*. 2021 Feb;63(2):406–22. doi:10.1002/bimj.201900376
36. Fernández Oliveira C, Martínez Roca C, Gómez Tellado M, Salvador Garrido M, Outeda Macías M, Martín Herranz I. Tratamiento con sirolimus oral o tópico en anomalías

vasculares complejas en pediatría. Experiencia en un hospital terciario. *Cir Pediatr.* 2023;36(2):60–6. doi:10.54847/cp.2023.02.03

37. Mack JM, Verkamp B, Richter GT, Nicholas R, Stewart K, Crary SE. Effect of sirolimus on coagulopathy of slow-flow vascular malformations. *Pediatric Blood & Cancer.* 2019 Oct;66(10):e27896. doi:10.1002/pbc.27896
38. Tian R, Liang Y, Zhang W, Wang J, Shan Y, Gao H, et al. Effectiveness of sirolimus in the treatment of complex lymphatic malformations: Single center report of 56 cases. *Journal of Pediatric Surgery.* 2020 Nov;55(11):2454–8. doi:10.1016/j.jpedsurg.2019.12.021
39. Harbers VEM, Rongen GAPJM, Van Der Vleuten CJM, Verhoeven BH, De Laat PCJ, Van Der Horst CMAM, et al. Patients with Congenital Low-Flow Vascular Malformation Treated with Low Dose Sirolimus. *Adv Ther.* 2021 Jun;38(6):3465–82. doi:10.1007/s12325-021-01758-y
40. Gabeff R, Boccara O, Soupre V, Lorette G, Bodemer C, Herbreteau D, et al. Efficacy and Tolerance of Sirolimus (Rapamycin) for Extracranial Arteriovenous Malformations in Children and Adults. *Acta Derm Venerol.* 2019;0. doi:10.2340/00015555-3273
41. Engel ER, Hammill A, Adams D, Phillips RJ, Jeng M, Tollefson MM, et al. Response to sirolimus in capillary lymphatic venous malformations and associated syndromes: Impact on symptomatology, quality of life, and radiographic response. *Pediatric Blood & Cancer.* 2023 Apr;70(4):e30215. doi:10.1002/pbc.30215
42. Nisbet R, Pang C, Evans N, Belhadj A, Khalifa M, Papadopoulou A, et al. Oral sirolimus therapy for patients with complex low-flow vascular malformations. *Journal of Vascular Surgery: Venous and Lymphatic Disorders.* 2025 Sep;13(5):102261. doi:10.1016/j.jvsv.2025.102261
43. Hammer J, Seront E, Duez S, Dupont S, Van Damme A, Schmitz S, et al. Sirolimus is efficacious in treatment for extensive and/or complex slow-flow vascular malformations: a monocentric prospective phase II study. *Orphanet J Rare Dis.* 2018 Dec;13(1):191. doi:10.1186/s13023-018-0934-z
44. Ozeki M, Endo S, Yasue S, Asada R, Saito AM, Hashimoto H, et al. Sirolimus treatment for intractable vascular anomalies (SIVA): An open-label, single-arm, multicenter, prospective trial. *Pediatrics International.* 2025 Jan;67(1):e70002. doi:10.1111/ped.70002
45. Adams DM, Trenor CC, Hammill AM, Vinks AA, Patel MN, Chaudry G, et al. Efficacy and Safety of Sirolimus in the Treatment of Complicated Vascular Anomalies. *Pediatrics.* 2016 Feb 1;137(2):e20153257. doi:10.1542/peds.2015-3257

46. Ji Y, Chen S, Yang K, Zhou J, Zhang X, Jiang X, et al. A prospective multicenter study of sirolimus for complicated vascular anomalies. *Journal of Vascular Surgery*. 2021 Nov;74(5):1673-1681.e3. doi:10.1016/j.jvs.2021.04.071
47. Pang C, Evans N, Jethwa P, Papadopoulou A, Khalifa M, Tsui J, et al. Single Center Experience of Sirolimus Therapy in Head and Neck Low-flow Vascular Malformations. *Vasc Endovascular Surg*. 2021 Jul;55(5):482–90. doi:10.1177/15385744211010378
48. Yesil S, Tanyildiz HG, Bozkurt C, Cakmakci E, Sahin G. Single-center experience with sirolimus therapy for vascular malformations. *Pediatric Hematology and Oncology*. 2016 Apr 2;33(3):219–25. doi:10.3109/08880018.2016.1160170
49. Kim M, Hong KT, Park HJ, Kim BK, Choi JY, Kim HY, et al. Clinical effectiveness and safety of sirolimus in pediatric patients with complex vascular anomalies: necessitating personalized and comprehensive approaches. *Front Pediatr*. 2023 Nov 16;11:1304133. doi:10.3389/fped.2023.1304133
50. Wang J, Meng Y, Zhang X, Li Y, Sun N, Liu Q, et al. A real-world study of sirolimus in the treatment of pediatric head and neck lymphatic malformations. *Journal of Vascular Surgery: Venous and Lymphatic Disorders*. 2025 Jul;13(4):102230. doi:10.1016/j.jvsv.2025.102230
51. Triana P, Dore M, Cerezo V, Cervantes M, Sánchez A, Ferrero M, et al. Sirolimus in the Treatment of Vascular Anomalies. *Eur J Pediatr Surg*. 2016 Oct 10;27(01):086–90. doi:10.1055/s-0036-1593383
52. Sandbank S, Molho-Pessach V, Farkas A, Barzilai A, Greenberger S. Oral and Topical Sirolimus for Vascular Anomalies: A Multicentre Study and Review. *Acta Derm Venerol*. 2019;99(11):990–6. doi:10.2340/00015555-3262
53. Wiegand S, Wichmann G, Dietz A. Treatment of Lymphatic Malformations with the mTOR Inhibitor Sirolimus: A Systematic Review. *Lymphatic Research and Biology*. 2018 Aug;16(4):330–9. doi:10.1089/lrb.2017.0062
54. Wiegand S, Dietz A, Wichmann G. Efficacy of sirolimus in children with lymphatic malformations of the head and neck. *Eur Arch Otorhinolaryngol*. 2022 Aug;279(8):3801–10. doi:10.1007/s00405-022-07378-8
55. National Heart, Lung, and Blood Institute (NHLBI). Study Quality Assessment Tools | NHLBI, NIH. Available from: <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>
56. Oxman AD. Grading quality of evidence and strength of recommendations. *BMJ*. 2004 Jun 19;328(7454):1490. doi:10.1136/bmj.328.7454.1490

57. Gower JC. A General Coefficient of Similarity and Some of Its Properties. *Biometrics*. 1971;27(4):857–71. doi:10.2307/2528823
58. OECD, European Union, Joint Research Centre - European Commission. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD; 2008. Available from: https://www.oecd.org/en/publications/handbook-on-constructing-composite-indicators-methodology-and-user-guide_9789264043466-en.html doi:10.1787/9789264043466-en
59. Morita S, Thall PF, Müller P. Determining the Effective Sample Size of a Parametric Prior. *Biometrics*. 2008 Jun;64(2):595–602. doi:10.1111/j.1541-0420.2007.00888.x
60. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*. 2014;13(1):41–54. doi:10.1002/pst.1589
61. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis* 3rd edition. 2025.
62. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. 2004.
63. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clinical Trials*. 2010 Feb 1;7(1):5–18. doi:10.1177/1740774509356002
64. Röver C, Wu Q, Loos A, Friede T. Bayesian Random-Effects Meta-Analysis of Aggregate Data on Clinical Events. *Statistics in Medicine*. 2026 Apr;45(8–9):e70508. doi:10.1002/sim.70508
65. Higgins JPT, Thompson SG, Spiegelhalter DJ. A Re-Evaluation of Random-Effects Meta-Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2009 Jan 1;172(1):137–59. doi:10.1111/j.1467-985X.2008.00552.x

11. Supporting Information

11.1 Supporting material I

[dataset high similarity n=20.xlsx](#)

[dataset medium similarity n=20.xlsx](#)

[dataset low similarity n=20.xlsx](#)

[dataset mixed similarity n=20.xlsx](#)

[dataset high similarity n=100.xlsx](#)

[dataset medium similarity n=100.xlsx](#)

[dataset low similarity n=100.xlsx](#)

[dataset mixed similarity n=100.xlsx](#)

11.2 Supporting material II

[data extraction.xlsx](#)

11.3 Supporting material III

[ROB.xlsx](#)

11.4 Supporting material IV

[Final dataset used in RWD analysis.xlsx](#)

11.5 Supporting material V

[rmap results n=100.xlsx](#)

[sam prior results n=100.xlsx](#)

[elastic prior results n=100.xlsx](#)

[nnhm results n=100.xlsx](#)

11.6 Supporting material VI

[rmap results n=20.xlsx](#)

[sam prior results n=20.xlsx](#)

[elastic prior results n=20.xlsx](#)

[nnhm results n=20.xlsx](#)

11.7 Supporting material VII

[sensitivity analysis 200 scenarios binary endpoint.xlsx](#)

11.8 Supporting material VIII

[sensitivity analysis 200 scenarios continuous endpoint.xlsx](#)

11.9 Supporting material IX

[Scripts used in the analysis of simulated data](#)

[Scripts used to analyze the binary endpoint](#)

[Scripts used to analyze the continuous endpoint](#)

12. Annex

12.1 Simulation

```
library(dplyr)
library(tidyr)
library(ggplot2)

set.seed(2024)

# =====
#
# PARAMETERS
# =====

CURRENT_N <- 100

# OUTCOME 1: Response proportion
CURRENT_TRUE_PROP <- 0.60
CURRENT_RESPONDERS <- rbinom(1, CURRENT_N, CURRENT_TRUE_PROP)
CURRENT_OBS_PROP <- CURRENT_RESPONDERS / CURRENT_N

# OUTCOME 2: Event rate
CURRENT_TRUE_EVENT_RATE <- 0.25
CURRENT_FOLLOW_UP_TIME <- CURRENT_N * 2
CURRENT_EVENTS <- rpois(1, CURRENT_TRUE_EVENT_RATE * CURRENT_FOLLOW_UP_TIME)
CURRENT_OBS_EVENT_RATE <- CURRENT_EVENTS / CURRENT_FOLLOW_UP_TIME

# OUTCOME 3: Continuous
CURRENT_TRUE_MEAN <- 25.0
CURRENT_TRUE_SD <- 12.0
CURRENT_CONTINUOUS_DATA <- rnorm(CURRENT_N, CURRENT_TRUE_MEAN, CURRENT_TRUE_S
D)
CURRENT_OBS_MEAN <- mean(CURRENT_CONTINUOUS_DATA)
CURRENT_OBS_SD <- sd(CURRENT_CONTINUOUS_DATA)

cat("=====\n")
cat("CURRENT STUDY (IDENTICAL across all 4 scenarios)\n")
cat("=====\n")
cat(sprintf("Sample size: %d\n\n", CURRENT_N))

cat("OUTCOME 1: Overall Response (Proportion)\n")
cat(sprintf("  True: %.2f, Observed: %.3f\n\n", CURRENT_TRUE_PROP, CURRENT_OB
S_PROP))
```

```

cat("OUTCOME 2: Time-to-Event (Event Rate)\n")
cat(sprintf(" True: %.3f, Observed: %.3f\n\n", CURRENT_TRUE_EVENT_RATE, CURR
ENT_OBS_EVENT_RATE))

cat("OUTCOME 3: Continuous Outcome\n")
cat(sprintf(" True mean: %.1f, Observed: %.2f (SD=%.2f)\n\n",
CURRENT_TRUE_MEAN, CURRENT_OBS_MEAN, CURRENT_OBS_SD))

HIST_SAMPLE_SIZES <- c(45, 87, 132, 56, 98, 71, 115, 39, 144, 63)
N_HISTORICAL <- 10

# =====
==
# HELPER FUNCTIONS
# =====
==

generate_proportion <- function(ref_prop, similarity) {
  if (similarity == "high") {
    deviation <- runif(1, -0.03, 0.03)
  } else if (similarity == "medium") {
    deviation <- sample(c(-1, 1), 1) * runif(1, 0.08, 0.12)
  } else {
    deviation <- sample(c(-1, 1), 1) * runif(1, 0.15, 0.25)
  }
  max(0.05, min(0.95, ref_prop + deviation))
}

generate_event_rate <- function(ref_rate, similarity) {
  if (similarity == "high") {
    deviation <- runif(1, -0.02, 0.02)
  } else if (similarity == "medium") {
    deviation <- sample(c(-1, 1), 1) * runif(1, 0.05, 0.10)
  } else {
    deviation <- sample(c(-1, 1), 1) * runif(1, 0.12, 0.20)
  }
  max(0.05, min(0.80, ref_rate + deviation))
}

generate_continuous_mean <- function(ref_mean, similarity) {
  if (similarity == "high") {
    deviation <- runif(1, -2, 2)
  } else if (similarity == "medium") {
    deviation <- sample(c(-1, 1), 1) * runif(1, 5, 10)
  } else {
    deviation <- sample(c(-1, 1), 1) * runif(1, 12, 18)
  }
  ref_mean + deviation
}

```

```

simulate_historical_study <- function(true_prop, true_event_rate, true_mean,
n, study_id, similarity) {
  # Outcome 1
  responders <- rbinom(1, n, true_prop)
  obs_prop <- responders / n
  se_prop <- sqrt(obs_prop * (1 - obs_prop) / n)
  if (se_prop == 0) se_prop <- 0.01

  # Outcome 2
  person_years <- n * 2
  events <- rpois(1, true_event_rate * person_years)
  obs_event_rate <- events / person_years
  se_rate <- sqrt(obs_event_rate / person_years)
  if (se_rate == 0) se_rate <- 0.01

  # Outcome 3
  continuous_data <- rnorm(n, true_mean, CURRENT_TRUE_SD)
  obs_mean <- mean(continuous_data)
  obs_sd <- sd(continuous_data)
  se_mean <- obs_sd / sqrt(n)

  data.frame(
    study_id = study_id,
    study_type = "historical",
    similarity = similarity,
    n = n,
    # Outcome 1
    responders = responders,
    true_proportion = true_prop,
    observed_proportion = obs_prop,
    se_proportion = se_prop,
    ci_lower_proportion = max(0, obs_prop - 1.96 * se_prop),
    ci_upper_proportion = min(1, obs_prop + 1.96 * se_prop),
    # Outcome 2
    person_years = person_years,
    events = events,
    true_event_rate = true_event_rate,
    observed_event_rate = obs_event_rate,
    se_event_rate = se_rate,
    ci_lower_event_rate = max(0, obs_event_rate - 1.96 * se_rate),
    ci_upper_event_rate = obs_event_rate + 1.96 * se_rate,
    # Outcome 3
    true_mean = true_mean,
    observed_mean = obs_mean,
    observed_sd = obs_sd,
    se_mean = se_mean,
    ci_lower_mean = obs_mean - 1.96 * se_mean,
    ci_upper_mean = obs_mean + 1.96 * se_mean,
  )
}

```

```

    stringsAsFactors = FALSE
  )
}

create_current_study_row <- function() {
  se_prop <- sqrt(CURRENT_OBS_PROP * (1 - CURRENT_OBS_PROP) / CURRENT_N)
  se_rate <- sqrt(CURRENT_OBS_EVENT_RATE / CURRENT_FOLLOW_UP_TIME)
  se_mean <- CURRENT_OBS_SD / sqrt(CURRENT_N)

  data.frame(
    study_id = "Current",
    study_type = "current",
    similarity = "reference",
    n = CURRENT_N,
    responders = CURRENT_RESPONDERS,
    true_proportion = CURRENT_TRUE_PROP,
    observed_proportion = CURRENT_OBS_PROP,
    se_proportion = se_prop,
    ci_lower_proportion = max(0, CURRENT_OBS_PROP - 1.96 * se_prop),
    ci_upper_proportion = min(1, CURRENT_OBS_PROP + 1.96 * se_prop),
    person_years = CURRENT_FOLLOW_UP_TIME,
    events = CURRENT_EVENTS,
    true_event_rate = CURRENT_TRUE_EVENT_RATE,
    observed_event_rate = CURRENT_OBS_EVENT_RATE,
    se_event_rate = se_rate,
    ci_lower_event_rate = max(0, CURRENT_OBS_EVENT_RATE - 1.96 * se_rate),
    ci_upper_event_rate = CURRENT_OBS_EVENT_RATE + 1.96 * se_rate,
    true_mean = CURRENT_TRUE_MEAN,
    observed_mean = CURRENT_OBS_MEAN,
    observed_sd = CURRENT_OBS_SD,
    se_mean = se_mean,
    ci_lower_mean = CURRENT_OBS_MEAN - 1.96 * se_mean,
    ci_upper_mean = CURRENT_OBS_MEAN + 1.96 * se_mean,
    stringsAsFactors = FALSE
  )
}

simulate_dataset <- function(scenario_name, similarity_pattern) {
  results <- create_current_study_row()

  for (i in 1:N_HISTORICAL) {
    sim_level <- similarity_pattern[i]
    hist_study <- simulate_historical_study(
      true_prop = generate_proportion(CURRENT_TRUE_PROP, sim_level),
      true_event_rate = generate_event_rate(CURRENT_TRUE_EVENT_RATE, sim_level),
      true_mean = generate_continuous_mean(CURRENT_TRUE_MEAN, sim_level),
      n = HIST_SAMPLE_SIZES[i],
      study_id = paste0("Historical_", i),

```

```

        similarity = sim_level
    )
    results <- rbind(results, hist_study)
}

results$scenario <- scenario_name
return(results)
}

# =====
==
# SIMULATE ALL FOUR SCENARIOS
# =====
==

dataset_high <- simulate_dataset("Scenario 1: High Similarity", rep("high", 1
0))
dataset_low <- simulate_dataset("Scenario 2: Low Similarity", rep("low", 10))
dataset_medium <- simulate_dataset("Scenario 3: Medium Similarity", rep("medi
um", 10))
dataset_mixed <- simulate_dataset("Scenario 4: Mixed Similarity",
                                c(rep("high", 3), rep("low", 3), rep("medi
um", 4)))

all_data <- rbind(dataset_high, dataset_low, dataset_medium, dataset_mixed)

# =====
==
# EXPORT DATA
# =====
==

write.csv(dataset_high, "C:/Users/marin/Downloads/Simulation/dataset_high_sim
ilarity.csv", row.names = FALSE)
write.csv(dataset_low, "C:/Users/marin/Downloads/Simulation/dataset_low_simil
arity.csv", row.names = FALSE)
write.csv(dataset_medium, "C:/Users/marin/Downloads/Simulation/dataset_medium
_similarity.csv", row.names = FALSE)
write.csv(dataset_mixed, "C:/Users/marin/Downloads/Simulation/dataset_mixed_s
imilarity.csv", row.names = FALSE)
write.csv(all_data, "C:/Users/marin/Downloads/Simulation/all_datasets_combine
d.csv", row.names = FALSE)

```

12.2 Bayesian simulated data

```

# =====
==
# Robust Meta-Analytic Predictive (rMAP) Prior Analysis
# =====
==

```

```
# Purpose:
#   Apply MAP and robust MAP (rMAP) borrowing to 4 simulated datasets
#   with 3 endpoints using the RBest package.
#
# MAP framework:
#    $\pi(\theta_c \mid D_c, D_{h1}, \dots, D_{hJ}) \propto L(\theta_c \mid D_c) \times \pi_{\text{MAP}}(\theta_c \mid D_{h1}, \dots, D_{hJ})$ 
#
# Historical MAP prior:
#    $\pi_{\text{MAP}}(\theta_c \mid D_{h1}, \dots, D_{hJ})$ 
#    $= \int N(\theta_c \mid \mu_h, \sigma_h^2) \times \pi(\mu_h, \sigma_h \mid D_{h1}, \dots, D_{hJ}) d\mu_h d\sigma_h$ 
#
# where:
#    $\theta_c$       = true parameter in the current study
#    $\mu_h$       = mean effect across historical studies
#    $\sigma_h$    = between-study heterogeneity among historical studies
#
# For binary and Poisson data:
#   - Binary outcomes are handled in RBest through the Beta-Binomial framework.
#   - Poisson/event-rate outcomes are modeled using a normal approximation on the log scale.
#
# rMAP framework:
#    $\pi_{\text{rMAP}}(\theta_c) = w \times \pi_{\text{MAP}}(\theta_c \mid \text{historical}) + (1 - w) \times \pi_V(\theta_c)$ 
#
# where:
#   w          = prior weight assigned to the MAP component
#    $\pi_V(\theta_c)$  = vague prior
# =====
==

# Install RBest
if (!require("RBest", quietly = TRUE)) {
  install.packages("RBest")
}

library(RBest)
library(dplyr)
library(ggplot2)
library(tidyr)

set.seed(2024)

# =====
==
# LOAD THE SIMULATED DATASETS
# =====
```

```

==
dataset_high <- read.csv("C:/Users/marin/Downloads/Simulation/dataset_high_si
milarity.csv")
dataset_low <- read.csv("C:/Users/marin/Downloads/Simulation/dataset_low_simi
larity.csv")
dataset_medium <- read.csv("C:/Users/marin/Downloads/Simulation/dataset_mediu
m_similarity.csv")
dataset_mixed <- read.csv("C:/Users/marin/Downloads/Simulation/dataset_mixed_
similarity.csv")

all_datasets <- list(
  "High Similarity" = dataset_high,
  "Low Similarity" = dataset_low,
  "Medium Similarity" = dataset_medium,
  "Mixed Similarity" = dataset_mixed
)

# =====
==
# DESIGN CHOICE FOR rMAP
# =====
==
# Conservative robustification weight as recommended in practice
w_map <- 0.3

# =====
==
# OUTCOME 1: RESPONSE PROPORTION (Binary Endpoint)
# =====
==
# For binary endpoints, RBest uses a Beta-Binomial MAP prior.
# Internally, the historical model is based on a random-effects meta-analysis
,
# and the resulting MAP prior is approximated as a mixture of Beta distributi
ons.

results_outcome1 <- list()

for (scenario_name in names(all_datasets)) {

  cat(sprintf("\n--- %s ---\n", scenario_name))

  data <- all_datasets[[scenario_name]]

  current <- data %>% filter(study_type == "current")
  historical <- data %>% filter(study_type == "historical")

```

```

# -----
--
# Step 1: Historical random-effects meta-analysis
# -----
--
# This yields the posterior distribution of ( $\mu_h$ ,  $\sigma_h$ )
# for the historical studies.

map_fit <- gMAP(
  cbind(responders, n - responders) ~ 1 | study_id,
  data = historical,
  family = binomial,
  tau.dist = "HalfNormal", # RBest argument name; represents  $\sigma_h$  prior
  tau.prior = 1,
  beta.prior = 2
)

sigma_h_median <- median(map_fit$tau)

cat("Historical MAP model fitted successfully.\n")
cat(sprintf("Posterior median of historical heterogeneity  $\sigma_h$ : %.3f\n",
sigma_h_median))

# -----
--
# Step 2: Approximate the MAP prior
# -----
--
map_prior <- automixfit(map_fit)
print(map_prior)

# -----
--
# Step 3: Construct robust MAP prior
# -----
--
vague_prior <- mixbeta(c(1, 1, 1)) # Beta(1,1), weak/vague prior

rmap_prior <- mixcombine(
  informative = map_prior,
  vague = vague_prior,
  weight = c(w_map, 1 - w_map)
)

print(rmap_prior)

```

```
# -----
--
# Step 4: Update priors with current study data
# -----
--
posterior_vague <- postmix(vague_prior, r = current$responders, n = current
$n)
posterior_map    <- postmix(map_prior,    r = current$responders, n = current
$n)
posterior_rmap   <- postmix(rmap_prior,   r = current$responders, n = current
$n)

# -----
--
# Step 5: Posterior summaries
# -----
--
summary_vague <- summary(posterior_vague)
summary_map   <- summary(posterior_map)
summary_rmap  <- summary(posterior_rmap)

results_outcome1[[scenario_name]] <- list(
  map_fit = map_fit,
  map_prior = map_prior,
  rmap_prior = rmap_prior,
  vague_prior = vague_prior,
  posterior_vague = posterior_vague,
  posterior_map = posterior_map,
  posterior_rmap = posterior_rmap,
  current = current,
  historical = historical,
  sigma_h = sigma_h_median
)
}

# =====
==
# OUTCOME 2: EVENT RATE (Poisson Endpoint)
# =====
==
# For Poisson/event-rate data, the model is formulated using a normal
# approximation on the log scale:
#   log(rate_j) approximately follows a normal model
# with standard error approximately equal to 1 / sqrt(events_j)

results_outcome2 <- list()

for (scenario_name in names(all_datasets)) {
```

```

cat(sprintf("\n--- %s ---\n", scenario_name))

data <- all_datasets[[scenario_name]]

current <- data %>% filter(study_type == "current")
historical <- data %>% filter(study_type == "historical")

historical <- historical %>%
  mutate(
    log_rate = log(observed_event_rate + 0.001),
    se_log_rate = 1 / sqrt(pmax(events, 1))
  )

current <- current %>%
  mutate(
    log_rate = log(observed_event_rate + 0.001),
    se_log_rate = 1 / sqrt(pmax(events, 1))
  )

# Historical random-effects meta-analysis on the log-rate scale
map_fit <- gMAP(
  log_rate ~ 1 | study_id,
  weights = se_log_rate^(-2),
  data = historical,
  family = gaussian,
  tau.dist = "HalfNormal", # RBest argument name; represents sigma_h prio
  tau.prior = 0.5,
  beta.prior = 2
)

sigma_h_median <- median(map_fit$tau)

# MAP prior approximation as a mixture of normals
map_prior <- automixfit(map_fit)
cat("MAP prior on the log-rate scale:\n")
print(map_prior)

# Vague prior on log-rate scale
vague_prior <- mixnorm(c(1, log(0.25), 2))

# Robust MAP prior
rmap_prior <- mixcombine(
  informative = map_prior,
  vague = vague_prior,

```

```

    weight = c(w_map, 1 - w_map)
  )

  # Update with current study data
  posterior_vague <- postmix(vague_prior, m = current$log_rate, se = current$
se_log_rate)
  posterior_map <- postmix(map_prior, m = current$log_rate, se = current$
se_log_rate)
  posterior_rmap <- postmix(rmap_prior, m = current$log_rate, se = current$
se_log_rate)

  # Summaries back-transformed to the rate scale

  results_outcome2[[scenario_name]] <- list(
    map_fit = map_fit,
    map_prior = map_prior,
    rmap_prior = rmap_prior,
    vague_prior = vague_prior,
    posterior_vague = posterior_vague,
    posterior_map = posterior_map,
    posterior_rmap = posterior_rmap,
    current = current,
    historical = historical,
    sigma_h = sigma_h_median
  )
}

# =====
==
# OUTCOME 3: CONTINUOUS MEAN
# =====
==
# For continuous outcomes, the model is formulated directly on the normal sca
le.

results_outcome3 <- list()

for (scenario_name in names(all_datasets)) {

  cat(sprintf("\n--- %s ---\n", scenario_name))

  data <- all_datasets[[scenario_name]]

  current <- data %>% filter(study_type == "current")
  historical <- data %>% filter(study_type == "historical")

  # Historical random-effects meta-analysis on the observed mean scale

```

```

map_fit <- gMAP(
  observed_mean ~ 1 | study_id,
  weights = n / observed_sd^2,
  data = historical,
  family = gaussian,
  tau.dist = "HalfNormal", # RBest argument name; represents sigma_h prio
r
  tau.prior = 10,
  beta.prior = 2
)

sigma_h_median <- median(map_fit$tau)

map_prior <- automixfit(map_fit)
cat("MAP prior on the normal scale:\n")
print(map_prior)

# Vague prior
vague_prior <- mixnorm(c(1, 25, 20))

# Robust MAP prior
rmap_prior <- mixcombine(
  informative = map_prior,
  vague = vague_prior,
  weight = c(w_map, 1 - w_map)
)

# Current study standard error
current_se <- current$observed_sd / sqrt(current$n)

# Update with current study data
posterior_vague <- postmix(vague_prior, m = current$observed_mean, se = current_se)
posterior_map <- postmix(map_prior, m = current$observed_mean, se = current_se)
posterior_rmap <- postmix(rmap_prior, m = current$observed_mean, se = current_se)

results_outcome3[[scenario_name]] <- list(
  map_fit = map_fit,
  map_prior = map_prior,
  rmap_prior = rmap_prior,
  vague_prior = vague_prior,
  posterior_vague = posterior_vague,
  posterior_map = posterior_map,
  posterior_rmap = posterior_rmap,
  current = current,

```

```

    historical = historical,
    sigma_h = sigma_h_median
  )
}

# =====
==
# CREATE SUMMARY TABLE
# =====
==

summary_table <- data.frame()

for (scenario_name in names(all_datasets)) {

  # -----
  --
  # Outcome 1: Response Proportion
  # -----
  --
  res1 <- results_outcome1[[scenario_name]]
  summary_table <- rbind(summary_table, data.frame(
    Scenario = scenario_name,
    Outcome = "Response Proportion",
    Method = c("No Borrowing", "MAP", "rMAP"),
    Posterior_Mean = c(
      summary(res1$posterior_vague)["mean"],
      summary(res1$posterior_map)["mean"],
      summary(res1$posterior_rmap)["mean"]
    ),
    CI_Lower = c(
      qmix(res1$posterior_vague, 0.025),
      qmix(res1$posterior_map, 0.025),
      qmix(res1$posterior_rmap, 0.025)
    ),
    CI_Upper = c(
      qmix(res1$posterior_vague, 0.975),
      qmix(res1$posterior_map, 0.975),
      qmix(res1$posterior_rmap, 0.975)
    ),
    Historical_Heterogeneity_sigma_h = res1$sigma_h
  ))

  # -----
  --
  # Outcome 2: Event Rate
  # -----
  --
  res2 <- results_outcome2[[scenario_name]]

```

```
summary_table <- rbind(summary_table, data.frame(
  Scenario = scenario_name,
  Outcome = "Event Rate",
  Method = c("No Borrowing", "MAP", "rMAP"),
  Posterior_Mean = c(
    exp(summary(res2$posterior_vague)["mean"]),
    exp(summary(res2$posterior_map)["mean"]),
    exp(summary(res2$posterior_rmap)["mean"])
  ),
  CI_Lower = c(
    exp(qmix(res2$posterior_vague, 0.025)),
    exp(qmix(res2$posterior_map, 0.025)),
    exp(qmix(res2$posterior_rmap, 0.025))
  ),
  CI_Upper = c(
    exp(qmix(res2$posterior_vague, 0.975)),
    exp(qmix(res2$posterior_map, 0.975)),
    exp(qmix(res2$posterior_rmap, 0.975))
  ),
  Historical_Heterogeneity_sigma_h = res2$sigma_h
))

# -----
--
# Outcome 3: Continuous Mean
# -----
--
res3 <- results_outcome3[[scenario_name]]
summary_table <- rbind(summary_table, data.frame(
  Scenario = scenario_name,
  Outcome = "Continuous Mean",
  Method = c("No Borrowing", "MAP", "rMAP"),
  Posterior_Mean = c(
    summary(res3$posterior_vague)["mean"],
    summary(res3$posterior_map)["mean"],
    summary(res3$posterior_rmap)["mean"]
  ),
  CI_Lower = c(
    qmix(res3$posterior_vague, 0.025),
    qmix(res3$posterior_map, 0.025),
    qmix(res3$posterior_rmap, 0.025)
  ),
  CI_Upper = c(
    qmix(res3$posterior_vague, 0.975),
    qmix(res3$posterior_map, 0.975),
    qmix(res3$posterior_rmap, 0.975)
  ),
  Historical_Heterogeneity_sigma_h = res3$sigma_h
))
}
```

```
# Round for display
summary_table$Posterior_Mean <- round(summary_table$Posterior_Mean, 3)
summary_table$CI_Lower <- round(summary_table$CI_Lower, 3)
summary_table$CI_Upper <- round(summary_table$CI_Upper, 3)
summary_table$Historical_Heterogeneity_sigma_h <- round(summary_table$Historical_Heterogeneity_sigma_h, 3)

print(summary_table)

# Save summary table
write.csv(
  summary_table,
  "C:/Users/marin/Downloads/R scripts/rMAP/rmap_summary_table.csv",
  row.names = FALSE
)

save(
  results_outcome1, results_outcome2, results_outcome3, summary_table,
  file = "C:/Users/marin/Downloads/R scripts/rMAP/rmap_results.RData"
)

#####
#####
# =====
==
# SAM Prior (Self-Adapting Mixture) Analysis
# =====
==
#
# THEORY-ALIGNED IMPLEMENTATION
#
# SAM prior:
#  $\pi(\theta_c \mid D_c, D_h) \propto L(\theta_c \mid D_c) \times \pi_{\text{SAM}}(\theta_c \mid D_h)$ 
#
#  $\pi_{\text{SAM}}(\theta_c \mid D_h) = w \times \pi_{\text{inf}}(\theta_c \mid D_h) + (1 - w) \times \pi_V(\theta_c)$ 
#
# where:
#  $\pi_{\text{inf}}(\theta_c \mid D_h) = \pi_{\text{MAP}}(\theta_c \mid D_{h1}, \dots, D_{hJ})$ 
#
# In this implementation:
# - The informative prior is the MAP prior obtained from the historical studies
#   using RBest.
# - For the binary endpoint, the MAP prior is implemented using the
#   Beta-Binomial framework in RBest.
# - For the event-rate endpoint, log(event rate) is treated as approximately
#   normal, exactly as in the rMAP and Elastic sections.
```

```
# - For the continuous endpoint, the model is formulated on the normal scale.
# - The vague prior is Beta(1,1) for the binary endpoint and a vague normal prior for the approximately normal endpoints.
# - The SAM prior weight w is computed with the SAMprior package.
#
# =====
==

if (!require("SAMprior", quietly = TRUE)) {
  install.packages("SAMprior")
}

library(RBest)
library(SAMprior)
library(dplyr)
library(ggplot2)

set.seed(2024)

# =====
==
# HELPERS
# =====
==

# Recover the posterior informative mixture weight  $w^*$ 
# by summing the posterior weights corresponding to the informative component s.
get_w_posterior_from_postmix <- function(sam_post, informative_prior) {
  post_mat <- as.matrix(sam_post)
  inf_mat <- as.matrix(informative_prior)

  if (is.null(dim(post_mat))) post_mat <- matrix(post_mat, nrow = 1)
  if (is.null(dim(inf_mat))) inf_mat <- matrix(inf_mat, nrow = 1)

  n_inf <- nrow(inf_mat)

  if (nrow(post_mat) < n_inf) {
    stop("Posterior mixture has fewer components than informative prior.")
  }

  sum(post_mat[seq_len(n_inf), 1])
}

# Deterministic pseudo-data for normal endpoints:
# returns a vector with exactly the requested sample mean and sample SD.
make_normal_pseudodata <- function(mean_target, sd_target, n) {
  n <- max(as.integer(round(n)), 2L)
}
```

```

z <- qnorm(ppoints(n))
z <- (z - mean(z)) / sd(z)

mean_target + sd_target * z
}

# =====
==
# LOAD DATA AND RUN ANALYSIS
# =====
==

datasets <- list(
  "High Similarity" = read.csv("C:/Users/marin/Downloads/Simulation/dataset_high_similarity.csv"),
  "Low Similarity" = read.csv("C:/Users/marin/Downloads/Simulation/dataset_low_similarity.csv"),
  "Medium Similarity" = read.csv("C:/Users/marin/Downloads/Simulation/dataset_medium_similarity.csv"),
  "Mixed Similarity" = read.csv("C:/Users/marin/Downloads/Simulation/dataset_mixed_similarity.csv")
)

all_results <- list()

for (scenario_name in names(datasets)) {

  cat(sprintf("\n%s\n%s\n", scenario_name, strrep("=", nchar(scenario_name))))
)

  data <- datasets[[scenario_name]]
  current <- data %>% filter(study_type == "current")
  historical <- data %>% filter(study_type == "historical")

  scenario_results <- list()

  # =====
==
  # OUTCOME 1: RESPONSE PROPORTION (binary endpoint)
  # =====
==

  responders_h <- round(historical$observed_proportion * historical$n)
  nonresponders_h <- historical$n - responders_h

  historical_bin <- data.frame(

```

```

    responders = responders_h,
    nonresponders = nonresponders_h,
    study_id = factor(seq_len(nrow(historical)))
  )

map_fit_1 <- gMAP(
  cbind(responders, nonresponders) ~ 1 | study_id,
  data = historical_bin,
  family = binomial,
  tau.dist = "HalfNormal",
  tau.prior = 1,
  beta.prior = 2
)

map_prior_1 <- automixfit(map_fit_1)
vague_prior_1 <- mixbeta(c(1, 1, 1)) # Beta(1,1)

theta_hat_h_1 <- unname(summary(map_prior_1)["mean"])
Delta_1 <- 0.10

x_c_1 <- round(current$observed_proportion * current$n)
n_c_1 <- current$n

# Current binary data as a 0/1 vector for SAMprior
y_current_1 <- c(rep(1, x_c_1), rep(0, n_c_1 - x_c_1))

sam_w_1 <- SAM_weight(
  if.prior = map_prior_1,
  nf.prior = vague_prior_1,
  delta    = Delta_1,
  data     = y_current_1
)

sam_w_1 <- min(max(sam_w_1, 1e-8), 1 - 1e-8)

sam_prior_1 <- SAM_prior(
  if.prior = map_prior_1,
  nf.prior = vague_prior_1,
  weight   = sam_w_1
)

sam_post_1 <- postmix(sam_prior_1, r = x_c_1, n = n_c_1)

sam_w_post_1 <- get_w_posterior_from_postmix(
  sam_post = sam_post_1,
  informative_prior = map_prior_1
)

```

```

scenario_results$outcome1 <- list(
  map_fit = map_fit_1,
  map_prior = map_prior_1,
  vague_prior = vague_prior_1,
  sam_prior = sam_prior_1,
  sam_post = sam_post_1,
  theta_hat_h = theta_hat_h_1,
  sigma_h = median(map_fit_1$tau),
  Delta = Delta_1,
  w_prior = sam_w_1,
  w_posterior = sam_w_post_1,
  mean_transformed = unname(summary(sam_post_1)["mean"]),
  ci_lower_transformed = qmix(sam_post_1, 0.025),
  ci_upper_transformed = qmix(sam_post_1, 0.975)
)

# =====
==
# OUTCOME 2: EVENT RATE (treated as normal on the log-rate scale)
# =====
==

cat("\n--- OUTCOME 2: Event Rate ---\n")

historical_2 <- historical %>%
  mutate(
    log_rate = log(observed_event_rate + 0.001),
    se_log_rate = 1 / sqrt(pmax(events, 1)),
    study_id = factor(seq_len(n()))
  )

current_2 <- current %>%
  mutate(
    log_rate = log(observed_event_rate + 0.001),
    se_log_rate = 1 / sqrt(pmax(events, 1))
  )

map_fit_2 <- gMAP(
  log_rate ~ 1 | study_id,
  weights = se_log_rate^(-2),
  data = historical_2,
  family = gaussian,
  tau.dist = "HalfNormal",
  tau.prior = 0.5,
  beta.prior = 2
)

```

```

map_prior_2 <- automixfit(map_fit_2)

# Vague normal prior on the log-rate scale
vague_prior_2 <- mixnorm(c(1, log(0.25), 2))

mu_h_hat_2 <- unname(summary(map_prior_2)["mean"])
Delta_2 <- 0.30

# Use n_eff = events so that the current SE is respected under the normal approximation.
n_eff_2 <- max(round(1 / current_2$se_log_rate^2), 2L)
y_current_2 <- make_normal_pseudodata(
  mean_target = current_2$log_rate,
  sd_target   = 1,
  n           = n_eff_2
)

sam_w_2 <- SAM_weight(
  if.prior = map_prior_2,
  nf.prior = vague_prior_2,
  delta    = Delta_2,
  data     = y_current_2
)

sam_w_2 <- min(max(sam_w_2, 1e-8), 1 - 1e-8)

sam_prior_2 <- SAM_prior(
  if.prior = map_prior_2,
  nf.prior = vague_prior_2,
  weight   = sam_w_2
)

sam_post_2 <- postmix(
  sam_prior_2,
  m = current_2$log_rate,
  se = current_2$se_log_rate
)

sam_w_post_2 <- get_w_posterior_from_postmix(
  sam_post = sam_post_2,
  informative_prior = map_prior_2
)

scenario_results$outcome2 <- list(

```

```

map_fit = map_fit_2,
map_prior = map_prior_2,
vague_prior = vague_prior_2,
sam_prior = sam_prior_2,
sam_post = sam_post_2,
mu_h_hat = mu_h_hat_2,
sigma_h = median(map_fit_2$tau),
Delta = Delta_2,
w_prior = sam_w_2,
w_posterior = sam_w_post_2,
mean_transformed = exp(unnamed(summary(sam_post_2)["mean"])),
ci_lower_transformed = exp(qmix(sam_post_2, 0.025)),
ci_upper_transformed = exp(qmix(sam_post_2, 0.975))
)

# =====
==
# OUTCOME 3: CONTINUOUS MEAN
# =====
==

historical_3 <- historical %>%
  mutate(
    se_mean = observed_sd / sqrt(n),
    study_id = factor(seq_len(n()))
  )

current_3 <- current %>%
  mutate(
    se_mean = observed_sd / sqrt(n)
  )

map_fit_3 <- gMAP(
  observed_mean ~ 1 | study_id,
  weights = se_mean^(-2),
  data = historical_3,
  family = gaussian,
  tau.dist = "HalfNormal",
  tau.prior = 10,
  beta.prior = 2
)

map_prior_3 <- automixfit(map_fit_3)

# Vague normal prior on the mean scale
vague_prior_3 <- mixnorm(c(1, 25, 20))

mu_h_hat_3 <- unnamed(summary(map_prior_3)["mean"])

```

```

Delta_3 <- 3

y_current_3 <- make_normal_pseudodata(
  mean_target = current_3$observed_mean,
  sd_target   = current_3$observed_sd,
  n           = current_3$n
)

sam_w_3 <- SAM_weight(
  if.prior = map_prior_3,
  nf.prior = vague_prior_3,
  delta    = Delta_3,
  data     = y_current_3
)

sam_w_3 <- min(max(sam_w_3, 1e-8), 1 - 1e-8)

sam_prior_3 <- SAM_prior(
  if.prior = map_prior_3,
  nf.prior = vague_prior_3,
  weight   = sam_w_3
)

sam_post_3 <- postmix(
  sam_prior_3,
  m = current_3$observed_mean,
  se = current_3$se_mean
)

sam_w_post_3 <- get_w_posterior_from_postmix(
  sam_post = sam_post_3,
  informative_prior = map_prior_3
)

scenario_results$outcome3 <- list(
  map_fit = map_fit_3,
  map_prior = map_prior_3,
  vague_prior = vague_prior_3,
  sam_prior = sam_prior_3,
  sam_post = sam_post_3,
  mu_h_hat = mu_h_hat_3,
  sigma_h = median(map_fit_3$tau),
  Delta = Delta_3,
  w_prior = sam_w_3,
  w_posterior = sam_w_post_3,
  mean_transformed = unname(summary(sam_post_3)["mean"]),
  ci_lower_transformed = qmix(sam_post_3, 0.025),

```

```

    ci_upper_transformed = qmix(sam_post_3, 0.975)
  )

  all_results[[scenario_name]] <- scenario_results
}

# =====
#
# CREATE SUMMARY TABLE
# =====
#

cat("\n\n")
cat("===== \n")
cat("SAM PRIOR SUMMARY TABLE\n")
cat("===== \n\n")

summary_rows <- list()
row_idx <- 1

for (scenario_name in names(all_results)) {
  res <- all_results[[scenario_name]]

  for (outcome_info in list(
    list(name = "Response Proportion", key = "outcome1"),
    list(name = "Event Rate", key = "outcome2"),
    list(name = "Continuous Mean", key = "outcome3")
  )) {
    outcome_res <- res[[outcome_info$key]]

    summary_rows[[row_idx]] <- data.frame(
      Scenario = scenario_name,
      Outcome = outcome_info$name,
      historical_mean = round(
        if (!is.null(outcome_res$theta_hat_h)) outcome_res$theta_hat_h else outcome_res$mu_h_hat,
        3
      ),
      historical_sigma_h = round(outcome_res$sigma_h, 3),
      Delta = round(outcome_res$Delta, 3),
      w = round(outcome_res$w_prior, 3),
      w_posterior = round(outcome_res$w_posterior, 3),
      Posterior_Mean = round(outcome_res$mean_transformed, 3),
      CI_Lower = round(outcome_res$ci_lower_transformed, 3),
      CI_Upper = round(outcome_res$ci_upper_transformed, 3)
    )
  }
}

```

```

    row_idx <- row_idx + 1
  }
}

summary_table <- do.call(rbind, summary_rows)
print(summary_table)

write.csv(
  summary_table,
  "C:/Users/marin/Downloads/R scripts/SAM/sam_prior_results.csv",
  row.names = FALSE
)

outcome1 <- all_results[[names(all_results)[1]]]$outcome1
outcome2 <- all_results[[names(all_results)[1]]]$outcome2
outcome3 <- all_results[[names(all_results)[1]]]$outcome3

save(
  all_results, summary_table,
  file = "C:/Users/marin/Downloads/R scripts/SAM/SAM_results.RData"
)

#####
#####
# =====
==
# Elastic Prior Analysis
# =====
==
#
# Elastic Prior Framework:
#  $\pi(\theta_c \mid D_c, D_h) \propto L(\theta_c \mid D_c) \times \pi_{\text{Elastic}}(\theta_c \mid D_h)$ 
#
# Where:
#    $\theta_c$            = true effect parameter in current study
#    $D_c$              = current study data
#    $D_h$              = historical data
#    $L(\theta_c \mid D_c)$  = likelihood from current study
#    $\pi_{\text{Elastic}}(\theta_c \mid D_h)$  = elastic-adjusted prior from historical data
#
# Construction in 3 Steps:
#
# STEP 1:
#   Construct a posterior distribution using only the historical data:
#    $\pi(\theta_h \mid D_h) \propto L(\theta_h \mid D_h) \times \pi_V(\theta_h)$ 
#
#   In this implementation,  $\theta_h$  is summarized using a fully Bayesian MAP
#   approach applied to the historical studies, yielding:

```

```
#       $\mu_{\hat{h}}$       = mean of the historical posterior
#       $\sigma_{\hat{h}}^2$   = variance of the historical posterior
#
# STEP 2:
#   Quantify discrepancy and compute discount function:
#    $T = (\theta_{\hat{c}} - \mu_{\hat{h}})^2 / (\text{Var}(\theta_{\hat{c}}) + \sigma_{\hat{h}}^2)$ 
#    $g(T) = 1 / (1 + \exp(a + b \times \log(T)))$ 
#
# STEP 3:
#   Inflate historical posterior variance:
#    $\pi_{\text{Elastic}}(\theta_c \mid D_h) \sim N(\mu_{\hat{h}}, \sigma_{\hat{h}}^2 \times g(T)^{(-1)})$ 
#
# Notes:
#   - For binary outcomes, the model is formulated on the logit scale.
#   - For event-rate outcomes, the model is formulated on the log scale.
#   - For continuous outcomes, the model is formulated on the original scale.
# =====
==

library(RBest)
library(dplyr)
library(ggplot2)

set.seed(2024)

# =====
==
# HELPER FUNCTIONS
# =====
==

adjust_prop <- function(prop, n) {
  out <- prop
  out[prop <= 0] <- 0.5 / (n[prop <= 0] + 1)
  out[prop >= 1] <- (n[prop >= 1] + 0.5) / (n[prop >= 1] + 1)
  out
}

logit <- function(p) log(p / (1 - p))
inv_logit <- function(x) 1 / (1 + exp(-x))

# -----
--
# Step 1: Extract historical posterior summaries from MAP
# -----
--
# For binary data, the MAP prior is a Beta mixture on the probability scale.
# To keep the Elastic prior formulation consistent, we transform samples to
# the logit scale and summarize them by  $\mu_{\hat{h}}$  and  $\sigma_{\hat{h}}^2$ .
```

```
#
# For Poisson/event-rate and continuous outcomes, the MAP prior is summarized
# directly on the modeling scale.
# -----
--

extract_historical_posterior_from_map <- function(map_prior,
                                                outcome_type = c("binary",
                                                                "poisson", "continuous"),
                                                n_sim = 50000) {
  outcome_type <- match.arg(outcome_type)

  if (outcome_type == "binary") {
    p_samp <- rmix(map_prior, n_sim)
    p_samp <- pmin(pmax(p_samp, 1e-8), 1 - 1e-8)
    theta_samp <- logit(p_samp)
  } else {
    theta_samp <- rmix(map_prior, n_sim)
  }

  mu_h_hat <- mean(theta_samp)
  sigma_h_sq_hat <- var(theta_samp)

  list(
    mu_h_hat = mu_h_hat,
    sigma_h_sq_hat = sigma_h_sq_hat,
    sigma_h_hat = sqrt(sigma_h_sq_hat)
  )
}

# -----
--
# Step 2: Calculate discrepancy statistic and discount function
# -----
--

calculate_discount <- function(theta_c_hat, var_theta_c_hat,
                              mu_h_hat, sigma_h_sq_hat,
                              a = 0, b = 2) {

  T_stat <- (theta_c_hat - mu_h_hat)^2 / (var_theta_c_hat + sigma_h_sq_hat)
  T_stat <- max(T_stat, 1e-10)

  g_T <- 1 / (1 + exp(a + b * log(T_stat)))

  list(
    T = T_stat,
    g_T = g_T,
    a = a,

```

```

    b = b
  )
}

# -----
--
# Step 3: Compute elastic posterior
# -----
--

compute_elastic_posterior <- function(theta_c_hat, var_theta_c_hat,
                                     mu_h_hat, sigma_h_sq_hat, g_T) {

  g_T_safe <- max(g_T, 0.001)

  sigma_Elastic_sq <- sigma_h_sq_hat / g_T_safe
  sigma_Elastic <- sqrt(sigma_Elastic_sq)

  sigma_post_sq <- 1 / (1 / sigma_Elastic_sq + 1 / var_theta_c_hat)
  mu_post <- sigma_post_sq * (mu_h_hat / sigma_Elastic_sq + theta_c_hat / var_theta_c_hat)
  sigma_post <- sqrt(sigma_post_sq)

  ci_lower <- qnorm(0.025, mu_post, sigma_post)
  ci_upper <- qnorm(0.975, mu_post, sigma_post)

  list(
    mean = mu_post,
    sd = sigma_post,
    ci_lower = ci_lower,
    ci_upper = ci_upper,
    sigma_Elastic_sq = sigma_Elastic_sq,
    sigma_Elastic = sigma_Elastic,
    g_T = g_T_safe
  )
}

a_param <- 0
b_param <- 2

cat(sprintf("Tuning parameters: a = %.1f, b = %.1f\n\n", a_param, b_param))

all_results <- list()

for (scenario_name in names(datasets)) {

  cat(sprintf("\n%s\n%s\n", scenario_name, strrep("=", nchar(scenario_name)))

```

```

)

data <- datasets[[scenario_name]]
current <- data %>% filter(study_type == "current")
historical <- data %>% filter(study_type == "historical")

scenario_results <- list()

# =====
==
# OUTCOME 1: RESPONSE PROPORTION
# =====
==

cat("\n--- OUTCOME 1: Response Proportion ---\n")

hist_resp <- historical %>%
  mutate(
    responders = round(observed_proportion * n),
    nonresponders = n - responders,
    study_id = factor(seq_len(n()))
  )

map_fit_1 <- gMAP(
  cbind(responders, nonresponders) ~ 1 | study_id,
  data = hist_resp,
  family = binomial,
  tau.dist = "HalfNormal",
  tau.prior = 1,
  beta.prior = 2
)

map_prior_1 <- automixfit(map_fit_1)
hist_post_1 <- extract_historical_posterior_from_map(
  map_prior = map_prior_1,
  outcome_type = "binary"
)

p_c_adj <- adjust_prop(current$observed_proportion, current$n)
theta_c_hat_1 <- logit(p_c_adj)
var_theta_c_hat_1 <- 1 / (current$n * p_c_adj * (1 - p_c_adj))

cat(sprintf(" Step 1:  $\mu_h = %.3f$ ,  $\sigma_h = %.3f$ \n",
  hist_post_1$mu_h_hat, hist_post_1$sigma_h_hat))

discount_1 <- calculate_discount(
  theta_c_hat = theta_c_hat_1,
  var_theta_c_hat = var_theta_c_hat_1,
  mu_h_hat = hist_post_1$mu_h_hat,

```

```

    sigma_h_sq_hat = hist_post_1$sigma_h_sq_hat,
    a = a_param,
    b = b_param
  )

  cat(sprintf(" Step 2: T = %.3f, g(T) = %.3f\n",
              discount_1$T, discount_1$g_T))

  elastic_post_1 <- compute_elastic_posterior(
    theta_c_hat = theta_c_hat_1,
    var_theta_c_hat = var_theta_c_hat_1,
    mu_h_hat = hist_post_1$mu_h_hat,
    sigma_h_sq_hat = hist_post_1$sigma_h_sq_hat,
    g_T = discount_1$g_T
  )

  cat(sprintf(" Step 3:  $\sigma^2_h / g(T) = %.3f$ \n", elastic_post_1$sigma_Elastic_sq))

  elastic_mean_prob <- inv_logit(elastic_post_1$mean)
  elastic_ci_prob <- inv_logit(c(elastic_post_1$ci_lower, elastic_post_1$ci_upper))

  cat(sprintf(" Posterior: %.3f [%.3f, %.3f]\n",
              elastic_mean_prob, elastic_ci_prob[1], elastic_ci_prob[2]))

  scenario_results$outcome1 <- list(
    map_fit = map_fit_1,
    map_prior = map_prior_1,
    hist_post = hist_post_1,
    discount = discount_1,
    elastic_post = elastic_post_1,
    mean_transformed = elastic_mean_prob,
    ci_lower_transformed = elastic_ci_prob[1],
    ci_upper_transformed = elastic_ci_prob[2]
  )

  # =====
  ==
  # OUTCOME 2: EVENT RATE
  # =====
  ==

  cat("\n--- OUTCOME 2: Event Rate ---\n")

  historical_2 <- historical %>%
    mutate(
      log_rate = log(observed_event_rate + 0.001),
      se_log_rate = 1 / sqrt(pmax(events, 1)),

```

```

    study_id = factor(seq_len(n()))
  )

current_2 <- current %>%
  mutate(
    log_rate = log(observed_event_rate + 0.001),
    se_log_rate = 1 / sqrt(pmax(events, 1))
  )

map_fit_2 <- gMAP(
  log_rate ~ 1 | study_id,
  weights = se_log_rate^(-2),
  data = historical_2,
  family = gaussian,
  tau.dist = "HalfNormal",
  tau.prior = 0.5,
  beta.prior = 2
)

map_prior_2 <- automixfit(map_fit_2)
hist_post_2 <- extract_historical_posterior_from_map(
  map_prior = map_prior_2,
  outcome_type = "poisson"
)

theta_c_hat_2 <- current_2$log_rate
var_theta_c_hat_2 <- current_2$se_log_rate^2

cat(sprintf(" Step 1:  $\mu_h = %.3f$ ,  $\sigma_h = %.3f$ \n",
  hist_post_2$mu_h_hat, hist_post_2$sigma_h_hat))

discount_2 <- calculate_discount(
  theta_c_hat = theta_c_hat_2,
  var_theta_c_hat = var_theta_c_hat_2,
  mu_h_hat = hist_post_2$mu_h_hat,
  sigma_h_sq_hat = hist_post_2$sigma_h_sq_hat,
  a = a_param,
  b = b_param
)

cat(sprintf(" Step 2: T = %.3f, g(T) = %.3f\n",
  discount_2$T, discount_2$g_T))

elastic_post_2 <- compute_elastic_posterior(
  theta_c_hat = theta_c_hat_2,
  var_theta_c_hat = var_theta_c_hat_2,
  mu_h_hat = hist_post_2$mu_h_hat,
  sigma_h_sq_hat = hist_post_2$sigma_h_sq_hat,
  g_T = discount_2$g_T
)

```

```

)

elastic_mean_rate <- exp(elastic_post_2$mean)
elastic_ci_rate <- exp(c(elastic_post_2$ci_lower, elastic_post_2$ci_upper))

cat(sprintf(" Posterior: %.3f [%.3f, %.3f]\n",
            elastic_mean_rate, elastic_ci_rate[1], elastic_ci_rate[2]))

scenario_results$outcome2 <- list(
  map_fit = map_fit_2,
  map_prior = map_prior_2,
  hist_post = hist_post_2,
  discount = discount_2,
  elastic_post = elastic_post_2,
  mean_transformed = elastic_mean_rate,
  ci_lower_transformed = elastic_ci_rate[1],
  ci_upper_transformed = elastic_ci_rate[2]
)

# =====
==
# OUTCOME 3: CONTINUOUS MEAN
# =====
==

cat("\n--- OUTCOME 3: Continuous Mean ---\n")

historical_3 <- historical %>%
  mutate(
    se_mean = observed_sd / sqrt(n),
    study_id = factor(seq_len(n()))
  )

current_3 <- current %>%
  mutate(
    se_mean = observed_sd / sqrt(n)
  )

map_fit_3 <- gMAP(
  observed_mean ~ 1 | study_id,
  weights = se_mean^(-2),
  data = historical_3,
  family = gaussian,
  tau.dist = "HalfNormal",
  tau.prior = 10,
  beta.prior = 2
)

map_prior_3 <- automixfit(map_fit_3)

```

```

hist_post_3 <- extract_historical_posterior_from_map(
  map_prior = map_prior_3,
  outcome_type = "continuous"
)

theta_c_hat_3 <- current_3$observed_mean
var_theta_c_hat_3 <- current_3$se_mean^2

cat(sprintf(" Step 1:  $\mu^h = %.2f$ ,  $\sigma^h = %.2f$ \n",
  hist_post_3$mu_h_hat, hist_post_3$sigma_h_hat))

discount_3 <- calculate_discount(
  theta_c_hat = theta_c_hat_3,
  var_theta_c_hat = var_theta_c_hat_3,
  mu_h_hat = hist_post_3$mu_h_hat,
  sigma_h_sq_hat = hist_post_3$sigma_h_sq_hat,
  a = a_param,
  b = b_param
)

cat(sprintf(" Step 2: T = %.3f, g(T) = %.3f\n",
  discount_3$T, discount_3$g_T))

elastic_post_3 <- compute_elastic_posterior(
  theta_c_hat = theta_c_hat_3,
  var_theta_c_hat = var_theta_c_hat_3,
  mu_h_hat = hist_post_3$mu_h_hat,
  sigma_h_sq_hat = hist_post_3$sigma_h_sq_hat,
  g_T = discount_3$g_T
)

cat(sprintf(" Posterior: %.2f [%.2f, %.2f]\n",
  elastic_post_3$mean, elastic_post_3$ci_lower, elastic_post_3$ci_upper))

scenario_results$outcome3 <- list(
  map_fit = map_fit_3,
  map_prior = map_prior_3,
  hist_post = hist_post_3,
  discount = discount_3,
  elastic_post = elastic_post_3,
  mean_transformed = elastic_post_3$mean,
  ci_lower_transformed = elastic_post_3$ci_lower,
  ci_upper_transformed = elastic_post_3$ci_upper
)

all_results[[scenario_name]] <- scenario_results
}

```

```
# =====
==
# CREATE SUMMARY TABLE
# =====
==

summary_rows <- list()
row_idx <- 1

for (scenario_name in names(all_results)) {
  res <- all_results[[scenario_name]]

  for (outcome_info in list(
    list(name = "Response Proportion", key = "outcome1"),
    list(name = "Event Rate", key = "outcome2"),
    list(name = "Continuous Mean", key = "outcome3")
  )) {
    outcome_res <- res[[outcome_info$key]]

    summary_rows[[row_idx]] <- data.frame(
      Scenario = scenario_name,
      Outcome = outcome_info$name,
      mu_h_hat = round(outcome_res$hist_post$mu_h_hat, 3),
      sigma_h_hat = round(outcome_res$hist_post$sigma_h_hat, 3),
      sigma_h_sq_hat = round(outcome_res$hist_post$sigma_h_sq_hat, 3),
      T = round(outcome_res$discount$T, 3),
      g_T = round(outcome_res$discount$g_T, 3),
      sigma_Elastic = round(outcome_res$elastic_post$sigma_Elastic, 3),
      sigma_Elastic_sq = round(outcome_res$elastic_post$sigma_Elastic_sq, 3),
      Posterior_Mean = round(outcome_res$mean_transformed, 3),
      CI_Lower = round(outcome_res$ci_lower_transformed, 3),
      CI_Upper = round(outcome_res$ci_upper_transformed, 3)
    )

    row_idx <- row_idx + 1
  }
}

summary_table <- do.call(rbind, summary_rows)
print(summary_table)

write.csv(
  summary_table,
  "C:/Users/marin/Downloads/R scripts/Elastic/elastic_prior_results.csv",
  row.names = FALSE
)

outcome1 <- scenario_results$outcome1
outcome2 <- scenario_results$outcome2
```

```

outcome3 <- scenario_results$outcome3

save(
  outcome1, outcome2, outcome3,
  file = "C:/Users/marin/Downloads/R scripts/Elastic/elastic_results.RData"
)

# BETWEEN-STUDY MODEL:
#  $\theta_j \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ 
#
# Where:
#  $\mu$  = overall average effect across studies
#  $\sigma^2$  = between-study heterogeneity variance
#  $\sigma$  = between-study heterogeneity standard deviation
#
# PRIORS:
#  $\mu \sim N(\mu_0, \sigma_\mu^2)$  [weakly informative normal prior]
#  $\sigma \sim \text{Half-Normal}(0, \sigma_\sigma)$  or  $\text{Half-Cauchy}(0, \sigma_\sigma)$ 
#
# CONDITIONAL POSTERIOR MEAN (given  $\mu$  and  $\sigma^2$ ):
#  $E[\theta_j \mid \hat{\theta}_j, \mu, \sigma^2] = w_j \times \hat{\theta}_j + (1 - w_j) \times \mu$ 
#
# Where the shrinkage weight is:
#  $w_j = (1/\sigma_j^2) / (1/\sigma_j^2 + 1/\sigma^2)$ 
#  $= \sigma^2 / (\sigma^2 + \sigma_j^2)$ 
#
# Interpretation:
# - Small  $\sigma^2 \rightarrow w_j \approx 0 \rightarrow$  strong shrinkage toward  $\mu$  (strong borrowing)
# - Large  $\sigma^2 \rightarrow w_j \approx 1 \rightarrow$  weak shrinkage (each study uses own data)
# - Large  $\sigma_j^2 \rightarrow$  stronger shrinkage toward  $\mu$  (imprecise studies borrow more)
#
# Note:
# The weighted-average formula above is exact conditionally on  $\mu$  and  $\sigma^2$ .
# In the full Bayesian model below, posterior summaries integrate over
# uncertainty in both  $\mu$  and  $\sigma$ .

library(dplyr)

set.seed(2024)

# =====
# NNHM FUNCTIONS WITH PROPER BAYESIAN INFERENCE
# =====

```

```

==
#' NNHM: Normal-Normal Hierarchical Model with MCMC
#'
#' Implements the full Bayesian hierarchical model:
#'   Within-study:  $\theta_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$ 
#'   Between-study:  $\theta_j \mid \mu, \sigma \sim N(\mu, \sigma^2)$ 
#'   Prior on  $\mu$ :  $\mu \sim N(\mu_0, \sigma_\mu^2)$ 
#'   Prior on  $\sigma$ :  $\sigma \sim \text{Half-Normal}(\theta, \sigma_\sigma)$ 
#'
#' Uses Gibbs sampling for posterior inference.

NNHM_posterior <- function(theta_hat_j, sigma_j,
                           mu_0 = 0, sigma_mu = 10, sigma_sigma = 1,
                           n_iter = 10000, n_burnin = 2000,
                           current_idx = 1) {

  J <- length(theta_hat_j) # Total number of studies (current + historical)
  sigma_j_sq <- sigma_j^2 # Within-study variances  $\sigma_j^2$ 

  # =====
=
  # Initialize parameters
  # =====
=

  # Initialize  $\mu$  (overall mean)
  mu <- mean(theta_hat_j)

  # Initialize  $\sigma^2$  (between-study variance)
  sigma_sq <- var(theta_hat_j)
  sigma_sq <- max(sigma_sq, 0.01)

  # Initialize  $\theta_j$  (study-specific effects)
  theta_j <- theta_hat_j

  # Storage for posterior samples
  mu_samples <- numeric(n_iter)
  sigma_samples <- numeric(n_iter)
  theta_j_samples <- matrix(0, nrow = n_iter, ncol = J)

  # =====
=
  # Gibbs Sampling
  # =====
=

  for (iter in 1:n_iter) {

```

```

# -----
# Step 1: Sample  $\theta_j \mid \mu, \sigma^2, \hat{\theta}_j, \sigma_j^2$  for each study j
# -----
# Posterior for  $\theta_j$  is Normal with:
#   Precision:  $1/\sigma_j^2 + 1/\sigma^2$ 
#   Mean:  $(\hat{\theta}_j/\sigma_j^2 + \mu/\sigma^2) / (1/\sigma_j^2 + 1/\sigma^2)$ 

for (j in 1:J) {
  post_precision_j <- 1 / sigma_j_sq[j] + 1 / sigma_sq
  post_var_j <- 1 / post_precision_j
  post_mean_j <- post_var_j * (theta_hat_j[j] / sigma_j_sq[j] + mu / sigma_sq)
  theta_j[j] <- rnorm(1, post_mean_j, sqrt(post_var_j))
}

# -----
# Step 2: Sample  $\mu \mid \theta_j, \sigma^2$ 
# -----
# Prior:  $\mu \sim N(\mu_0, \sigma_\mu^2)$ 
# Likelihood:  $\theta_j \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ 
#
# Posterior precision:  $J/\sigma^2 + 1/\sigma_\mu^2$ 
# Posterior mean:  $(\sum(\theta_j)/\sigma^2 + \mu_0/\sigma_\mu^2) / (J/\sigma^2 + 1/\sigma_\mu^2)$ 

post_precision_mu <- J / sigma_sq + 1 / sigma_mu^2
post_var_mu <- 1 / post_precision_mu
post_mean_mu <- post_var_mu * (sum(theta_j) / sigma_sq + mu_0 / sigma_mu^2)
mu <- rnorm(1, post_mean_mu, sqrt(post_var_mu))

# -----
# Step 3: Sample  $\sigma^2 \mid \theta_j, \mu$  using Metropolis-Hastings
# -----
# Prior:  $\sigma \sim \text{Half-Normal}(0, \sigma_\sigma)$ , so  $\sigma^2$  has corresponding distribution
# Likelihood:  $\theta_j \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ 

# Propose new  $\sigma^2$  (log-normal random walk)
sigma_sq_proposed <- exp(log(sigma_sq) + rnorm(1, 0, 0.3))

# Log-posterior for current  $\sigma^2$ 
log_post_current <- sum(dnorm(theta_j, mu, sqrt(sigma_sq), log = TRUE)) +
  dnorm(sqrt(sigma_sq), 0, sigma_sigma, log = TRUE) -
  0.5 * log(sigma_sq) # Jacobian for  $\sigma \rightarrow \sigma^2$ 

# Log-posterior for proposed  $\sigma^2$ 
log_post_proposed <- sum(dnorm(theta_j, mu, sqrt(sigma_sq_proposed), log = TRUE)) +

```

```

dnorm(sqrt(sigma_sq_proposed), 0, sigma_sigma, log = TRUE) -
0.5 * log(sigma_sq_proposed)

# Acceptance probability (with proposal ratio for log-normal RW)
log_accept <- log_post_proposed - log_post_current +
  log(sigma_sq_proposed) - log(sigma_sq)

if (log(runif(1)) < log_accept) {
  sigma_sq <- sigma_sq_proposed
}

# Store samples
mu_samples[iter] <- mu
sigma_samples[iter] <- sqrt(sigma_sq)
theta_j_samples[iter, ] <- theta_j
}

# =====
=
# Discard burn-in and compute summaries
# =====
=

keep_idx <- (n_burnin + 1):n_iter
mu_post <- mu_samples[keep_idx]
sigma_post <- sigma_samples[keep_idx]
theta_j_post <- theta_j_samples[keep_idx, ]

# Posterior summaries for current study (theta_c)
theta_c_samples <- theta_j_post[, current_idx]

# Compute shrinkage weights using posterior mean of sigma
sigma_hat <- mean(sigma_post)
sigma_sq_hat <- sigma_hat^2
w_j <- sigma_sq_hat / (sigma_sq_hat + sigma_j_sq) # Shrinkage weights

# Borrowing weight for current study (1 - w_j = weight on overall mean)
borrowing_weight_c <- 1 - w_j[current_idx]

return(list(
  # Current study posterior
  theta_c_mean = mean(theta_c_samples),
  theta_c_sd = sd(theta_c_samples),
  theta_c_ci_lower = quantile(theta_c_samples, 0.025),
  theta_c_ci_upper = quantile(theta_c_samples, 0.975),
  theta_c_samples = theta_c_samples,

  # Hyperparameter posteriors
  mu_mean = mean(mu_post),

```

```

mu_sd = sd(mu_post),
mu_ci_lower = quantile(mu_post, 0.025),
mu_ci_upper = quantile(mu_post, 0.975),

sigma_mean = mean(sigma_post),
sigma_sd = sd(sigma_post),
sigma_ci_lower = quantile(sigma_post, 0.025),
sigma_ci_upper = quantile(sigma_post, 0.975),

# Shrinkage weights
w_j = w_j,
w_c = w_j[current_idx],          # Weight on current study's own d
ata
    borrowing_weight = borrowing_weight_c, # Weight on overall mean (1 - w_c)

# All study posteriors
theta_j_means = colMeans(theta_j_post),
theta_j_sds = apply(theta_j_post, 2, sd),

# Full samples
mu_samples = mu_post,
sigma_samples = sigma_post,
theta_j_samples = theta_j_post
))
}

for (scenario_name in names(datasets)) {

  cat(sprintf("\n%s\n%s\n", scenario_name, strrep("=", nchar(scenario_name)))
  )

  data <- datasets[[scenario_name]]
  current <- data %>% filter(study_type == "current")
  historical <- data %>% filter(study_type == "historical")

  # Combine current (first) + historical studies
  all_studies <- rbind(current, historical)
  J <- nrow(all_studies)

  scenario_results <- list()

  # =====
  # OUTCOME 1: RESPONSE PROPORTION (logit scale)
  # =====

  # Transform to logit scale

```

```

props <- all_studies$observed_proportion
ns <- all_studies$n

theta_hat_j <- log(props / (1 - props))           #  $\hat{\theta}_j$  on logit scale
sigma_j <- 1 / sqrt(ns * props * (1 - props))     # sigma_j (within-study SD)

# Prior specification
mu_0_1 <- 0
sigma_mu_1 <- 3
sigma_sigma_1 <- 1

# Run NNHM
nnhm_1 <- NNHM_posterior(
  theta_hat_j, sigma_j,
  mu_0 = mu_0_1, sigma_mu = sigma_mu_1,
  sigma_sigma = sigma_sigma_1,
  n_iter = 15000, n_burnin = 3000,
  current_idx = 1
)

# Transform back to probability
theta_c_prob <- 1 / (1 + exp(-nnhm_1$theta_c_samples))
nnhm_mean_prob <- mean(theta_c_prob)
nnhm_ci_prob <- quantile(theta_c_prob, c(0.025, 0.975))

scenario_results$outcome1 <- list(
  nnhm = nnhm_1,
  mean_transformed = nnhm_mean_prob,
  ci_lower_transformed = nnhm_ci_prob[1],
  ci_upper_transformed = nnhm_ci_prob[2]
)

# =====
# OUTCOME 2: EVENT RATE (log scale)
# =====

rates <- all_studies$observed_event_rate
events <- all_studies$events

theta_hat_j_2 <- log(rates + 0.001)               #  $\hat{\theta}_j$  on log scale
sigma_j_2 <- 1 / sqrt(pmax(events, 1))           # sigma_j

```

```

mu_0_2 <- log(0.25)
sigma_mu_2 <- 2
sigma_sigma_2 <- 1

nnhm_2 <- NNHM_posterior(
  theta_hat_j_2, sigma_j_2,
  mu_0 = mu_0_2, sigma_mu = sigma_mu_2,
  sigma_sigma = sigma_sigma_2,
  n_iter = 15000, n_burnin = 3000,
  current_idx = 1
)

# Transform back to rate
theta_c_rate <- exp(nnhm_2$theta_c_samples)
nnhm_mean_rate <- mean(theta_c_rate)
nnhm_ci_rate <- quantile(theta_c_rate, c(0.025, 0.975))

scenario_results$outcome2 <- list(
  nnhm = nnhm_2,
  mean_transformed = nnhm_mean_rate,
  ci_lower_transformed = nnhm_ci_rate[1],
  ci_upper_transformed = nnhm_ci_rate[2]
)

# =====
# OUTCOME 3: CONTINUOUS MEAN
# =====

means <- all_studies$observed_mean
sds <- all_studies$observed_sd
ns <- all_studies$n

theta_hat_j_3 <- means
sigma_j_3 <- sds / sqrt(ns)           # sigma_j = SD / sqrt(n)

mu_0_3 <- 25
sigma_mu_3 <- 20
sigma_sigma_3 <- 10

nnhm_3 <- NNHM_posterior(
  theta_hat_j_3, sigma_j_3,
  mu_0 = mu_0_3, sigma_mu = sigma_mu_3,
  sigma_sigma = sigma_sigma_3,
  n_iter = 15000, n_burnin = 3000,
  current_idx = 1
)

```

```

scenario_results$outcome3 <- list(
  nnhm = nnhm_3,
  mean_transformed = nnhm_3$theta_c_mean,
  ci_lower_transformed = nnhm_3$theta_c_ci_lower,
  ci_upper_transformed = nnhm_3$theta_c_ci_upper
)

all_results[[scenario_name]] <- scenario_results
}

# =====
#
# CREATE SUMMARY TABLE
# =====

summary_rows <- list()
row_idx <- 1

for (scenario_name in names(all_results)) {
  res <- all_results[[scenario_name]]

  for (outcome_info in list(
    list(name = "Response Proportion", key = "outcome1"),
    list(name = "Event Rate", key = "outcome2"),
    list(name = "Continuous Mean", key = "outcome3")
  )) {
    outcome_res <- res[[outcome_info$key]]
    nnhm <- outcome_res$nnhm

    summary_rows[[row_idx]] <- data.frame(
      Scenario = scenario_name,
      Outcome = outcome_info$name,
      mu_posterior = round(nnhm$mu_mean, 3),
      sigma_posterior = round(nnhm$sigma_mean, 3),
      w_c = round(nnhm$w_c, 3),
      borrowing_weight = round(nnhm$borrowing_weight, 3),
      theta_c_mean = round(outcome_res$mean_transformed, 3),
      CI_Lower = round(outcome_res$ci_lower_transformed, 3),
      CI_Upper = round(outcome_res$ci_upper_transformed, 3)
    )
    row_idx <- row_idx + 1
  }
}

summary_table <- do.call(rbind, summary_rows)
print(summary_table)

```

```
write.csv(summary_table, "C:/Users/marin/Downloads/R scripts/NNHM/nnhm_results.csv", row.names = FALSE)

outcome1 <- scenario_results$outcome1
outcome2 <- scenario_results$outcome2
outcome3 <- scenario_results$outcome3

save(outcome1, outcome2, outcome3,
      file = "C:/Users/marin/Downloads/R scripts/NNHM/NNHM_results.RData")
```

12.3 pre borrowing weights

```
-----
# STEP 1: LOAD REQUIRED PACKAGES
# -----
--

library(cluster)
library(dplyr)
library(readr)

# -----
--
# STEP 2: LOAD YOUR DATA
# -----
--

data <- read.csv(
  "C:/Users/marin/Downloads/Datasets/demographics.csv",
  fileEncoding = "latin1",
  stringsAsFactors = FALSE,
  na.strings = c("", "NA")
)

# =====
==
# PART A: RISK OF BIAS (ROB) WEIGHTS
# =====
==
#
# WHY? Studies with high risk of bias may give misleading results.
#       We want to DOWNWEIGHT biased studies so they contribute less.
#
# HOW? We assign a weight between 0 and 1 based on the ROB rating:
#       - "Good" quality = 1.0
#       - "Fair" quality = 0.7
```

```
# - "Poor" quality = 0.3
#
# -----
--

calculate_rob_weights <- function(rob_ratings) {
  # Create a named vector of weights for each ROB category
  rob_weight_map <- c(
    "Good" = 1.0,    # No bias concerns - full weight
    "Fair" = 0.7,    # Some bias concerns - reduced weight
    "Poor" = 0.3     # Serious bias concerns - heavily downweighted
  )

  # Apply the mapping to each study's ROB rating
  weights <- rob_weight_map[rob_ratings]

  # If any ROB rating is not recognized, assign 0.5 (uncertain)
  weights[is.na(weights)] <- 0.5

  return(weights)
}

# Calculate ROB weights for all studies
rob_weights <- calculate_rob_weights(data$ROB)

cat("=== RISK OF BIAS WEIGHTS ===\n")
cat("Logic: Good=1.0, Fair=0.7, Poor=0.3\n\n")
for (i in 1:nrow(data)) {
  cat(sprintf("%-45s | ROB: %-4s | Weight: %.1f\n",
              substr(data$Studies[i], 1, 45),
              data$ROB[i],
              rob_weights[i]))
}

# =====
==
# PART B: STUDY DESIGN WEIGHTS
# =====
==
#
# WHY? Different study designs have different levels of evidence quality.
#       Randomized trials are generally more reliable than case series.
#
# HOW? We assign weights based on the evidence hierarchy:
#       - Randomized Trial = 1.0 (highest quality)
#       - Prospective observational = 0.85
#       - Systematic review = 0.80
#       - Multicenter study = 0.75
#       - Retrospective case-control = 0.70
```

```
# - Other observational = 0.60
# - Case series = 0.50 (lowest quality)
#
# -----
--

calculate_design_weights <- function(designs) {

  weights <- sapply(designs, function(design) {
    design_lower <- tolower(design)

    # Check design type in order of specificity
    if (grepl("randomised|randomized|rct", design_lower)) {
      return(1.0) # Randomized trials - highest weight
    } else if (grepl("prospective|trial", design_lower) && !grepl("retrospect
ive", design_lower)) {
      return(0.85) # Prospective only
    } else if (grepl("systematic review", design_lower)) {
      return(0.80) # Systematic reviews
    } else if (grepl("multicenter", design_lower)) {
      return(0.75) # Multicenter studies
    } else if (grepl("retrospective case control|case control", design_lower)
) {
      return(0.70) # Case-control studies
    } else if (grepl("observational", design_lower)) {
      return(0.60) # Other observational studies
    } else if (grepl("case series", design_lower)) {
      return(0.50) # Case series - lowest weight
    } else {
      return(0.50) # Unknown - conservative weight
    }
  })

  return(weights)
}

# Calculate design weights for all studies
design_weights <- calculate_design_weights(data$Design)

# =====
==
# PART C: DEMOGRAPHIC SIMILARITY WEIGHTS (GOWER DISTANCE WITHOUT IMPUTATION)
# =====
==
#
# For each historical study, similarity is calculated using only the variable
s
# reported in both the current study and the historical study.
```

```
#
# A completeness penalty is added:
#   Completeness_j = number of available variables / total number of planned
#   variables
#
# Adjusted similarity:
#   S_adjusted_j = S_j * Completeness_j
#

# =====
==

calculate_demographic_similarity <- function(data) {

  n_studies <- nrow(data)
  n_historical <- n_studies - 1

  cat("Current study:", data$Studies[1], "\n")
  cat("Historical studies:", n_historical, "\n\n")

  # =====
  # SELECT KEY DEMOGRAPHIC / CLINICAL VARIABLES
  # =====

  demo_data <- data.frame(
    Total = data$Total,
    Age = data$Age_median,
    Duration = data$Median.duration.of.treatment.in.months,
    Male_prop = data$Males / data$Total
  )

  if ("Antiplatelet" %in% names(data)) {
    demo_data$Antiplatelet_prop <- data$Antiplatelet / data$Total
  }
  if ("Oral..venous.anticoagulant" %in% names(data)) {
    demo_data$Anticoagulant_prop <- data$Oral..venous.anticoagulant / data$Total
  }
  if ("Antineoplastic..vincristine." %in% names(data)) {
    demo_data$Antineoplastic_prop <- data$Antineoplastic..vincristine. / data$Total
  }
  if ("Antibiotic.antibiotherapy" %in% names(data)) {
    demo_data$Antibiotic_prop <- data$Antibiotic.antibiotherapy / data$Total
  }
  if ("Steroids" %in% names(data)) {
    demo_data$Steroids_prop <- data$Steroids / data$Total
  }
  if ("Surgery" %in% names(data)) {
```

```

demo_data$Surgery_prop <- data$Surgery / data$Total
}
if ("Sclerotherapy.embolization" %in% names(data)) {
  demo_data$Sclero_prop <- data$Sclerotherapy.embolization / data$Total
}
if ("Venous.malformation" %in% names(data)) {
  demo_data$VM_prop <- data$Venous.malformation / data$Total
}
if ("Capillary.malformation" %in% names(data)) {
  demo_data$Capillary_prop <- data$Capillary.malformation / data$Total
}
if ("Lymphatic.malformations" %in% names(data)) {
  demo_data$LM_prop <- data$Lymphatic.malformations / data$Total
}
if ("Combined" %in% names(data)) {
  demo_data$Combined_prop <- data$Combined / data$Total
}
if ("Cervico.facial" %in% names(data)) {
  demo_data$Cervicofacial_prop <- data$Cervico.facial / data$Total
}
if ("Trunk" %in% names(data)) {
  demo_data$Trunk_prop <- data$Trunk / data$Total
}
if ("Limb" %in% names(data)) {
  demo_data$Limb_prop <- data$Limb / data$Total
}
if ("Cervicofacial.and.trunk" %in% names(data)) {
  demo_data$Cervi_trunk_prop <- data$Cervicofacial.and.trunk / data$Total
}
if ("Trunk.and.limb" %in% names(data)) {
  demo_data$Trunk_limb_prop <- data$Trunk.and.limb / data$Total
}
if ("Cervicofacial..trunk.and.limb" %in% names(data)) {
  demo_data$Cervi_trunk_limb_prop <- data$Cervicofacial..trunk.and.limb / d
ata$Total
}
if ("Internal" %in% names(data)) {
  demo_data$Internal_prop <- data$Internal / data$Total
}
if ("Pain" %in% names(data)) {
  demo_data$Pain_prop <- data$Pain / data$Total
}
if ("Functional.limitation" %in% names(data)) {
  demo_data$Functional_limitation_prop <- data$Functional.limitation / data
$Total
}
if ("Bleeding" %in% names(data)) {
  demo_data$Bleeding_prop <- data$Bleeding / data$Total
}
if ("Cellulitis" %in% names(data)) {

```

```

demo_data$Cellulitis_prop <- data$Cellulitis / data$Total
}
if ("Coagulopathy" %in% names(data)) {
  demo_data$Coagulopathy_prop <- data$Coagulopathy / data$Total
}
if ("Infection" %in% names(data)) {
  demo_data$Infection_prop <- data$Infection / data$Total
}
if ("Oozing" %in% names(data)) {
  demo_data$Oozing_prop <- data$Oozing / data$Total
}
if ("Ulceration" %in% names(data)) {
  demo_data$Ulceration_prop <- data$Ulceration / data$Total
}
if ("Problems.with.the.speech.tongue" %in% names(data)) {
  demo_data$Speech_tongue_prop <- data$Problems.with.the.speech.tongue / da
ta$Total
}
if ("Upper.airway.obstruction" %in% names(data)) {
  demo_data$Upper_airway_obstruction_prop <- data$Upper.airway.obstruction
/ data$Total
}
if ("Facial.deformity" %in% names(data)) {
  demo_data$Facial_deformity_prop <- data$Facial.deformity / data$Total
}
if ("Swelling" %in% names(data)) {
  demo_data$Swelling_prop <- data$Swelling / data$Total
}

# =====
# CALCULATE GOWER DISTANCE WITH MISSING VALUES
# =====
# daisy() handles missing values by excluding unavailable variables
# from each pairwise distance calculation.

gower_dist <- daisy(demo_data, metric = "gower")
dist_matrix <- as.matrix(gower_dist)

# Similarity between current study and each historical study
raw_similarity <- 1 - dist_matrix[2:n_studies, 1]

# =====
# COMPLETENESS PENALTY
# =====
# Count variables available in both current and historical study.

total_variables <- ncol(demo_data)

completeness <- numeric(n_historical)

```

```

for (j in 2:n_studies) {
  available_vars <- !is.na(demo_data[1, ]) & !is.na(demo_data[j, ])
  completeness[j - 1] <- sum(available_vars) / total_variables
}

adjusted_similarity <- raw_similarity * completeness

cat("Similarity calculated without imputation.\n")
cat("Completeness penalty applied.\n\n")

return(list(
  historical_studies = data$Studies[2:n_studies],
  raw_similarity = raw_similarity,
  completeness = completeness,
  adjusted_similarity = adjusted_similarity,
  variables_used = names(demo_data),
  total_variables = total_variables
))
}

demo_result <- calculate_demographic_similarity(data)

# Use adjusted similarity for the primary analysis
demographic_similarity <- demo_result$adjusted_similarity

# Optional: keep raw similarity and completeness separately
raw_demographic_similarity <- demo_result$raw_similarity
completeness_weight <- demo_result$completeness

rob_weights <- calculate_rob_weights(data$ROB[2:nrow(data)])
design_weights <- calculate_design_weights(data$Design[2:nrow(data)])

# =====
#
# PART D: COMBINING THE WEIGHTS
# =====
#

# Define importance weights for the additive method
w_demo <- 0.40
w_rob <- 0.35
w_design <- 0.25

# METHOD 1: Weighted Average
combined_additive <- w_demo * demographic_similarity +
  w_rob * rob_weights +

```

```

w_design * design_weights

# METHOD 2: Multiplicative
combined_multiplicative <- demographic_similarity * rob_weights * design_weights

# METHOD 3: Quality-Adjusted Similarity
quality_adjusted <- demographic_similarity * sqrt(rob_weights * design_weights)

# =====
==
# CREATE RESULTS TABLE WITH ALL 20 STUDIES
# Row 1 = Current study (weights = NA)
# Rows 2-20 = Historical studies (weights calculated)
# =====
==

results <- data.frame(
  Study = data$Studies,
  Raw_Demographic_Similarity = c(NA, round(raw_demographic_similarity, 3)),
  Completeness_Weight = c(NA, round(completeness_weight, 3)),
  Adjusted_Demographic_Similarity = c(NA, round(demographic_similarity, 3)),
  ROB_Weight = c(NA, round(rob_weights, 3)),
  Design_Weight = c(NA, round(design_weights, 3)),
  Combined_Additive = c(NA, round(combined_additive, 3)),
  Combined_Multiplicative = c(NA, round(combined_multiplicative, 3)),
  Quality_Adjusted = c(NA, round(quality_adjusted, 3)),
  ROB = data$ROB,
  Design = data$Design
)

# =====
==
# PART E: NORMALIZE WEIGHTS FOR HISTORICAL STUDIES
# =====
==

# Get weights for historical studies only (rows 2:20)
historical_weights <- results$Quality_Adjusted[2:nrow(results)]

# Normalize so they sum to 1
normalized_weights <- historical_weights / sum(historical_weights)

# Add to results (NA for current study, normalized for historical)
results$Normalized_Weight <- c(NA, round(normalized_weights, 4))

# =====

```

```

==
# PART F: RANK HISTORICAL STUDIES
# =====
==

cat("\n\n=== HISTORICAL STUDIES RANKED BY WEIGHT ===\n\n")

# Sort historical studies by quality-adjusted weight (descending)
historical_results <- results[2:nrow(results), ]
historical_results <- historical_results[order(-historical_results$Quality_Adjusted), ]
historical_results$Rank <- 1:nrow(historical_results)

# Print ranked results
print(historical_results[, c("Rank", "Study", "Quality_Adjusted",
                           "Normalized_Weight", "ROB", "Design")])

# =====
==
# PART G: SAVE RESULTS
# =====
==

results$Study_type<- data$Study_type
results$N_pain<- data$N_pain
results$Response_pain <- data$Response_pain

results$N_d_dimer <-data$N_d_dimer
results$Mean_decrease <-data$Mean_decrease
results$SD_decrease<-data$SD_decrease

# Save complete results to CSV
write.csv(results, "C:/Users/marin/Downloads/Datasets/calculated_weights.csv"
, row.names = FALSE)

```

12.4 RWD analysis (Binary endpoint)

```

# =====
==
# BAYESIAN BORROWING ANALYSIS - QUALITY-ADJUSTED THEORY-CONSISTENT IMPLEMENTATION
# =====
==
#
# METHODS REPORTED:
# 1. rMAP - Robust MAP prior
# 2. SAM - Self-Adapting Mixture Prior
# 3. Elastic Prior

```

```
# 4. NNHM - Normal-Normal Hierarchical Model
#
# INTERNAL METHOD:
# - MAP prior is fitted internally because it is required to construct:
#   (a) rMAP
#   (b) SAM for multiple historical studies
#   (c) Elastic prior
#
# QUALITY ADJUSTMENT:
#   q_j = Similarity_j × sqrt(ROB_j × Design_j)
#
# HOW QUALITY WEIGHTS ENTER:
#   MAP / rMAP / Elastic / SAM-informative-prior:
#   via quality-adjusted effective counts:
#     n_j,eff = round(n_j * q_j)
#     x_j,eff = round(x_j * q_j)
#
# NNHM:
#   via similarity-weighted second level:
#     theta_j | mu, sigma^2 ~ N(mu, sigma^2 / q_j)
#
# ADDITIONAL ESS:
#   For every borrowing method, the reported ESS is the posterior variance-based
#   additional ESS on the probability scale:
#
#     ESS_add = max(0, n_c * (Var_no_borrow / Var_method - 1))
#
# =====
==

library(RBest)
library(SAMprior)
library(tidyverse)

set.seed(2024)

# =====
==
# 1. LOAD DATA
# =====
==

data <- read.csv(
  "C:/Users/marin/Downloads/Datasets/calculated_weights.csv",
  stringsAsFactors = FALSE
)
```

```

current <- data %>%
  filter(Study_type == "current")

historical <- data %>%
  filter(Study_type == "historical") %>%
  filter(
    !is.na(N_pain),
    !is.na(Response_pain),
    !is.na(Quality_Adjusted)
  )

stopifnot(nrow(current) == 1)

current <- current %>%
  mutate(prop = Response_pain / N_pain)

historical <- historical %>%
  mutate(
    prop = Response_pain / N_pain,
    q_j = pmin(pmax(Quality_Adjusted, 1e-6), 1)
  )

x_c <- current$Response_pain
n_c <- current$N_pain
p_c <- current$prop

cat(sprintf(
  "Current study: %s | N=%d | Resp=%d | Rate=%.1f%%\n",
  current$Study, n_c, x_c, 100 * p_c
))
cat(sprintf(
  "Historical studies: %d | raw N=%d | raw Resp=%d\n\n",
  nrow(historical),
  sum(historical$N_pain),
  sum(historical$Response_pain)
))

# =====
#
# 2. SHARED HELPERS
# =====
#

adjust_prop <- function(prop, n) {
  out <- prop
  out[prop <= 0] <- 0.5 / (n[prop <= 0] + 1)
  out[prop >= 1] <- (n[prop >= 1] + 0.5) / (n[prop >= 1] + 1)
  out
}

```

```

logit <- function(p) log(p / (1 - p))
inv_logit <- function(x) 1 / (1 + exp(-x))

make_w_key <- function(w) sprintf("w_%.2f", w)
make_delta_key <- function(d) sprintf("Delta_%.4f", d)
make_sigma_key <- function(x) sprintf("sigma_sigma_%.3f", x)
make_el_key <- function(a, b) sprintf("a_%.2f_b_%.2f", a, b)

to_percent <- function(x) 100 * x

# -----
--
# Sample from beta mixture
# -----
--
rbeta_mix <- function(mix_obj, nsim = 100000) {
  mat <- as.matrix(t(mix_obj))
  if (is.null(dim(mat))) {
    mat <- matrix(mat, nrow = 1)
  }

  w <- mat[, 1]
  a <- mat[, 2]
  b <- mat[, 3]

  comp <- sample(seq_along(w), size = nsim, replace = TRUE, prob = w)
  rbeta(nsim, shape1 = a[comp], shape2 = b[comp])
}

# -----
--
# Posterior variance-based additional ESS
# -----
--
compute_additional_ess_variance <- function(var_method, var_no_borrow, n_current) {
  if (is.na(var_method) || is.na(var_no_borrow) || var_method <= 0 || var_no_borrow <= 0) {
    return(NA_real_)
  }
  max(0, n_current * (var_no_borrow / var_method - 1))
}

# =====
==
# 3. QUALITY-ADJUSTED HISTORICAL DATA FOR MAP / rMAP / SAM / ELASTIC
# =====
==

```

```

historical_eff <- historical %>%
  mutate(
    n_effective = pmax(round(N_pain * q_j), 1L),
    r_effective = round(Response_pain * q_j),
    r_effective = pmin(r_effective, n_effective),
    study_id = factor(seq_len(n()))
  )

cat(sprintf(
  "Quality-adjusted historical total N = %d\n\n",
  sum(historical_eff$n_effective)
))

# =====
#
# 4. INTERNAL MAP PRIOR
# =====

cat("Fitting quality-adjusted MAP prior...\n")

map_fit <- gMAP(
  cbind(r_effective, n_effective - r_effective) ~ 1 | study_id,
  data      = historical_eff,
  family    = binomial,
  tau.dist  = "HalfNormal",
  tau.prior = 1,
  beta.prior = 2
)

map_beta <- automixfit(map_fit)
posterior_map <- postmix(map_beta, r = x_c, n = n_c)

map_results <- list(
  tau_median      = median(map_fit$tau),
  prior_mean      = unname(summary(map_beta)["mean"]),
  prior_sd        = unname(summary(map_beta)["sd"]),
  posterior_mean  = unname(summary(posterior_map)["mean"]),
  posterior_sd    = unname(summary(posterior_map)["sd"]),
  ci_lower        = qmix(posterior_map, 0.025),
  ci_upper        = qmix(posterior_map, 0.975)
)

# =====
#
# 5. NO BORROWING REFERENCE POSTERIOR
# =====

```

```

vague_prior <- mixbeta(c(1, 1, 1)) # Beta(1,1)
posterior_vague <- postmix(vague_prior, r = x_c, n = n_c)

n_mc_ess <- 100000

no_borrow_samples <- rbeta_mix(posterior_vague, nsim = n_mc_ess)
no_borrow_samples <- pmin(pmax(no_borrow_samples, 1e-8), 1 - 1e-8)
posterior_var_no_borrow <- var(no_borrow_samples)

no_borrow_results <- list(
  mean      = unname(summary(posterior_vague)["mean"]),
  sd        = unname(summary(posterior_vague)["sd"]),
  ci_lower  = qmix(posterior_vague, 0.025),
  ci_upper  = qmix(posterior_vague, 0.975),
  ESS       = 0,
  var_post  = posterior_var_no_borrow
)

# =====
#
# 6. rMAP PRIOR
# =====
#

w_values <- seq(0.02, 1.00, by = 0.02)

rmap_results <- list(
  no_borrowing = no_borrow_results
)

cat("Running rMAP scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(w_values), style = 3)

for (i in seq_along(w_values)) {
  w_r <- w_values[i]

  rmap_prior <- mixcombine(
    informative = map_beta,
    vague      = vague_prior,
    weight     = c(w_r, 1 - w_r)
  )

  post_rmap <- postmix(rmap_prior, r = x_c, n = n_c)

  post_samples <- rbeta_mix(post_rmap, nsim = n_mc_ess)
  post_samples <- pmin(pmax(post_samples, 1e-8), 1 - 1e-8)
  post_var <- var(post_samples)
}

```

```

ESS_add <- compute_additional_ess_variance(
  var_method = post_var,
  var_no_borrow = posterior_var_no_borrow,
  n_current = n_c
)

key <- make_w_key(w_r)

rmap_results[[key]] <- list(
  w_robust = w_r,
  mean      = unname(summary(post_rmap)["mean"]),
  sd        = unname(summary(post_rmap)["sd"]),
  ci_lower  = qmix(post_rmap, 0.025),
  ci_upper  = qmix(post_rmap, 0.975),
  var_post  = post_var,
  ESS       = ESS_add
)

setTxtProgressBar(pb, i)
}
close(pb)

# =====
#
# 7. SAM PRIOR (MULTIPLE HISTORICAL STUDIES SUMMARIZED BY MAP)
# =====
#

sam_vague_prior <- mixbeta(c(1, 1, 1)) # Beta(1,1)
theta_hat_h <- unname(summary(map_beta)["mean"])

cat(sprintf(
  "SAM informative prior summarized by MAP: theta_hat_h = %.4f (%.1f%%)\n",
  theta_hat_h, 100 * theta_hat_h
))

y_current <- c(rep(1, x_c), rep(0, n_c - x_c))

# -----
#
# Posterior informative weight w*
# -----
#

get_w_posterior_from_postmix <- function(sam_post, informative_prior) {
  post_mat <- as.matrix(sam_post)
  inf_mat <- as.matrix(informative_prior)

  if (is.null(dim(post_mat))) post_mat <- matrix(post_mat, nrow = 1)

```

```

if (is.null(dim(inf_mat))) inf_mat <- matrix(inf_mat, nrow = 1)

n_inf <- nrow(inf_mat)

if (nrow(post_mat) < n_inf) {
  stop("Posterior mixture has fewer components than informative prior.")
}

sum(post_mat[seq_len(n_inf), 1])
}

# -----
--
# Main SAM function
# -----
--
compute_SAM_binary_map_samprior <- function(
  x_c,
  n_c,
  y_current,
  informative_prior,
  vague_prior = mixbeta(c(1, 1, 1)),
  Delta,
  var_no_borrow,
  n_mc = 100000
) {

  theta_hat_h <- unname(summary(informative_prior)["mean"])

  w_sam <- SAM_weight(
    if.prior = informative_prior,
    nf.prior = vague_prior,
    delta    = Delta,
    data     = y_current
  )

  w_sam <- min(max(w_sam, 1e-8), 1 - 1e-8)

  sam_prior <- SAM_prior(
    if.prior = informative_prior,
    nf.prior = vague_prior,
    weight   = w_sam
  )

  sam_post <- postmix(sam_prior, r = x_c, n = n_c)

  w_post <- get_w_posterior_from_postmix(

```

```

    sam_post = sam_post,
    informative_prior = informative_prior
  )

post_samples <- rbeta_mix(sam_post, nsim = n_mc)
post_samples <- pmin(pmax(post_samples, 1e-8), 1 - 1e-8)
post_var <- var(post_samples)

ESS_add <- compute_additional_ess_variance(
  var_method = post_var,
  var_no_borrow = var_no_borrow,
  n_current = n_c
)

list(
  Delta      = Delta,
  theta_hat_h = theta_hat_h,
  w_prior    = w_sam,
  w_posterior = w_post,
  mean       = unname(summary(sam_post)["mean"]),
  sd         = unname(summary(sam_post)["sd"]),
  ci_lower   = qmix(sam_post, 0.025),
  ci_upper   = qmix(sam_post, 0.975),
  var_post   = post_var,
  ESS        = ESS_add,
  sam_prior  = sam_prior,
  sam_post   = sam_post
)
}
Delta_values <- seq(0.01, 0.50, length.out = 50)

sam_results <- list(
  informative_prior = list(
    theta_hat_h = theta_hat_h,
    map_prior_mean = unname(summary(map_beta)["mean"]),
    map_prior_sd = unname(summary(map_beta)["sd"])
  )
)

cat("Running SAM scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(Delta_values), style = 3)

for (i in seq_along(Delta_values)) {
  Delta <- Delta_values[i]
  key <- make_delta_key(Delta)

  sam_results[[key]] <- compute_SAM_binary_map_samprior(
    x_c = x_c,
    n_c = n_c,

```

```

    y_current      = y_current,
    informative_prior = map_beta,
    vague_prior     = sam_vague_prior,
    Delta          = Delta,
    var_no_borrow   = posterior_var_no_borrow,
    n_mc            = n_mc_ess
  )

  setTxtProgressBar(pb, i)
}
close(pb)

sam_results_df <- purrr::map_dfr(
  Delta_values,
  function(Delta) {
    key <- make_delta_key(Delta)
    res <- sam_results[[key]]

    tibble::tibble(
      Delta      = res$Delta,
      theta_hat_h = res$theta_hat_h,
      w_prior    = res$w_prior,
      w_posterior = res$w_posterior,
      posterior_mean = res$mean,
      posterior_sd  = res$sd,
      ci_lower     = res$ci_lower,
      ci_upper     = res$ci_upper,
      ESS         = res$ESS
    )
  }
)

print(sam_results_df)

# =====
#
# 8. ELASTIC PRIOR
# =====
#

map_prior_samples_prob <- rmix(map_beta, 100000)
map_prior_samples_prob <- pmin(pmax(map_prior_samples_prob, 1e-8), 1 - 1e-8)
map_prior_samples_logit <- logit(map_prior_samples_prob)

mu_h_hat <- mean(map_prior_samples_logit)
sigma_h_sq_hat <- var(map_prior_samples_logit)
sigma_h_hat <- sqrt(sigma_h_sq_hat)

p_c_adj <- adjust_prop(p_c, n_c)

```

```

theta_c_hat <- logit(p_c_adj)
var_theta_c_hat <- 1 / (n_c * p_c_adj * (1 - p_c_adj))

calculate_discount_elastic <- function(theta_c_hat, var_theta_c_hat,
                                     mu_h_hat, sigma_h_sq_hat,
                                     a = 0, b = 2) {
  T_stat <- (theta_c_hat - mu_h_hat)^2 / (var_theta_c_hat + sigma_h_sq_hat)
  T_stat <- max(T_stat, 1e-10)
  g_T <- 1 / (1 + exp(a + b * log(T_stat)))
  list(T = T_stat, g_T = g_T)
}

compute_elastic_posterior <- function(theta_c_hat, var_theta_c_hat,
                                     mu_h_hat, sigma_h_sq_hat,
                                     g_T, var_no_borrow,
                                     n_current, n_mc = 100000) {
  g_T_safe <- max(g_T, 1e-8)
  sigma_E_sq <- sigma_h_sq_hat / g_T_safe

  prec_post <- 1 / sigma_E_sq + 1 / var_theta_c_hat
  var_post_logit <- 1 / prec_post
  mean_post_logit <- var_post_logit * (
    mu_h_hat / sigma_E_sq + theta_c_hat / var_theta_c_hat
  )

  samps_logit <- rnorm(n_mc, mean_post_logit, sqrt(var_post_logit))
  samps_prob <- inv_logit(samps_logit)
  post_var_prob <- var(samps_prob)

  ESS_add <- compute_additional_ess_variance(
    var_method = post_var_prob,
    var_no_borrow = var_no_borrow,
    n_current = n_current
  )

  list(
    mean = mean(samps_prob),
    sd = sd(samps_prob),
    ci_lower = unname(quantile(samps_prob, 0.025)),
    ci_upper = unname(quantile(samps_prob, 0.975)),
    mu_post_logit = mean_post_logit,
    sigma_post_logit = sqrt(var_post_logit),
    sigma_E = sqrt(sigma_E_sq),
    g_T = g_T_safe,
    var_post = post_var_prob,
    ESS = ESS_add
  )
}

```

```

a_values <- seq(-2, 3, length.out = 10)
b_values <- seq(0.5, 4, length.out = 5)
elastic_grid <- expand.grid(a = a_values, b = b_values) %>%
  arrange(a, b)

elastic_results <- list(
  mu_h_hat = mu_h_hat,
  sigma_h_hat = sigma_h_hat,
  sigma_h_sq_hat = sigma_h_sq_hat
)

cat("Running Elastic scenarios...\n")
pb <- txtProgressBar(min = 0, max = nrow(elastic_grid), style = 3)

for (i in seq_len(nrow(elastic_grid))) {
  a_i <- elastic_grid$a[i]
  b_i <- elastic_grid$b[i]

  disc <- calculate_discount_elastic(
    theta_c_hat = theta_c_hat,
    var_theta_c_hat = var_theta_c_hat,
    mu_h_hat = mu_h_hat,
    sigma_h_sq_hat = sigma_h_sq_hat,
    a = a_i,
    b = b_i
  )

  epost <- compute_elastic_posterior(
    theta_c_hat = theta_c_hat,
    var_theta_c_hat = var_theta_c_hat,
    mu_h_hat = mu_h_hat,
    sigma_h_sq_hat = sigma_h_sq_hat,
    g_T = disc$g_T,
    var_no_borrow = posterior_var_no_borrow,
    n_current = n_c,
    n_mc = n_mc_ess
  )

  key <- make_el_key(a_i, b_i)
  elastic_results[[key]] <- c(
    list(a = a_i, b = b_i, T_stat = disc$T),
    epost
  )

  setTxtProgressBar(pb, i)
}
close(pb)

# =====

```

```

==
# 9. NNHM
# =====
==
# QUALITY-ADJUSTED NNHM:
#   Level 1:
#      $y_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$ 
#
#   Level 2:
#      $\theta_j \mid \mu, \sigma^2 \sim N(\mu, \sigma^2 / q_j)$ 
#
# where  $q_j$  in  $(0,1]$  is the study-specific similarity weight.
# The current study uses  $q_j = 1$ .
# =====
==

NNHM_posterior <- function(y_j, sigma_j, similarity_weights = NULL,
                           mu_0 = 0, sigma_mu = 10, sigma_sigma = 1,
                           n_iter = 10000, n_burnin = 2000,
                           current_idx = 1) {

  J <- length(y_j)

  if (is.null(similarity_weights)) {
    similarity_weights <- rep(1, J)
  }

  similarity_weights <- pmax(similarity_weights, 1e-6)
  similarity_weights[current_idx] <- 1

  theta_j <- y_j
  mu <- mean(y_j)
  sigma <- max(sd(y_j), 0.1)

  mu_samps <- numeric(n_iter)
  sigma_samps <- numeric(n_iter)
  theta_samps <- matrix(NA_real_, nrow = n_iter, ncol = J)

  for (iter in seq_len(n_iter)) {
    for (j in seq_len(J)) {
      sigma2_j <- sigma_j[j]^2
      sigma2_between_j <- sigma^2 / similarity_weights[j]

      prec_j <- 1 / sigma2_j + 1 / sigma2_between_j
      var_j <- 1 / prec_j
      mean_j <- var_j * (y_j[j] / sigma2_j + mu / sigma2_between_j)

      theta_j[j] <- rnorm(1, mean_j, sqrt(var_j))
    }
  }
}

```

```

prec_mu <- sum(similarity_weights / sigma^2) + 1 / sigma_mu^2
var_mu <- 1 / prec_mu
mean_mu <- var_mu * (
  sum(similarity_weights * theta_j) / sigma^2 + mu_0 / sigma_mu^2
)

mu <- rnorm(1, mean_mu, sqrt(var_mu))

sigma_prop <- exp(log(sigma) + rnorm(1, 0, 0.25))

logpost_sigma <- function(sigma_val) {
  if (sigma_val <= 0) return(-Inf)

  loglik_theta <- sum(
    dnorm(
      theta_j,
      mean = mu,
      sd = sqrt(sigma_val^2 / similarity_weights),
      log = TRUE
    )
  )

  logprior_sigma <- dnorm(sigma_val, mean = 0, sd = sigma_sigma, log = TR
UE)

  loglik_theta + logprior_sigma + log(sigma_val)
}

if (log(runif(1)) < (logpost_sigma(sigma_prop) - logpost_sigma(sigma))) {
  sigma <- sigma_prop
}

mu_samps[iter] <- mu
sigma_samps[iter] <- sigma
theta_samps[iter, ] <- theta_j
}

keep <- (n_burnin + 1):n_iter

list(
  theta_c_samples = theta_samps[keep, current_idx],
  mu_mean = mean(mu_samps[keep]),
  sigma_mean = mean(sigma_samps[keep])
)
}

all_props <- c(current$prop, historical$prop)

```

```

all_N <- c(n_c, historical$N_pain)
all_q <- c(1, historical$q_j)

all_props_adj <- adjust_prop(all_props, all_N)
y_j_nnhm <- logit(all_props_adj)
sigma_j_nnhm <- sqrt(1 / (all_N * all_props_adj * (1 - all_props_adj)))

sigma_sigma_values <- seq(0.1, 5.0, length.out = 50)
nnhm_results <- list()

cat("Running NNHM scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(sigma_sigma_values), style = 3)

for (i in seq_along(sigma_sigma_values)) {
  sigma_sigma_i <- sigma_sigma_values[i]

  fit <- NNHM_posterior(
    y_j = y_j_nnhm,
    sigma_j = sigma_j_nnhm,
    similarity_weights = all_q,
    mu_0 = 0,
    sigma_mu = 10,
    sigma_sigma = sigma_sigma_i,
    n_iter = 8000,
    n_burnin = 1500,
    current_idx = 1
  )

  samps_prob <- inv_logit(fit$theta_c_samples)
  post_var_prob <- var(samps_prob)

  ESS_add <- compute_additional_ess_variance(
    var_method = post_var_prob,
    var_no_borrow = posterior_var_no_borrow,
    n_current = n_c
  )

  key <- make_sigma_key(sigma_sigma_i)

  nnhm_results[[key]] <- list(
    sigma_sigma = sigma_sigma_i,
    mu_posterior = fit$mu_mean,
    sigma_posterior = fit$sigma_mean,
    mean = mean(samps_prob),
    ci_lower = unname(quantile(samps_prob, 0.025)),
    ci_upper = unname(quantile(samps_prob, 0.975)),
    var_post = post_var_prob,
    ESS = ESS_add
  )
}

```

```

    setTxtProgressBar(pb, i)
  }
  close(pb)

# =====
#
# 10. SUMMARY TABLE
# =====
#

closest_key <- function(values, target, key_fun) {
  key_fun(values[which.min(abs(values - target))])
}

sam_key_010 <- closest_key(Delta_values, 0.10, make_delta_key)
sam_key_025 <- closest_key(Delta_values, 0.25, make_delta_key)
nnhm_key_05 <- closest_key(sigma_sigma_values, 0.5, make_sigma_key)
nnhm_key_10 <- closest_key(sigma_sigma_values, 1.0, make_sigma_key)

find_elastic_key <- function(target_a, target_b, grid) {
  idx <- which.min(abs(grid$a - target_a) + abs(grid$b - target_b))
  make_el_key(grid$a[idx], grid$b[idx])
}

ek_hi <- find_elastic_key(-1, 1, elastic_grid)
ek_lo <- find_elastic_key(2, 3, elastic_grid)

summary_table <- tibble(
  Method = c(
    "No Borrowing",
    "rMAP (w=0.50)",
    "rMAP (w=0.80)",
    "SAM (Delta~0.10)",
    "SAM (Delta~0.25)",
    "Elastic (a=-1, b=1) ",
    "Elastic (a=2, b=3)",
    "NNHM (sigma_sigma~0.5)",
    "NNHM (sigma_sigma~1.0)"
  ),
  Posterior_Mean = c(
    rmap_results$no_borrowing$mean,
    rmap_results[[make_w_key(0.50)]]$mean,
    rmap_results[[make_w_key(0.80)]]$mean,
    sam_results[[sam_key_010]]$mean,
    sam_results[[sam_key_025]]$mean,
    elastic_results[[ek_hi]]$mean,
    elastic_results[[ek_lo]]$mean,
    nnhm_results[[nnhm_key_05]]$mean,

```

```

    nnhm_results[[nnhm_key_10]]$mean
  ),
  CI_Lower = c(
    rmap_results$no_borrowing$ci_lower,
    rmap_results[[make_w_key(0.50)]]$ci_lower,
    rmap_results[[make_w_key(0.80)]]$ci_lower,
    sam_results[[sam_key_010]]$ci_lower,
    sam_results[[sam_key_025]]$ci_lower,
    elastic_results[[ek_hi]]$ci_lower,
    elastic_results[[ek_lo]]$ci_lower,
    nnhm_results[[nnhm_key_05]]$ci_lower,
    nnhm_results[[nnhm_key_10]]$ci_lower
  ),
  CI_Upper = c(
    rmap_results$no_borrowing$ci_upper,
    rmap_results[[make_w_key(0.50)]]$ci_upper,
    rmap_results[[make_w_key(0.80)]]$ci_upper,
    sam_results[[sam_key_010]]$ci_upper,
    sam_results[[sam_key_025]]$ci_upper,
    elastic_results[[ek_hi]]$ci_upper,
    elastic_results[[ek_lo]]$ci_upper,
    nnhm_results[[nnhm_key_05]]$ci_upper,
    nnhm_results[[nnhm_key_10]]$ci_upper
  ),
  ESS = c(
    0,
    rmap_results[[make_w_key(0.50)]]$ESS,
    rmap_results[[make_w_key(0.80)]]$ESS,
    sam_results[[sam_key_010]]$ESS,
    sam_results[[sam_key_025]]$ESS,
    elastic_results[[ek_hi]]$ESS,
    elastic_results[[ek_lo]]$ESS,
    nnhm_results[[nnhm_key_05]]$ESS,
    nnhm_results[[nnhm_key_10]]$ESS
  )
) %>%
mutate(
  Posterior_Mean = round(to_percent(Posterior_Mean), 2),
  CI_Lower       = round(to_percent(CI_Lower), 2),
  CI_Upper       = round(to_percent(CI_Upper), 2),
  ESS            = round(ESS, 3),
  CI_Width       = round(CI_Upper - CI_Lower, 2)
)

print(summary_table)

# =====
#
# 11. SENSITIVITY TABLES
# =====

```

```

==

sensitivity_rmap <- tibble(
  Method = "rMAP",
  Parameter = paste0("w=", sprintf("%.2f", w_values)),
  Parameter_Value = w_values,
  Mean = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$mean),
  CI_Lower = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ci_lo
wer),
  CI_Upper = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ci_up
per),
  ESS = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ESS)
) %>%
  mutate(
    Mean = round(to_percent(Mean), 2),
    CI_Lower = round(to_percent(CI_Lower), 2),
    CI_Upper = round(to_percent(CI_Upper), 2),
    ESS = round(ESS, 3),
    CI_Width = round(CI_Upper - CI_Lower, 2)
  )

sensitivity_sam <- tibble(
  Method = "SAM",
  Parameter = paste0("Delta=", sprintf("%.3f", Delta_values)),
  Parameter_Value = Delta_values,
  Mean = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$me
an),
  CI_Lower = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)
]]$ci_lower),
  CI_Upper = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)
]]$ci_upper),
  w_prior = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]
$w_prior),
  w_posterior = sapply(Delta_values, function(d) sam_results[[make_delta_key(
d)]]$w_posterior),
  ESS = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$ESS
)
) %>%
  mutate(
    Mean = round(to_percent(Mean), 2),
    CI_Lower = round(to_percent(CI_Lower), 2),
    CI_Upper = round(to_percent(CI_Upper), 2),
    w_prior = round(w_prior, 4),
    w_posterior = round(w_posterior, 4),
    ESS = round(ESS, 3),
    CI_Width = round(CI_Upper - CI_Lower, 2)
  )

elastic_keys <- mapply(make_el_key, elastic_grid$a, elastic_grid$b)

```

```
sensitivity_elastic <- tibble(
  Method = "Elastic",
  Parameter = sprintf("a=%.2f,b=%.2f", elastic_grid$a, elastic_grid$b),
  Parameter_Value = seq_len(nrow(elastic_grid)),
  a = elastic_grid$a,
  b = elastic_grid$b,
  Mean = sapply(elastic_keys, function(k) elastic_results[[k]]$mean),
  CI_Lower = sapply(elastic_keys, function(k) elastic_results[[k]]$ci_lower),
  CI_Upper = sapply(elastic_keys, function(k) elastic_results[[k]]$ci_upper),
  g_T = sapply(elastic_keys, function(k) elastic_results[[k]]$g_T),
  ESS = sapply(elastic_keys, function(k) elastic_results[[k]]$ESS)
) %>%
mutate(
  Mean = round(to_percent(Mean), 2),
  CI_Lower = round(to_percent(CI_Lower), 2),
  CI_Upper = round(to_percent(CI_Upper), 2),
  g_T = round(g_T, 4),
  ESS = round(ESS, 3),
  CI_Width = round(CI_Upper - CI_Lower, 2)
)

sensitivity_nnhm <- tibble(
  Method = "NNHM",
  Parameter = paste0("sigma_sigma=", sprintf("%.2f", sigma_sigma_values)),
  Parameter_Value = sigma_sigma_values,
  Mean = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_key(
x)]]$mean),
  CI_Lower = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_
key(x)]]$ci_lower),
  CI_Upper = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_
key(x)]]$ci_upper),
  sigma_posterior = sapply(sigma_sigma_values, function(x) nnhm_results[[make
_sigma_key(x)]]$sigma_posterior),
  ESS = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_key(x
)]]$ESS)
) %>%
mutate(
  Mean = round(to_percent(Mean), 2),
  CI_Lower = round(to_percent(CI_Lower), 2),
  CI_Upper = round(to_percent(CI_Upper), 2),
  sigma_posterior = round(sigma_posterior, 4),
  ESS = round(ESS, 3),
  CI_Width = round(CI_Upper - CI_Lower, 2)
)

sensitivity_full <- bind_rows(
  sensitivity_rmap %>% select(Method, Parameter, Parameter_Value, Mean, CI_Lo
wer, CI_Upper, ESS, CI_Width),
  sensitivity_sam %>% select(Method, Parameter, Parameter_Value, Mean, CI_Low
```

```

er, CI_Upper, ESS, CI_Width),
  sensitivity_elastic %>% select(Method, Parameter, Parameter_Value, Mean, CI
_Lower, CI_Upper, ESS, CI_Width),
  sensitivity_nnhm %>% select(Method, Parameter, Parameter_Value, Mean, CI_Lo
wer, CI_Upper, ESS, CI_Width)
)

# =====
==
# 12. SAVE
# =====
==

save_path <- "C:/Users/marin/Downloads/RWD results2/"
if (!dir.exists(save_path)) dir.create(save_path, recursive = TRUE)

all_results <- list(
  current_study = current,
  historical_studies = historical,
  historical_effective = historical_eff,
  map = map_results,
  no_borrowing = no_borrow_results,
  rmap = rmap_results,
  sam = sam_results,
  elastic = elastic_results,
  nnhm = nnhm_results,
  summary_table = summary_table,
  sensitivity_rmap = sensitivity_rmap,
  sensitivity_sam = sensitivity_sam,
  sensitivity_elastic = sensitivity_elastic,
  sensitivity_nnhm = sensitivity_nnhm,
  sensitivity_full = sensitivity_full,
  parameters = list(
    w_values = w_values,
    Delta_values = Delta_values,
    elastic_grid = elastic_grid,
    sigma_sigma_values = sigma_sigma_values
  )
)

write.csv(
  summary_table,
  file.path(save_path, "bayesian_borrowing_summary.csv"),
  row.names = FALSE
)

write.csv(

```

```
sensitivity_full,
file.path(save_path, "sensitivity_analysis_200scenarios.csv"),
row.names = FALSE
)

write.csv(
  sensitivity_rmap,
  file.path(save_path, "sensitivity_rmap_50.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_sam,
  file.path(save_path, "sensitivity_sam_50.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_elastic,
  file.path(save_path, "sensitivity_elastic_50.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_nnhm,
  file.path(save_path, "sensitivity_nnhm_50.csv"),
  row.names = FALSE
)

save(
  all_results,
  file = file.path(save_path, "bayesian_borrowing_complete_200scenarios.RData")
)

cat("\nAll files saved to", save_path, "\n")
```

12.5 RWD analysis (Continuous endpoint)

```
# =====
==
# BAYESIAN BORROWING ANALYSIS - CONTINUOUS ENDPOINT
# QUALITY-ADJUSTED THEORY-CONSISTENT IMPLEMENTATION
# =====
==
#
# ENDPOINT:
#   Mean_decrease = mean decrease of D-dimer from baseline
#   SD_decrease   = SD of decrease from baseline
#   N_d_dimer     = number with pre/post D-dimer values
```

```
#
# METHODS REPORTED:
# 1. rMAP - Robust MAP prior
# 2. SAM - Self-Adapting Mixture Prior
# 3. Elastic Prior
# 4. NNHM - Normal-Normal Hierarchical Model
#
# INTERNAL METHOD:
# - MAP prior is fitted internally because it is required to construct:
#   (a) rMAP
#   (b) SAM for multiple historical studies
#   (c) Elastic prior
#
# QUALITY ADJUSTMENT:
# q_j = Similarity_j × sqrt(ROB_j × Design_j)
#
# HOW QUALITY WEIGHTS ENTER:
# MAP / rMAP / Elastic / SAM-informative-prior:
#   via quality-adjusted effective sample size:
#   n_j,eff = round(n_j * q_j)
#   with within-study SE:
#   se_j,eff = SD_j / sqrt(n_j,eff)
#
# NNHM:
#   via similarity-weighted second level:
#   theta_j | mu, sigma^2 ~ N(mu, sigma^2 / q_j)
#
# ADDITIONAL ESS:
# For every borrowing method, the reported ESS is the posterior variance-based
# additional ESS on the mean scale:
#
# ESS_add = max(0, n_c * (Var_no_borrow / Var_method - 1))
# =====
==

library(RBest)
library(tidyverse)

set.seed(2024)

# =====
==
# 1. LOAD DATA
# =====
==

data <- read.csv(
  "C:/Users/marin/Downloads/Datasets/calculated_weights.csv",
```

```

    stringsAsFactors = FALSE
  )

current <- data %>%
  filter(Study_type == "current") %>%
  filter(
    !is.na(Mean_decrease),
    !is.na(SD_decrease),
    !is.na(N_d_dimer)
  )

historical <- data %>%
  filter(Study_type == "historical") %>%
  filter(
    !is.na(Mean_decrease),
    !is.na(SD_decrease),
    !is.na(N_d_dimer)
  )

stopifnot(nrow(current) == 1)

current <- current %>%
  mutate(
    q_j = 1
  )

historical <- historical %>%
  mutate(
    q_j = pmin(pmax(Quality_Adjusted, 1e-6), 1)
  )

mean_c <- current$Mean_decrease
sd_c    <- current$SD_decrease
n_c     <- current$N_d_dimer
se_c    <- sd_c / sqrt(n_c)

cat(sprintf(
  "Current study: %s | N=%d | Mean decrease=%.4f | SD=%.4f | SE=%.4f\n",
  current$Study, n_c, mean_c, sd_c, se_c
))
cat(sprintf(
  "Historical studies: %d | raw total N=%d\n\n",
  nrow(historical),
  sum(historical$N_d_dimer)
))

# =====
==
# 2. SHARED HELPERS

```

```
# =====
==

make_w_key      <- function(w) sprintf("w_%0.2f", w)
make_delta_key  <- function(d) sprintf("Delta_%0.4f", d)
make_sigma_key  <- function(x) sprintf("sigma_sigma_%0.3f", x)
make_el_key     <- function(a, b) sprintf("a_%0.2f_b_%0.2f", a, b)

# -----
--
# Sample from normal mixture
# -----
--
rnorm_mix <- function(mix_obj, nsim = 100000) {
  mat <- as.matrix(t(mix_obj))
  if (is.null(dim(mat))) {
    mat <- matrix(mat, nrow = 1)
  }

  w <- mat[, 1]
  mu <- mat[, 2]
  sd <- mat[, 3]

  comp <- sample(seq_along(w), size = nsim, replace = TRUE, prob = w)
  rnorm(nsim, mean = mu[comp], sd = sd[comp])
}

# -----
--
# Posterior variance-based additional ESS
# -----
--
compute_additional_ess_variance <- function(var_method, var_no_borrow, n_current) {
  if (is.na(var_method) || is.na(var_no_borrow) || var_method <= 0 || var_no_borrow <= 0) {
    return(NA_real_)
  }
  max(0, n_current * (var_no_borrow / var_method - 1))
}

# -----
--
# Posterior informative weight w*
# -----
--
get_w_posterior_from_postmix <- function(post_mix, informative_prior) {
  post_mat <- as.matrix(post_mix)
  inf_mat <- as.matrix(informative_prior)
}
```

```

if (is.null(dim(post_mat))) post_mat <- matrix(post_mat, nrow = 1)
if (is.null(dim(inf_mat))) inf_mat <- matrix(inf_mat, nrow = 1)

n_inf <- nrow(inf_mat)

if (nrow(post_mat) < n_inf) {
  stop("Posterior mixture has fewer components than informative prior.")
}

sum(post_mat[seq_len(n_inf), 1])
}

# =====
#
# 3. QUALITY-ADJUSTED HISTORICAL DATA FOR MAP / rMAP / SAM / ELASTIC
# =====
#

historical_eff <- historical %>%
  mutate(
    n_effective = pmax(round(N_d_dimer * q_j), 1L),
    se_effective = SD_decrease / sqrt(n_effective),
    study_id = factor(seq_len(n()))
  )

cat(sprintf(
  "Quality-adjusted historical total N = %d\n\n",
  sum(historical_eff$n_effective)
))

# =====
#
# 4. INTERNAL MAP PRIOR
# =====
#

cat("Fitting quality-adjusted MAP prior...\n")

map_fit <- gMAP(
  cbind(Mean_decrease, se_effective) ~ 1 | study_id,
  data = historical_eff,
  family = gaussian,
  tau.dist = "HalfNormal",
  tau.prior = 1,
  beta.prior = 2
)
map_norm <- automixfit(map_fit)
posterior_map <- postmix(map_norm, m = mean_c, se = se_c)

```

```

map_results <- list(
  tau_median      = median(map_fit$tau),
  prior_mean      = unname(summary(map_norm)["mean"]),
  prior_sd        = unname(summary(map_norm)["sd"]),
  posterior_mean   = unname(summary(posterior_map)["mean"]),
  posterior_sd     = unname(summary(posterior_map)["sd"]),
  ci_lower        = qmix(posterior_map, 0.025),
  ci_upper        = qmix(posterior_map, 0.975)
)

# =====
==
# 5. NO BORROWING REFERENCE POSTERIOR
# =====
==

vague_prior <- mixnorm(c(1, 0, 100)) # N(0, 100^2)
posterior_vague <- postmix(vague_prior, m = mean_c, se = se_c)

n_mc_ess <- 100000

no_borrow_samples <- rnorm_mix(posterior_vague, nsim = n_mc_ess)
posterior_var_no_borrow <- var(no_borrow_samples)

no_borrow_results <- list(
  mean      = unname(summary(posterior_vague)["mean"]),
  sd        = unname(summary(posterior_vague)["sd"]),
  ci_lower  = qmix(posterior_vague, 0.025),
  ci_upper  = qmix(posterior_vague, 0.975),
  ESS       = 0,
  var_post  = posterior_var_no_borrow
)

# =====
==
# 6. rMAP PRIOR
# =====
==

w_values <- seq(0.02, 1.00, by = 0.02)

rmap_results <- list(
  no_borrowing = no_borrow_results
)

cat("Running rMAP scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(w_values), style = 3)

```

```

for (i in seq_along(w_values)) {
  w_r <- w_values[i]

  rmap_prior <- mixcombine(
    informative = map_norm,
    vague = vague_prior,
    weight = c(w_r, 1 - w_r)
  )

  post_rmap <- postmix(rmap_prior, m = mean_c, se = se_c)

  post_samples <- rnorm_mix(post_rmap, nsim = n_mc_ess)
  post_var <- var(post_samples)

  ESS_add <- compute_additional_ess_variance(
    var_method = post_var,
    var_no_borrow = posterior_var_no_borrow,
    n_current = n_c
  )

  key <- make_w_key(w_r)

  rmap_results[[key]] <- list(
    w_robust = w_r,
    mean      = unname(summary(post_rmap)["mean"]),
    sd        = unname(summary(post_rmap)["sd"]),
    ci_lower  = qmix(post_rmap, 0.025),
    ci_upper  = qmix(post_rmap, 0.975),
    var_post  = post_var,
    ESS       = ESS_add
  )

  setTxtProgressBar(pb, i)
}
close(pb)

# =====
#
# 7. SAM PRIOR (MULTIPLE HISTORICAL STUDIES SUMMARIZED BY MAP)
# CONTINUOUS ENDPOINT VERSION
# =====
#

sam_vague_prior <- vague_prior
theta_hat_h <- unname(summary(map_norm)["mean"])

cat(sprintf(

```

```

"SAM informative prior summarized by MAP: theta_hat_h = %.4f\n",
theta_hat_h
))

# -----
--
# Normal-likelihood ratio for continuous endpoint
# -----
--
compute_R_continuous <- function(mean_c, se_c, theta_hat_h, Delta) {
  logL <- function(theta) dnorm(mean_c, mean = theta, sd = se_c, log = TRUE)

  logL_agree <- logL(theta_hat_h)
  logL_disagree <- max(logL(theta_hat_h + Delta), logL(theta_hat_h - Delta))

  exp(logL_agree - logL_disagree)
}

# -----
--
# Main SAM function
# -----
--
compute_SAM_continuous_map <- function(
  mean_c,
  se_c,
  n_c,
  informative_prior,
  vague_prior = mixnorm(c(1, 0, 100)),
  Delta,
  var_no_borrow,
  n_mc = 100000
) {

  theta_hat_h <- unname(summary(informative_prior)["mean"])

  R_val <- compute_R_continuous(
    mean_c = mean_c,
    se_c = se_c,
    theta_hat_h = theta_hat_h,
    Delta = Delta
  )

  w_sam <- R_val / (1 + R_val)
  w_sam <- min(max(w_sam, 1e-8), 1 - 1e-8)

  sam_prior <- mixcombine(
    informative = informative_prior,
    vague = vague_prior,

```

```

    weight = c(w_sam, 1 - w_sam)
  )

  sam_post <- postmix(sam_prior, m = mean_c, se = se_c)

  w_post <- get_w_posterior_from_postmix(
    post_mix = sam_post,
    informative_prior = informative_prior
  )

  post_samples <- rnorm_mix(sam_post, nsim = n_mc)
  post_var <- var(post_samples)

  ESS_add <- compute_additional_ess_variance(
    var_method = post_var,
    var_no_borrow = var_no_borrow,
    n_current = n_c
  )

  list(
    Delta      = Delta,
    theta_hat_h = theta_hat_h,
    w_prior    = w_sam,
    w_posterior = w_post,
    mean       = unname(summary(sam_post)["mean"]),
    sd         = unname(summary(sam_post)["sd"]),
    ci_lower   = qmix(sam_post, 0.025),
    ci_upper   = qmix(sam_post, 0.975),
    var_post   = post_var,
    ESS        = ESS_add,
    sam_prior  = sam_prior,
    sam_post   = sam_post
  )
}

Delta_values <- seq(0.1, 3, length.out = 50)

sam_results <- list(
  informative_prior = list(
    theta_hat_h      = theta_hat_h,
    map_prior_mean   = unname(summary(map_norm)["mean"]),
    map_prior_sd     = unname(summary(map_norm)["sd"])
  )
)

cat("Running SAM scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(Delta_values), style = 3)

for (i in seq_along(Delta_values)) {

```

```

Delta <- Delta_values[i]
key <- make_delta_key(Delta)

sam_results[[key]] <- compute_SAM_continuous_map(
  mean_c      = mean_c,
  se_c        = se_c,
  n_c         = n_c,
  informative_prior = map_norm,
  vague_prior  = sam_vague_prior,
  Delta       = Delta,
  var_no_borrow = posterior_var_no_borrow,
  n_mc        = n_mc_ess
)

setTxtProgressBar(pb, i)
}
close(pb)

sam_results_df <- purrr::map_dfr(
  Delta_values,
  function(Delta) {
    key <- make_delta_key(Delta)
    res <- sam_results[[key]]

    tibble::tibble(
      Delta           = res$Delta,
      theta_hat_h    = res$theta_hat_h,
      w_prior        = res$w_prior,
      w_posterior    = res$w_posterior,
      posterior_mean  = res$mean,
      posterior_sd    = res$sd,
      ci_lower       = res$ci_lower,
      ci_upper       = res$ci_upper,
      ESS            = res$ESS
    )
  }
)

print(sam_results_df)

# =====
#
# 8. ELASTIC PRIOR
# =====
#

map_prior_samples <- rmix(map_norm, 100000)

mu_h_hat <- mean(map_prior_samples)

```

```

sigma_h_sq_hat <- var(map_prior_samples)
sigma_h_hat <- sqrt(sigma_h_sq_hat)

theta_c_hat <- mean_c
var_theta_c_hat <- se_c^2

calculate_discount_elastic <- function(theta_c_hat, var_theta_c_hat,
                                       mu_h_hat, sigma_h_sq_hat,
                                       a = 0, b = 2) {
  T_stat <- (theta_c_hat - mu_h_hat)^2 / (var_theta_c_hat + sigma_h_sq_hat)
  T_stat <- max(T_stat, 1e-10)

  g_T <- 1 / (1 + exp(a + b * log(T_stat)))

  list(T = T_stat, g_T = g_T)
}

compute_elastic_posterior <- function(theta_c_hat, var_theta_c_hat,
                                       mu_h_hat, sigma_h_sq_hat,
                                       g_T, var_no_borrow,
                                       n_current) {

  g_T_safe <- max(g_T, 1e-8)

  sigma_E_sq <- sigma_h_sq_hat / g_T_safe
  sigma_E <- sqrt(sigma_E_sq)

  prec_post <- 1 / sigma_E_sq + 1 / var_theta_c_hat
  var_post <- 1 / prec_post
  sd_post <- sqrt(var_post)

  mean_post <- var_post * (
    mu_h_hat / sigma_E_sq + theta_c_hat / var_theta_c_hat
  )

  ESS_add <- compute_additional_ess_variance(
    var_method = var_post,
    var_no_borrow = var_no_borrow,
    n_current = n_current
  )

  list(
    mean = mean_post,
    sd = sd_post,
    ci_lower = qnorm(0.025, mean_post, sd_post),
    ci_upper = qnorm(0.975, mean_post, sd_post),
    mu_post = mean_post,
    sigma_post = sd_post,

```

```

    sigma_E = sigma_E,
    g_T = g_T_safe,
    var_post = var_post,
    ESS = ESS_add
  )
}

a_values <- seq(-2, 3, length.out = 10)
b_values <- seq(0.5, 4, length.out = 5)

elastic_grid <- expand.grid(a = a_values, b = b_values) %>%
  arrange(a, b)

elastic_results <- list(
  mu_h_hat = mu_h_hat,
  sigma_h_hat = sigma_h_hat,
  sigma_h_sq_hat = sigma_h_sq_hat
)

cat("Running Elastic scenarios...\n")
pb <- txtProgressBar(min = 0, max = nrow(elastic_grid), style = 3)

for (i in seq_len(nrow(elastic_grid))) {
  a_i <- elastic_grid$a[i]
  b_i <- elastic_grid$b[i]

  disc <- calculate_discount_elastic(
    theta_c_hat = theta_c_hat,
    var_theta_c_hat = var_theta_c_hat,
    mu_h_hat = mu_h_hat,
    sigma_h_sq_hat = sigma_h_sq_hat,
    a = a_i,
    b = b_i
  )

  epost <- compute_elastic_posterior(
    theta_c_hat = theta_c_hat,
    var_theta_c_hat = var_theta_c_hat,
    mu_h_hat = mu_h_hat,
    sigma_h_sq_hat = sigma_h_sq_hat,
    g_T = disc$g_T,
    var_no_borrow = posterior_var_no_borrow,
    n_current = n_c
  )

  key <- make_el_key(a_i, b_i)

  elastic_results[[key]] <- c(
    list(a = a_i, b = b_i, T_stat = disc$T),

```

```

    epost
  )

  setTxtProgressBar(pb, i)
}
close(pb)
# =====
==
# 9. NNHM
# =====
==
# QUALITY-ADJUSTED NNHM:
#   Level 1:
#     y_j | theta_j ~ N(theta_j, sigma_j^2)
#
#   Level 2 for historical studies only:
#     theta_j | mu, sigma^2 ~ N(mu, sigma^2 / q_j)
#
# =====
==

NNHM_posterior <- function(y_j, sigma_j, similarity_weights = NULL,
                           mu_0 = 0, sigma_mu = 10, sigma_sigma = 1,
                           n_iter = 10000, n_burnin = 2000,
                           current_idx = 1) {

  J <- length(y_j)

  if (is.null(similarity_weights)) {
    similarity_weights <- rep(1, J)
  }

  similarity_weights <- pmax(similarity_weights, 1e-6)
  similarity_weights[current_idx] <- 1

  theta_j <- y_j
  mu <- mean(y_j)
  sigma <- max(sd(y_j), 0.1)

  mu_samps <- numeric(n_iter)
  sigma_samps <- numeric(n_iter)
  theta_samps <- matrix(NA_real_, nrow = n_iter, ncol = J)

  for (iter in seq_len(n_iter)) {
    for (j in seq_len(J)) {
      sigma2_j <- sigma_j[j]^2
      sigma2_between_j <- sigma^2 / similarity_weights[j]

```

```

    prec_j <- 1 / sigma2_j + 1 / sigma2_between_j
    var_j <- 1 / prec_j
    mean_j <- var_j * (y_j[j] / sigma2_j + mu / sigma2_between_j)

    theta_j[j] <- rnorm(1, mean_j, sqrt(var_j))
  }

  prec_mu <- sum(similarity_weights / sigma^2) + 1 / sigma_mu^2
  var_mu <- 1 / prec_mu
  mean_mu <- var_mu * (
    sum(similarity_weights * theta_j) / sigma^2 + mu_0 / sigma_mu^2
  )

  mu <- rnorm(1, mean_mu, sqrt(var_mu))

  sigma_prop <- exp(log(sigma) + rnorm(1, 0, 0.25))

  logpost_sigma <- function(sigma_val) {
    if (sigma_val <= 0) return(-Inf)

    loglik_theta <- sum(
      dnorm(
        theta_j,
        mean = mu,
        sd = sqrt(sigma_val^2 / similarity_weights),
        log = TRUE
      )
    )

    logprior_sigma <- dnorm(sigma_val, mean = 0, sd = sigma_sigma, log = TR
UE)

    loglik_theta + logprior_sigma + log(sigma_val)
  }

  if (log(runif(1)) < (logpost_sigma(sigma_prop) - logpost_sigma(sigma))) {
    sigma <- sigma_prop
  }

  mu_samps[iter] <- mu
  sigma_samps[iter] <- sigma
  theta_samps[iter, ] <- theta_j
}

keep <- (n_burnin + 1):n_iter

list(
  theta_c_samples = theta_samps[keep, current_idx],

```

```

    mu_mean = mean(mu_samps[keep]),
    sigma_mean = mean(sigma_samps[keep])
  )
}

all_means <- c(mean_c, historical$Mean_decrease)
all_se <- c(se_c, historical$SD_decrease / sqrt(historical$N_d_dimer))
all_q <- c(1, historical$q_j)

sigma_sigma_values <- seq(0.1, 3.0, length.out = 50)
nnhm_results <- list()

cat("Running NNHM scenarios...\n")
pb <- txtProgressBar(min = 0, max = length(sigma_sigma_values), style = 3)

for (i in seq_along(sigma_sigma_values)) {
  sigma_sigma_i <- sigma_sigma_values[i]

  fit <- NNHM_posterior(
    y_j = all_means,
    sigma_j = all_se,
    similarity_weights = all_q,
    mu_0 = 0,
    sigma_mu = 10,
    sigma_sigma = sigma_sigma_i,
    n_iter = 8000,
    n_burnin = 1500,
    current_idx = 1
  )

  samps <- fit$theta_c_samples
  post_var <- var(samps)

  ESS_add <- compute_additional_ess_variance(
    var_method = post_var,
    var_no_borrow = posterior_var_no_borrow,
    n_current = n_c
  )

  key <- make_sigma_key(sigma_sigma_i)

  nnhm_results[[key]] <- list(
    sigma_sigma = sigma_sigma_i,
    mu_posterior = fit$mu_mean,
    sigma_posterior = fit$sigma_mean,
    mean = mean(samps),
    ci_lower = unname(quantile(samps, 0.025)),
    ci_upper = unname(quantile(samps, 0.975)),
    var_post = post_var,

```

```

    ESS = ESS_add
  )

  setTxtProgressBar(pb, i)
}
close(pb)

# =====
#
# 10. SUMMARY TABLE
# =====

closest_key <- function(values, target, key_fun) {
  key_fun(values[which.min(abs(values - target))])
}

sam_key_05 <- closest_key(Delta_values, 0.50, make_delta_key)
sam_key_10 <- closest_key(Delta_values, 1.00, make_delta_key)

nnhm_key_05 <- closest_key(sigma_sigma_values, 0.5, make_sigma_key)
nnhm_key_10 <- closest_key(sigma_sigma_values, 1.0, make_sigma_key)

find_elastic_key <- function(target_a, target_b, grid) {
  idx <- which.min(abs(grid$a - target_a) + abs(grid$b - target_b))
  make_el_key(grid$a[idx], grid$b[idx])
}

ek_hi <- find_elastic_key(-1, 1, elastic_grid)
ek_lo <- find_elastic_key(2, 3, elastic_grid)

summary_table <- tibble(
  Method = c(
    "No Borrowing",
    "rMAP (w=0.50)",
    "rMAP (w=0.80)",
    "SAM (Delta=0.50)",
    "SAM (Delta=1.00)",
    "Elastic (a=-1, b=1)",
    "Elastic (a=2, b=3)",
    "NNHM (sigma_sigma~0.5)",
    "NNHM (sigma_sigma~1.0)"
  ),
  Posterior_Mean = c(
    rmap_results$no_borrowing$mean,
    rmap_results[[make_w_key(0.50)]]$mean,
    rmap_results[[make_w_key(0.80)]]$mean,

```

```

sam_results[[sam_key_05]]$mean,
sam_results[[sam_key_10]]$mean,
elastic_results[[ek_hi]]$mean,
elastic_results[[ek_lo]]$mean,
nnhm_results[[nnhm_key_05]]$mean,
nnhm_results[[nnhm_key_10]]$mean
),
CI_Lower = c(
  rmap_results$no_borrowing$ci_lower,
  rmap_results[[make_w_key(0.50)]]$ci_lower,
  rmap_results[[make_w_key(0.80)]]$ci_lower,
  sam_results[[sam_key_05]]$ci_lower,
  sam_results[[sam_key_10]]$ci_lower,
  elastic_results[[ek_hi]]$ci_lower,
  elastic_results[[ek_lo]]$ci_lower,
  nnhm_results[[nnhm_key_05]]$ci_lower,
  nnhm_results[[nnhm_key_10]]$ci_lower
),
CI_Upper = c(
  rmap_results$no_borrowing$ci_upper,
  rmap_results[[make_w_key(0.50)]]$ci_upper,
  rmap_results[[make_w_key(0.80)]]$ci_upper,
  sam_results[[sam_key_05]]$ci_upper,
  sam_results[[sam_key_10]]$ci_upper,
  elastic_results[[ek_hi]]$ci_upper,
  elastic_results[[ek_lo]]$ci_upper,
  nnhm_results[[nnhm_key_05]]$ci_upper,
  nnhm_results[[nnhm_key_10]]$ci_upper
),
ESS = c(
  0,
  rmap_results[[make_w_key(0.50)]]$ESS,
  rmap_results[[make_w_key(0.80)]]$ESS,
  sam_results[[sam_key_05]]$ESS,
  sam_results[[sam_key_10]]$ESS,
  elastic_results[[ek_hi]]$ESS,
  elastic_results[[ek_lo]]$ESS,
  nnhm_results[[nnhm_key_05]]$ESS,
  nnhm_results[[nnhm_key_10]]$ESS
)
) %>%
mutate(
  Posterior_Mean = round(Posterior_Mean, 4),
  CI_Lower       = round(CI_Lower, 4),
  CI_Upper       = round(CI_Upper, 4),
  ESS            = round(ESS, 3),
  CI_Width       = round(CI_Upper - CI_Lower, 4)
)

print(summary_table)

```

```
# =====
==
# 11. SENSITIVITY TABLES
# =====
==

sensitivity_rmap <- tibble(
  Method = "rMAP",
  Parameter = paste0("w=", sprintf("%.2f", w_values)),
  Parameter_Value = w_values,
  Mean = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$mean),
  CI_Lower = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ci_lower),
  CI_Upper = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ci_upper),
  ESS = sapply(w_values, function(w) rmap_results[[make_w_key(w)]]$ESS)
) %>%
  mutate(
    Mean = round(Mean, 4),
    CI_Lower = round(CI_Lower, 4),
    CI_Upper = round(CI_Upper, 4),
    ESS = round(ESS, 3),
    CI_Width = round(CI_Upper - CI_Lower, 4)
  )

sensitivity_sam <- tibble(
  Method = "SAM",
  Parameter = paste0("Delta=", sprintf("%.3f", Delta_values)),
  Parameter_Value = Delta_values,
  Mean = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$mean),
  CI_Lower = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$ci_lower),
  CI_Upper = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$ci_upper),
  w_prior = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$w_prior),
  w_posterior = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$w_posterior),
  ESS = sapply(Delta_values, function(d) sam_results[[make_delta_key(d)]]$ESS)
) %>%
  mutate(
    Mean = round(Mean, 4),
    CI_Lower = round(CI_Lower, 4),
    CI_Upper = round(CI_Upper, 4),
    w_prior = round(w_prior, 4),
    w_posterior = round(w_posterior, 4),
    ESS = round(ESS, 3),
  )
```

```

    CI_Width = round(CI_Upper - CI_Lower, 4)
  )

elastic_keys <- mapply(make_el_key, elastic_grid$a, elastic_grid$b)

sensitivity_elastic <- tibble(
  Method = "Elastic",
  Parameter = sprintf("a=%.2f,b=%.2f", elastic_grid$a, elastic_grid$b),
  Parameter_Value = seq_len(nrow(elastic_grid)),
  a = elastic_grid$a,
  b = elastic_grid$b,
  Mean = sapply(elastic_keys, function(k) elastic_results[[k]]$mean),
  CI_Lower = sapply(elastic_keys, function(k) elastic_results[[k]]$ci_lower),
  CI_Upper = sapply(elastic_keys, function(k) elastic_results[[k]]$ci_upper),
  g_T = sapply(elastic_keys, function(k) elastic_results[[k]]$g_T),
  ESS = sapply(elastic_keys, function(k) elastic_results[[k]]$ESS)
) %>%
  mutate(
    Mean = round(Mean, 4),
    CI_Lower = round(CI_Lower, 4),
    CI_Upper = round(CI_Upper, 4),
    g_T = round(g_T, 4),
    ESS = round(ESS, 3),
    CI_Width = round(CI_Upper - CI_Lower, 4)
  )

sensitivity_nnhm <- tibble(
  Method = "NNHM",
  Parameter = paste0("sigma_sigma=", sprintf("%0.2f", sigma_sigma_values)),
  Parameter_Value = sigma_sigma_values,
  Mean = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_key(
x)]]$mean),
  CI_Lower = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_
key(x)]]$ci_lower),
  CI_Upper = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_
key(x)]]$ci_upper),
  sigma_posterior = sapply(sigma_sigma_values, function(x) nnhm_results[[make
_sigma_key(x)]]$sigma_posterior),
  ESS = sapply(sigma_sigma_values, function(x) nnhm_results[[make_sigma_key(x
)]]$ESS)
) %>%
  mutate(
    Mean = round(Mean, 4),
    CI_Lower = round(CI_Lower, 4),
    CI_Upper = round(CI_Upper, 4),
    sigma_posterior = round(sigma_posterior, 4),
    ESS = round(ESS, 3),
    CI_Width = round(CI_Upper - CI_Lower, 4)
  )

```

```
sensitivity_full <- bind_rows(
  sensitivity_rmap %>% select(Method, Parameter, Parameter_Value, Mean, CI_Lo
wer, CI_Upper, ESS, CI_Width),
  sensitivity_sam %>% select(Method, Parameter, Parameter_Value, Mean, CI_Low
er, CI_Upper, ESS, CI_Width),
  sensitivity_elastic %>% select(Method, Parameter, Parameter_Value, Mean, CI
_Lower, CI_Upper, ESS, CI_Width),
  sensitivity_nnhm %>% select(Method, Parameter, Parameter_Value, Mean, CI_Lo
wer, CI_Upper, ESS, CI_Width)
)

# =====
==
# 12. SAVE
# =====
==

save_path <- "C:/Users/marin/Downloads/D-dimer results/"
if (!dir.exists(save_path)) dir.create(save_path, recursive = TRUE)

all_results <- list(
  current_study = current,
  historical_studies = historical,
  historical_effective = historical_eff,
  map = map_results,
  no_borrowing = no_borrow_results,
  rmap = rmap_results,
  sam = sam_results,
  elastic = elastic_results,
  nnhm = nnhm_results,
  summary_table = summary_table,
  sensitivity_rmap = sensitivity_rmap,
  sensitivity_sam = sensitivity_sam,
  sensitivity_elastic = sensitivity_elastic,
  sensitivity_nnhm = sensitivity_nnhm,
  sensitivity_full = sensitivity_full,
  parameters = list(
    w_values = w_values,
    Delta_values = Delta_values,
    elastic_grid = elastic_grid,
    sigma_sigma_values = sigma_sigma_values
  )
)

write.csv(
  summary_table,
  file.path(save_path, "bayesian_borrowing_summary_continuous.csv"),
  row.names = FALSE
)
```

```

)

write.csv(
  sensitivity_full,
  file.path(save_path, "sensitivity_analysis_200scenarios_continuous.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_rmap,
  file.path(save_path, "sensitivity_rmap_50_continuous.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_sam,
  file.path(save_path, "sensitivity_sam_50_continuous.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_elastic,
  file.path(save_path, "sensitivity_elastic_50_continuous.csv"),
  row.names = FALSE
)

write.csv(
  sensitivity_nnhm,
  file.path(save_path, "sensitivity_nnhm_50_continuous.csv"),
  row.names = FALSE
)

save(
  all_results,
  file = file.path(save_path, "bayesian_borrowing_complete_200scenarios_continuous.RData")
)

cat("\nAll files saved to", save_path, "\n")

```