A Comparative Evaluation of Hierarchical Bayesian, Empirical Bayesian, and Frequentist Mediation Analysis Using Simulated Clinical Trial Data

BASS XXXIII

33RD ANNUAL BIOPHARMACEUTICAL APPLIED STATISTICS SYMPOSIUM

NOV 3, 2026

Ibrahim Turkoz, PhD

J&J Innovative Medicine

Isaac Nuamah, PhD; Marc Sobel, PhD

Johnson&Johnson
Innovative Medicine

Agenda

- Introduction
- Motivating Example
- Frequentist Solution
- Bayesian Solutions
- Discussion

Introduction

Mediation refers to the process by which a variable (the mediator) transmits the effect of an independent variable (X) on a dependent variable (Y).

- How is the association generated?
- Does the association only exist for specific subgroups?
- Mediation and pathway analysis address the how question.

Key question: *How does X influence Y indirectly through a mediator (M)?*

Assumptions of Mediation Analysis

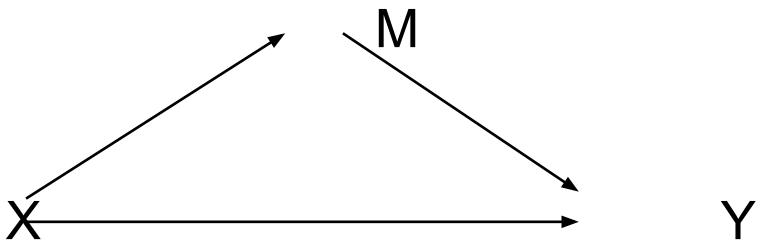
- No omitted confounders between X, M, and Y.
- Temporal order: X influences M, which influences Y.

Limitations and Challenges

- Causal Interpretation: Mediation analysis does not confirm causality.
- Measurement Error: Can bias the estimates of mediated effects.
- Complex Mediation: When multiple mediators or indirect effects are involved, more sophisticated models like Structural Equation Modeling (SEM) may be needed.

Introduction

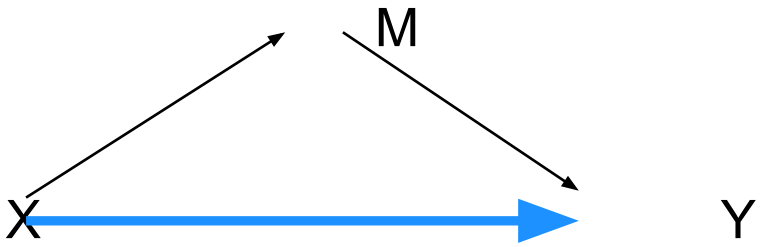
A mediator is a covariate that mediates the association between X and Y.



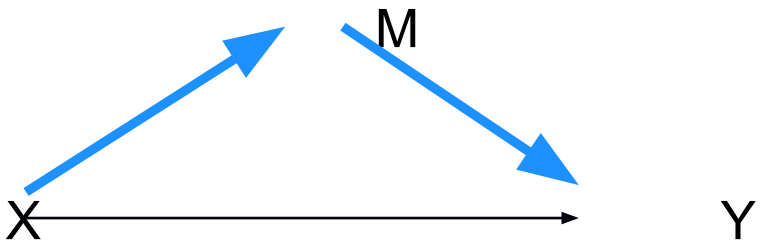
Part of the effect of X on Y is due to the fact that X influences M, which in turn affects Y.

Introduction

We aim to disentangle the total effect of X on Y into a direct effect that goes through all possible pathways but M.



… and an indirect effect that goes through M



Introduction

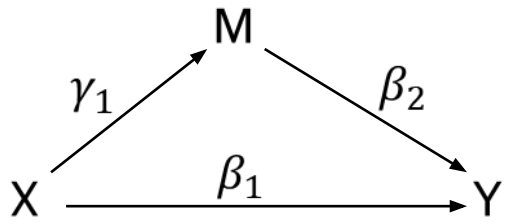
A very basic approach for mediation analysis is the difference method or use of product method based on fitting two regression models.

Let *X* be the exposure of interest, *Y* a continuous outcome, *M* a potential mediator, and *C* a set of confounders:

$$E[Y|x,m,c] = \beta_0 + \beta_1 x + \beta_2 m + \beta_3^T \mathrm{c}$$
$$E[M|x,c] = \gamma_0 + \gamma_1 x + \gamma_3^T \mathrm{c}$$

The indirect effect is calculated by taking the product of the effect of the exposure on the mediator times the effect of the mediator on the outcome.

Direct Effect = β_1
Indirect Effect = $\gamma_1 . \beta_2$
Total Effect = $DE + IE$



The product method is the common choice as it can be extended to binary as well as time-to-event outcomes.

Clinical Question – Related to Secondary Endpoints

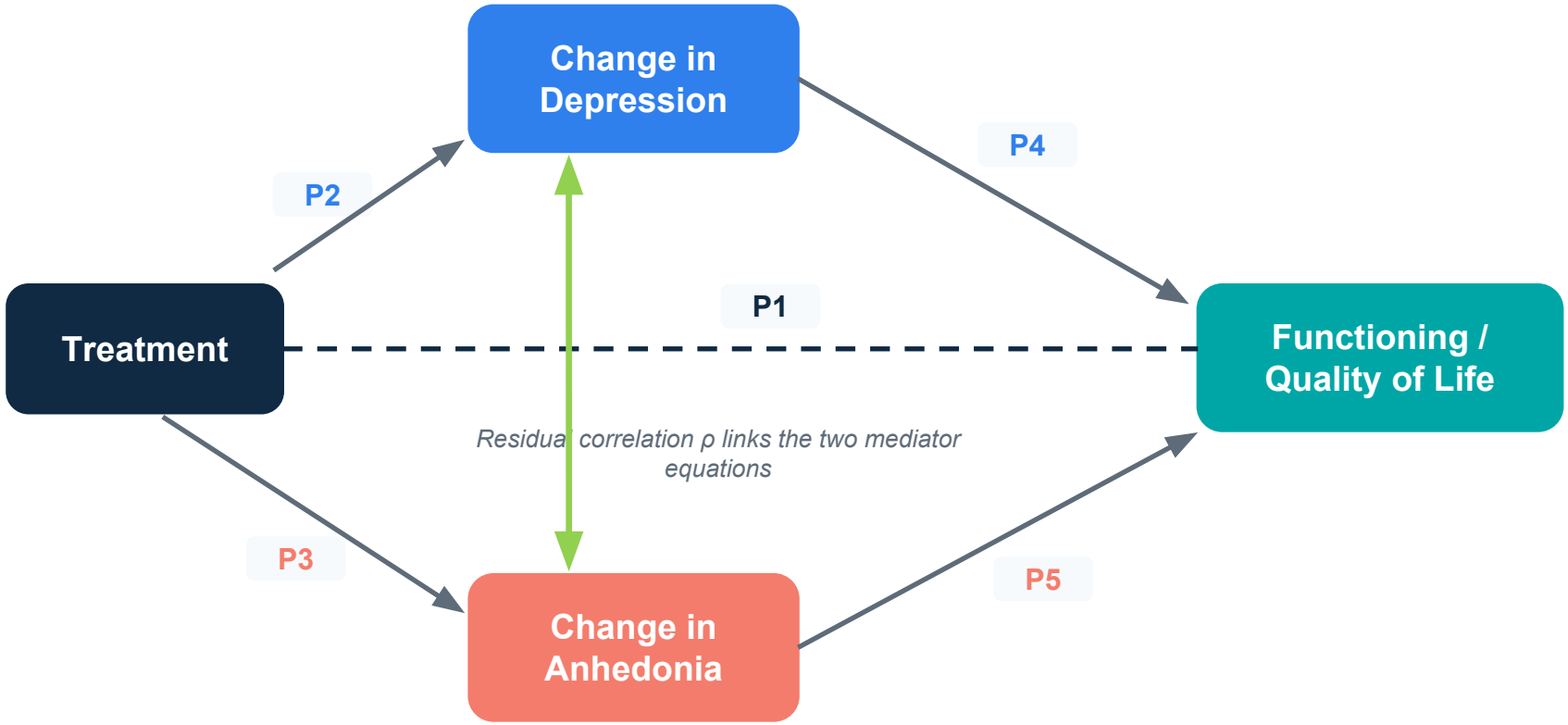
We evaluated improvements in a secondary endpoint, functional outcome (Quality of Life Scale (QLS)) in patients with Major Depressive Disorder (MDD). Patients treated with study treatment were compared to those given a placebo, while accounting for the magnitude of direct and indirect treatment effects on QLS total score (with indirect effects mediated through changes in MADRS and SHAPS total scores).

Mediation analysis was used to separate the total effects of treatment into direct and indirect impacts on QLS symptoms. Path coefficients were calculated to assess the direct and indirect effects of treatment compared to placebo.

The dependent variable was the change in Functioning total score at End Point, while the independent variables were the changes in MADRS (depression) and SHAPS (anhedonia) total scores at End Point. These variables were considered potential mediators of change in QLS total score based on clinical relationships suggested by experts. Additionally, potential baseline moderators such as age, gender, and baseline severity scores were included in the models to account for their influence on QLS score changes.

Clinical Question

Decomposing the total treatment effect into direct and indirect pathways



Effect decomposition

Direct effect

P1

Indirect through depression

P2 × P4

Indirect through anhedonia

P3 × P5

Total effect

P1 + P2×P4 + P3×P5

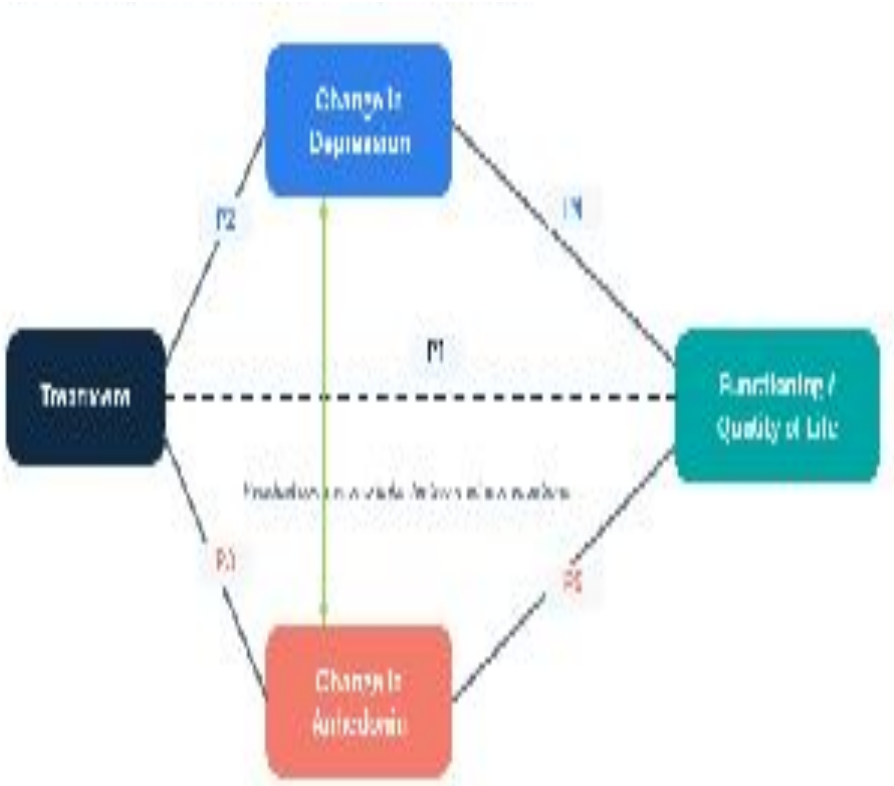
The total treatment effect is partitioned into one direct pathway and two mediator-specific indirect pathways.

J&J Innovative Medicine

Clinical Question

System of Regression Models		
	Change	Equation
Model 1	Quality-of-Life Score	BL+ Age + Gender + P1*Trt + P4*Δ_MADRS + P5*Δ_SHAPS
Model 2	Depression	BL+ Age + Gender + P2*Trt
Model 3	Anhedonia	BL+ Age + Gender + P3*Trt
The total effect of treatment is determined by P1 +(P2*P4)+(P3*P5).		

BL: baseline score of the respective dependent variable; Trt: Treatment; Δ_MADRS: Change in depression; Δ_SHAPS: Change in anhedonia; P1: Direct of treatment effect; P2 through P5: Coefficients of the intermediate paths.



Estimation Strategies Compared

Three approaches applied to the same linked mediation models

1 Frequentist OLS

Separate regression models provide a familiar benchmark for estimating P1–P5 and the derived
pathway contributions.

2 Empirical Bayes

Data-derived prior centers and variances are estimated from an independent preliminary simulation set, then treated as fixed in the main analyses.

3 Hierarchical Shrinkage Bayes

Coefficient-specific precision parameters allow the amount of regularization to adapt separately across regression coefficients.

J&J Innovative Medicine

Why Bayesian Estimation Strategies?

Motivation for empirical Bayes and coefficient-specific hierarchical shrinkage

Why use Bayesian regularization in this mediation setting?

• Correlated mediators can make pathway-specific coefficients unstable because the mediators explain overlapping portions of outcome variability.

• Small-to-moderate sample sizes can produce uncertain coefficient estimates and wide intervals, especially for pathway-specific effects formed from products of coefficients.

Why empirical Bayes?

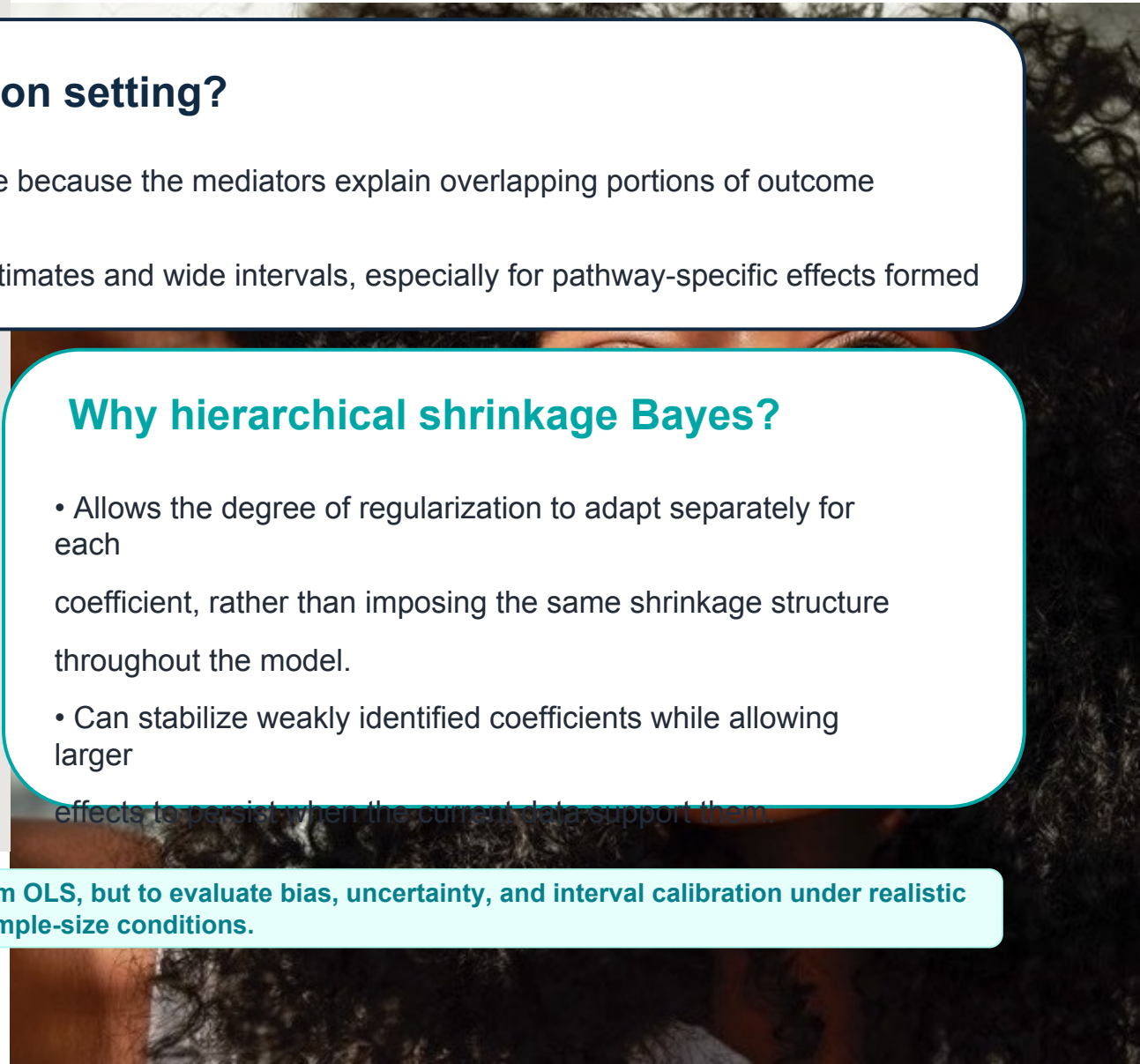
• Uses information from an independent preliminary simulation set to construct empirical prior centers and variance estimates.

• Provides fixed shrinkage toward an empirical center and a posterior distribution for coefficients and derived pathway contributions.

Why hierarchical shrinkage Bayes?

• Allows the degree of regularization to adapt separately for each coefficient, rather than imposing the same shrinkage structure throughout the model.

• Can stabilize weakly identified coefficients while allowing larger effects to persist when the current data support them.



The objective is not to assume that a Bayesian method will always outperform OLS, but to evaluate bias, uncertainty, and interval calibration under realistic correlation and sample-size conditions.

Simulation Design

- Data generating model based on the MDD trial structure.

- Two mediator correlation regimes: **ρ** = 0.3 (moderate) and **ρ** = 0.7 (strong).

- Moderate sample sizes, realistic covariates, many Monte Carlo replicates.

- Metrics: bias, interval coverage (nominal 90%), interval width (stability).

Evaluation Metrics

Performance assessed across repeated simulated clinical trials

1 Mean Absolute Error

$$MAE = (1/M) \Sigma |\hat{\theta}_j - \theta_j|$$

Accuracy of point estimation

Lower MAE indicates estimates are, on average, closer to the true simulated parameter.

2 Coverage Probability

Pr(true θ lies within the 90% interval)

Interval calibration

Calculated as the proportion of 90% confidence or credible intervals containing the true value.

3 Interval Width

Average upper limit − lower limit

Precision / stability

Narrower intervals indicate greater precision, but width should be interpreted together with coverage.

Evaluated separately by:

- **Mediator correlation** ρ = 0.3 and 0.7
- **Sample size** N = 100 and 200
- **Replicates** M = 100 trials
- **Interval level** Nominal 90%

Together, the metrics quantify the trade-off among estimation accuracy, interval calibration, and precision.

Solutions from Frequentist Perspective

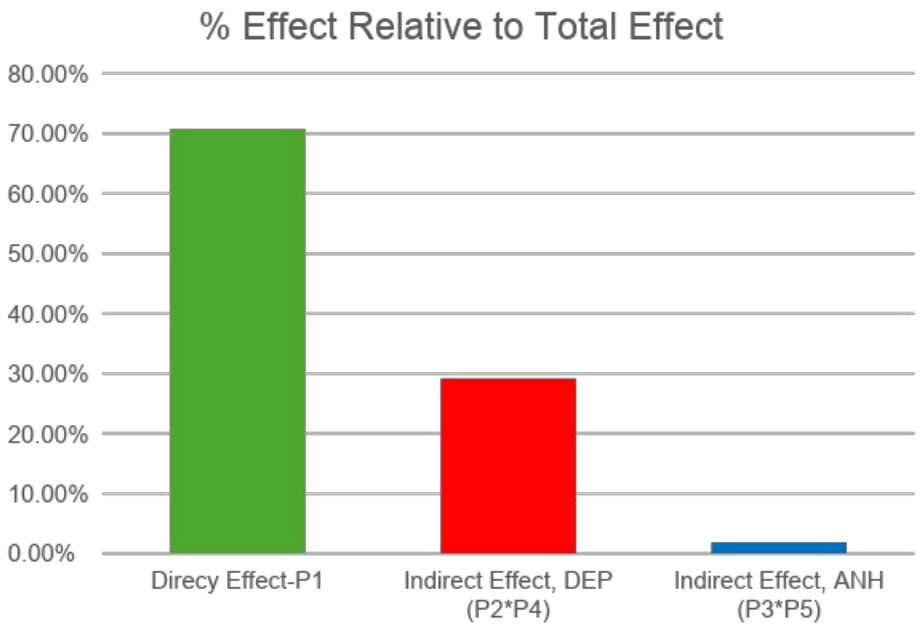
Linear Equations

	Parameter	Estimate	Standard Error	t Value	P-value
Effect					
Effect of Treatment on Change in QLS Total	P1	-0.631	0.657	-1.008	0.322
Effect of Treatment on Change in MADRS Total	P2	-2.170	1.630	-1.425	0.156
Effect of Treatment on Change in SHAPS Total	P3	-0.747	0.950	-0.790	0.460
Effect of Change in MADRS Total on Change in QLS Total	P4	0.120	0.040	3.266	0.001
Effect of Change in SHAPS Total on Change in QLS Total	P5	0.005	0.058	0.100	0.950

Total Effect of Treatment on Change in QLS Total	**-0.891**	**% Contribution**
	Direct Effect, P1 =-0.631	**70.8%**
	P2*P4 = -0.260	**29.2%**
	P3*P5 = -0.004	**0.30%**

Fit Summary

R-square	0.244
Goodness of Fit Index (GFI)	0.916



Model Specification

Three linked regression models for one outcome and two correlated mediators

Model 1: Functioning outcome

$$Z_i = L_i^T \Omega + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Ω contains baseline, age, gender, treatment P1, and mediator coefficients P4 and P5.

Model 2: Depression mediator

$$W_{2i} = F_{2i}^T R_2 + \eta_{2i}$$

R_2 contains baseline covariates and the treatment coefficient P2.

Model 3: Anhedonia mediator

$$W_{3i} = F_{3i}^T R_3 + \eta_{3i}$$

R_3 contains baseline covariates and the treatment coefficient P3.

Correlated mediator errors

$$(\eta_{2i}, \eta_{3i})^T \sim N_2(0, \Sigma)$$

Σ = covariance matrix
ρ = residual correlation between the two
mediator equations

Pathway effects

Direct: P1 **Indirect: P2×P4 and P3×P5**

The three equations define the same pathway structure used by OLS, empirical Bayes, and hierarchical Bayes.

J&J Innovative Medicine

Solutions from Bayesian Perspective

Likelihood and priors combine to produce posterior distributions

1 Likelihood

Information from the current dataset

×

2 Prior distributions

Information and regularization specified before updating

∝

3 Posterior

Updated distribution for coefficients and variance

Likelihood for Model 1

$$L(Z \mid \Omega, \sigma^2) \propto (\sigma^2)^{-n/2} \times \exp\{-[\Sigma_i(Z_i - L_i^T\Omega)^2]/(2\sigma^2)\}$$

Current data determine the fit of Ω and σ².

Prior specification

$$\Omega \mid \sigma^2 \sim N(\Omega_0, \sigma^2 V_0)$$
$$\sigma^2 \sim IG(a_0, b_0)$$

EB estimates Ω_0 and V_0 from preliminary simulations.

Posterior distribution

$$p(\Omega, \sigma^2 \mid Z) \propto L(Z \mid \Omega, \sigma^2) \times p(\Omega \mid \sigma^2) \times p(\sigma^2)$$

Posterior draws provide point estimates and credible intervals.

Bayesian estimation produces a distribution of plausible values rather than only a single coefficient estimate.

J&J Innovative Medicine

Solutions from Bayesian Perspective

$$LIKE_1 \propto (1/\sigma_1^2)^{n/2} \exp\left\{-\sum_{i=1}^{n} (Z[i] - L[i]'\Omega)^2 / 2\sigma_1^2\right\}$$

[illegible] =

$$\Omega = (P_{BS,1}, P_{A,1}, P_{G,1}, P_1, P_4, P_5)'$$

[illegible] = [illegible] = 1, … , [illegible]

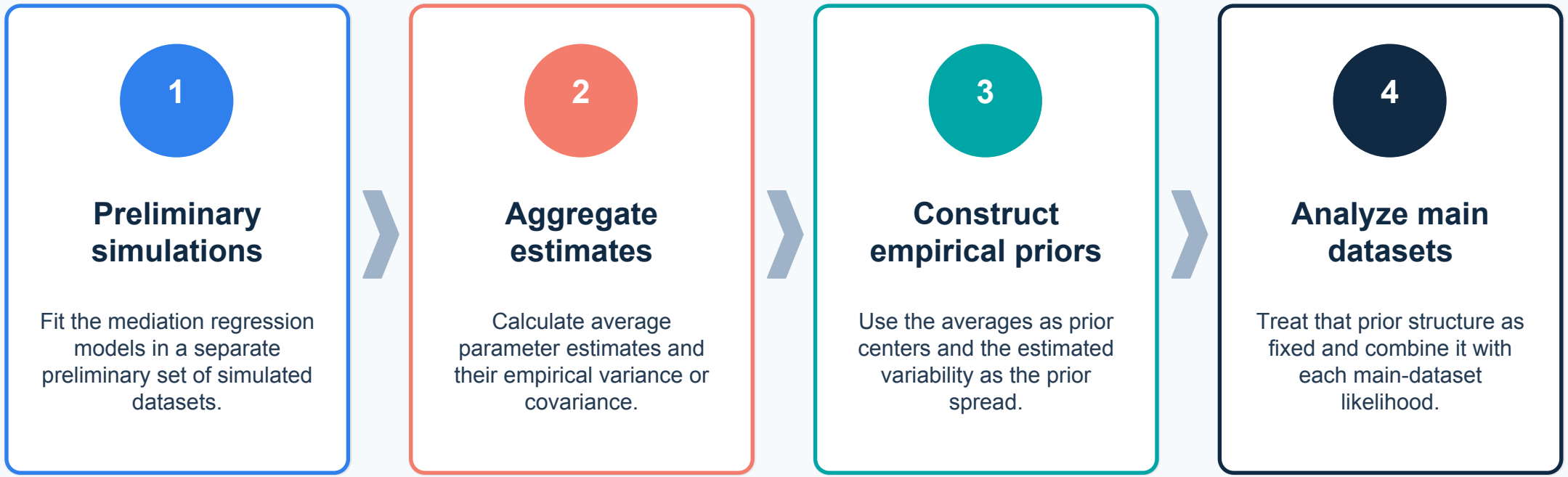
$$(\Omega|\sigma^2) \sim \mathcal{N}(\widehat{\Omega}, \sigma^2/\alpha_0)$$

$$(\sigma^2) \sim \mathcal{IG}(a_0, b_0)$$

$$(Z, \Omega^{(j)}, [\sigma^2]^{(j)}) \propto \frac{\exp\left\{-(1/2)\frac{\sum_{i=1}^{n}(Z_i - \Omega^{(j)} L_i)^2}{[\sigma^2]^{(j)}}\right\}}{([\sigma^2]^{(j)})^{(n)/2}} * \exp\left(-\frac{b_0}{[\sigma^2]^{(j)}}\right) \frac{1}{[[\sigma^2]^{(j)}]^{a_0+1}} * \exp\left\{-(1/2)\alpha_0 \frac{(\Omega^{(j)} - \widehat{\Omega})'(\Omega^{(j)} - \widehat{\Omega})}{[\sigma^{(j)}]^2}\right\}$$

Empirical Bayes: where the prior comes from

Prior information is estimated empirically before analyzing the main simulation datasets



Important for this paper: this is a controlled simulation comparison. The empirical prior is estimated from preliminary simulations, not from external historical clinical data.

How an empirical Bayes estimate is generated

The posterior is a precision-weighted compromise between current data and a fixed empirical prior

Current-data estimate

Likelihood contribution
Estimate and standard error from the current dataset

+

Empirical prior

Fixed prior center and spread estimated from preliminary simulations

→

Posterior estimate

Stabilized estimate plus a posterior credible interval

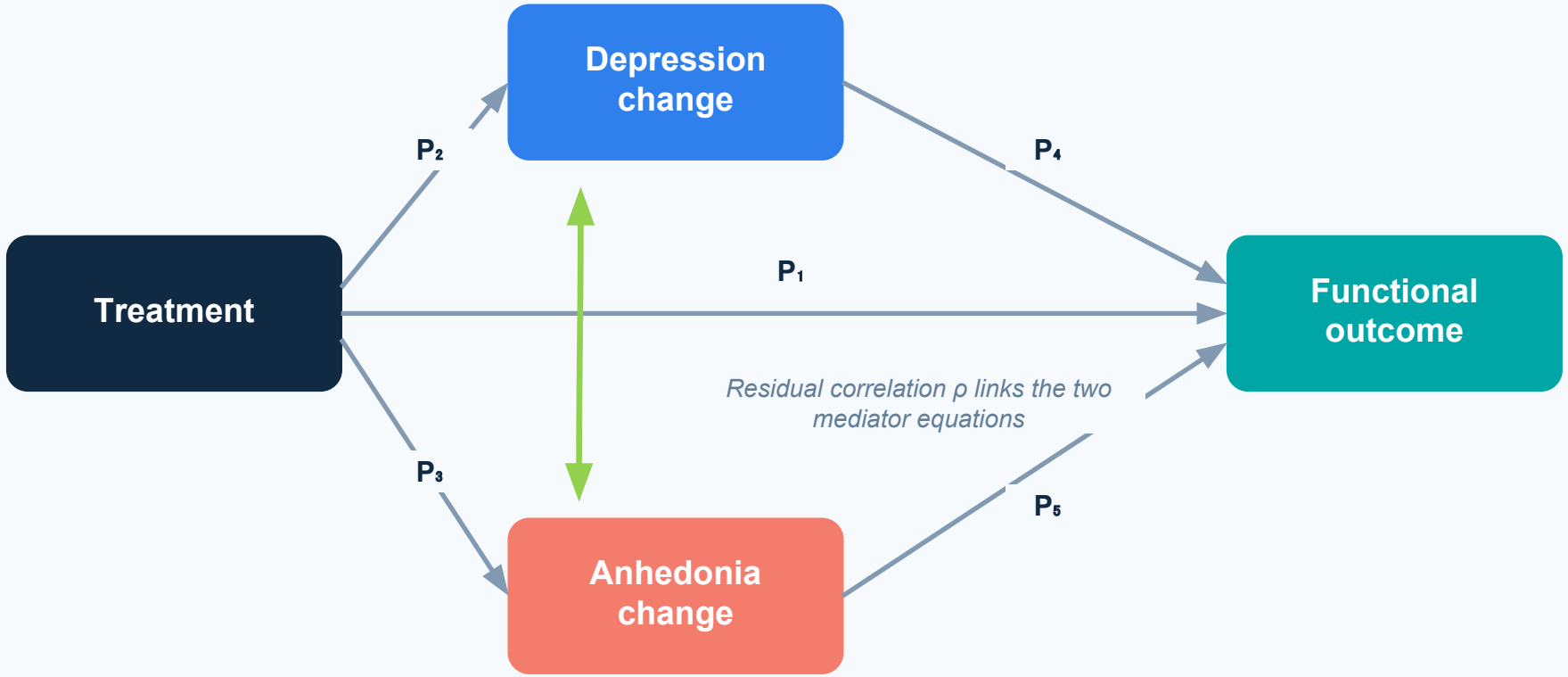


More precise current data → larger w → less shrinkage
Less precise current data → smaller w → more shrinkage

Empirical Bayes does not simply average two estimates; it gives more influence to the more precise source.

Empirical Bayes in the correlated-mediator model

Fixed empirical priors regularize the pathway coefficients while preserving the model structure



What EB supplies

- Prior centers for the regression coefficients
- Fixed prior variance or covariance from preliminary simulations
- Posterior draws obtained by combining each prior with the current likelihood

The pathway model stays the same; EB changes how uncertain coefficients are estimated by shrinking them toward empirical prior centers.

The bridge from empirical Bayes to hierarchical shrinkage

Same prior center, more flexible learning of regularization



Hierarchical shrinkage retains empirical calibration but learns the amount of shrinkage coefficient by coefficient.

What makes the model hierarchical?

A prior is placed on another prior parameter

LEVEL 1

Likelihood

The current data inform the regression coefficients and residual variance.

LEVEL 2

Coefficient prior

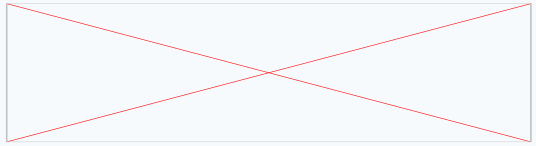
LEVEL 3

Hyperprior on shrinkage



The shrinkage parameter α is a precision

Precision is the inverse of variance



Large α□

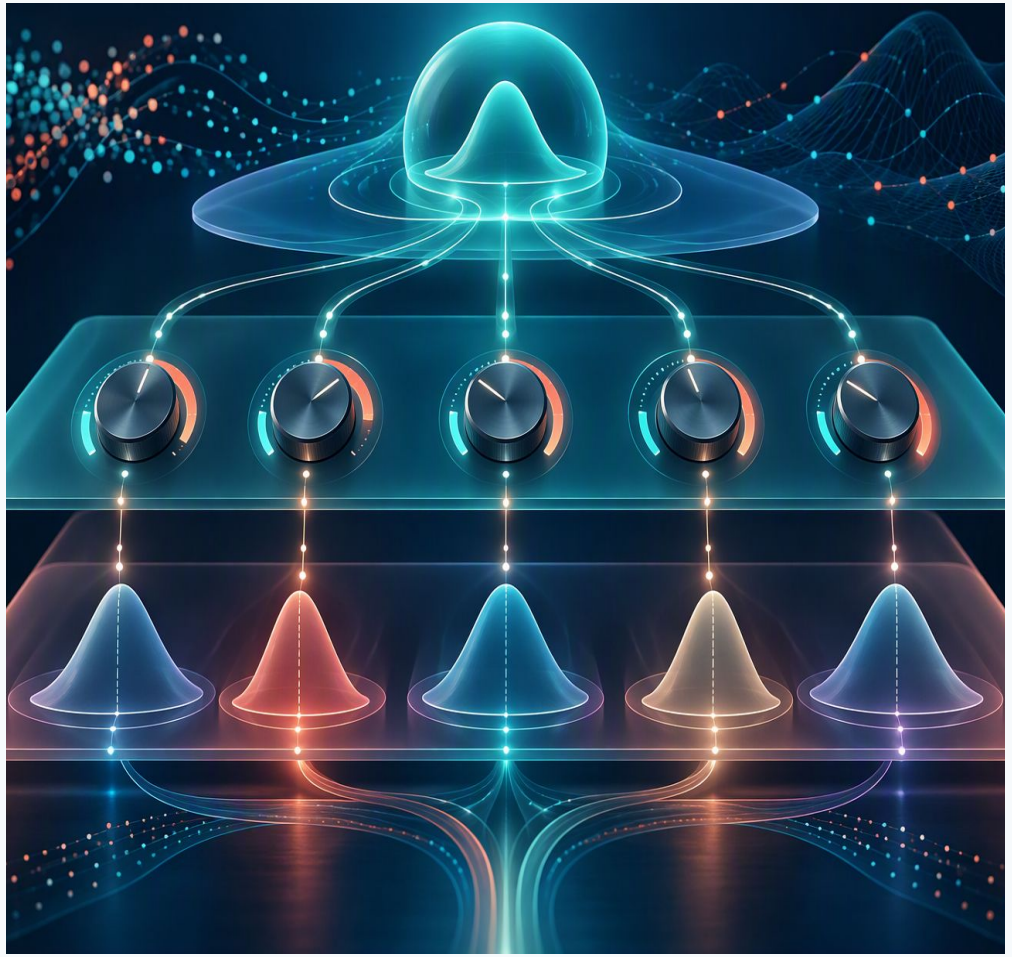
Small prior variance

Stronger pull toward the prior center

Small α□

Large prior variance

Weaker pull toward the prior center



α□ does not measure the effect size; it controls how tightly β□ is regularized around μ□.

How a coefficient estimate is generated

The posterior combines likelihood precision and prior precision

Current data

Estimate from the likelihood with uncertainty determined by sample size, design, and correlation

+

Hierarchical prior

Centered at empirical μ□ with current shrinkage precision α□

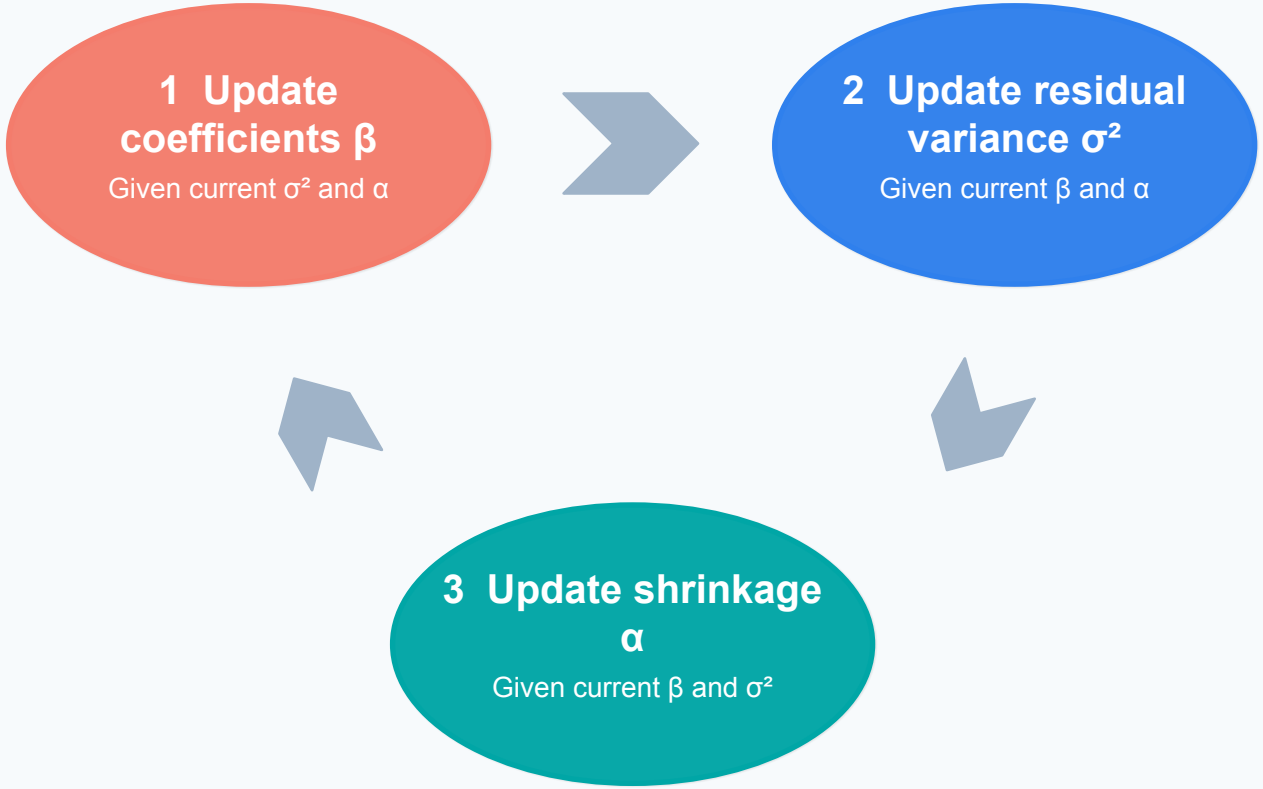
Posterior for β□

A full distribution of plausible values, summarized by a posterior mean or median and a credible interval

The data and prior are weighted by precision: the more precise source has greater influence.

Gibbs sampling turns the hierarchy into estimates

Each update uses the latest values of the other parameters



Repeat 5,000 iterations

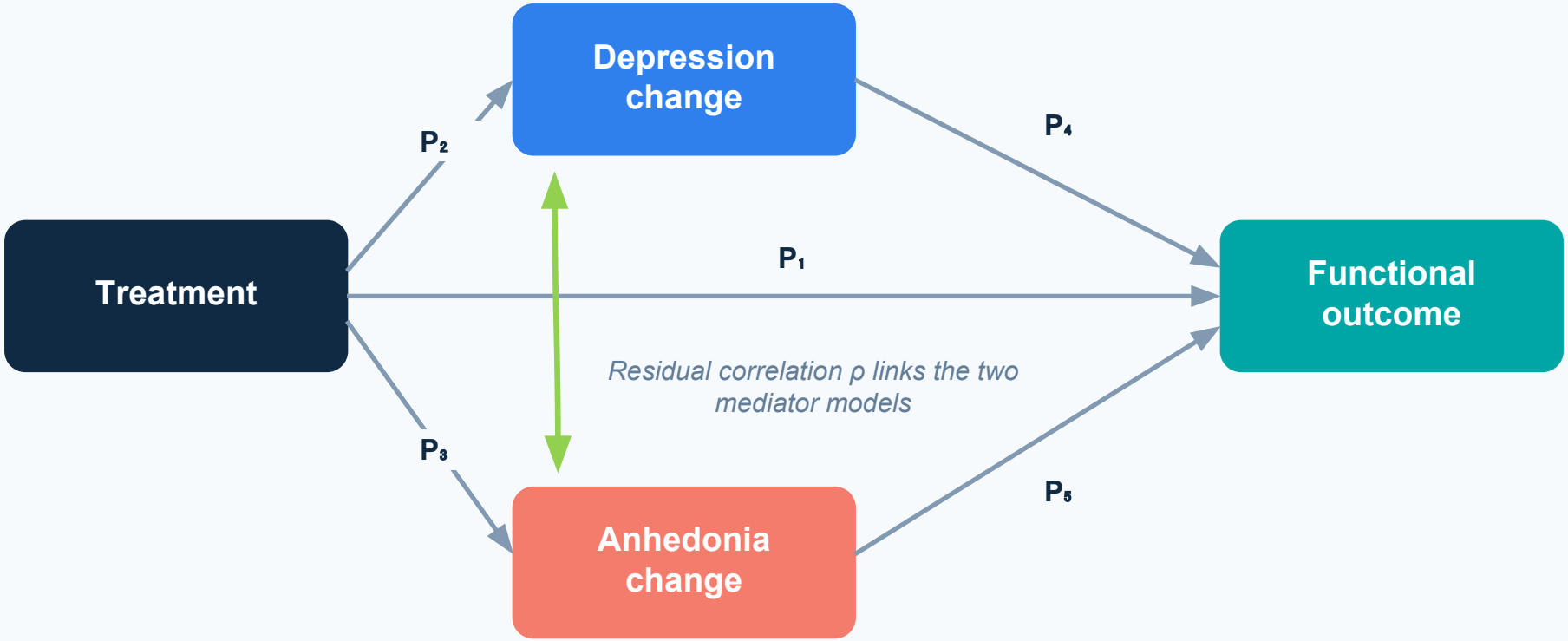
Discard first 1,000 as burn-in
Use remaining draws to approximate each posterior distribution

$$\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(4000)}$$

The final estimate is a summary of posterior draws, not the value from a single iteration.

Why coefficient-specific shrinkage matters here

The pathway coefficients are not equally identified



A separate α□ for each coefficient

$P_1 \leftrightarrow \alpha_1$

$P_2 \leftrightarrow \alpha_2$

$P_3 \leftrightarrow \alpha_3$

$P_4 \leftrightarrow \alpha_4$

$P_5 \leftrightarrow \alpha_5$

Different pathways can receive different levels of regularization, while correlation still governs identifiability.

Fixed versus adaptive shrinkage

The distinction to emphasize in the presentation

Empirical Bayes

- Empirically calibrated prior center
- Prior variance structure treated as fixed
- Same analysis combines prior + current data
- Often more conservative intervals in this study

Hierarchical shrinkage

- Empirically calibrated prior center
- α_j is learned for each coefficient
- Adaptive, parameter-specific regularization
- Often narrower intervals for selected parameters

Neither method uniformly dominates

Three messages to leave with you

1

Hierarchy means learning the shrinkage

The model estimates β, σ^2, and α together rather than fixing the amount of regularization in advance.

2

Shrinkage is coefficient-specific

Each α_j controls the prior precision for one coefficient, allowing stronger or weaker regularization across pathways.

3

Inference comes from posterior draws

Repeated Gibbs updates produce a posterior distribution summarized by a mean or median and a credible interval.

EB fixes the shrinkage structure; hierarchical shrinkage learns a different amount of shrinkage for each coefficient.

Result Summary: Mean Absolute Error

Table 1: MAE for EB, HS, and ML estimators by sample size and mediator correlation

N	ρ	Method	P1	P2	P3	P4	P5	π1	π2	π3
200	.7	EB Prior	.118	.149	.134	.009	.008	.021	.025	.021
200	.3	EB Prior	.092	.090	.083	.005	.008	.021	.025	.021
200	.7	MLE	.108	.145	.130	.009	.009	.021	.024	.024
200	.3	MLE	.085	.094	.082	.006	.008	.021	.024	.024
200	.7	HS Prior	.116	.147	.137	.008	.008	.022	.024	.023
200	.3	HS Prior	.087	.087	.084	.006	.008	.022	.025	.023
100	.7	EB Prior	.139	.189	.207	.008	.011	.022	.023	.027
100	.3	EB Prior	.111	.102	.124	.011	.012	.023	.016	.022
100	.7	MLE	.145	.231	.206	.008	.010	.024	.022	.028
100	.3	MLE	.146	.106	.134	.010	.012	.021	.016	.021
100	.7	HS Prior	.141	.209	.191	.008	.010	.021	.022	.028
100	.3	HS Prior	.132	.103	.127	.010	.012	.021	.016	.021

Mean absolute error across M=100 simulated clinical trials.

Key patterns

- Increasing N from 100 to 200 generally reduces MAE for P1, P2, and P3 across methods and correlation regimes.
- Higher correlation has its clearest adverse effect on P2 and P3; P4 and P5 remain small across conditions.
- Differences among EB, MLE, and HS are modest relative to the effects of sample size, correlation, and parameter.
- π1–π3 MAEs remain small, with no method uniformly best across scenarios.

Result Summary: Interval Coverage

Table 2: Coverage of nominal 90% Bayesian credible intervals by sample size and correlation

N	ρ	Method	P1	P2	P3	P4	P5
200	.7	Empirical Bayes Prior	.960	.880	.910	.980	.980
200	.3	Empirical Bayes Prior	.962	.870	.920	.970	.980
200	.7	Hierarchical Shrinkage Prior	.880	.880	.910	.800	.930
200	.3	Hierarchical Shrinkage Prior	.870	.860	.910	.880	.910
100	.7	Empirical Bayes Prior	.960	.890	.850	.960	.980
100	.3	Empirical Bayes Prior	.950	.910	.920	.920	.930
100	.7	Hierarchical Shrinkage Prior	.930	.870	.870	.890	.910
100	.3	Hierarchical Shrinkage Prior	.900	.910	.900	.850	.860

Coverage probability based on M = 100 simulated clinical trials.

Key patterns

- EB coverage is equal to or greater than HS for
Nearly every parameter within each simulation
condition.
- EB consistently overcovers P1, while HS coverage
for P1 shifts from about nominal at N = 100 to under coverage at N = 200.
- At N = 100 and ρ = .7, EB coverage for P3 is 0.850,
showing sensitivity to the combination of small Sample and high correlation.
- Because M = 100 trials, a 0.01 coverage change
represents one simulation; small differences require cautious interpretation.

Result Summary: Interval Width & Stability

Table 3: Average length of 90% credible intervals by sample size and mediator correlation

N	ρ	Method	P1	P2	P3	P4	P5
200	.7	Empirical Bayes Prior	.689	.669	.664	.050	.070
200	.3	Empirical Bayes Prior	.559	.389	.384	.050	.071
200	.7	Hierarchical Shrinkage Prior	.453	.663	.661	.033	.045
200	.3	Hierarchical Shrinkage Prior	.349	.387	.383	.032	.045
100	.7	Empirical Bayes Prior	.893	.948	.940	.062	.084
100	.3	Empirical Bayes Prior	.622	.549	.560	.067	.090
100	.7	Hierarchical Shrinkage Prior	.651	.941	.934	.046	.061
100	.3	Hierarchical Shrinkage Prior	.561	.543	.564	.046	.062

Average credible-interval length across M= 100 simulated trials.

Key patterns

- Intervals generally narrow when N increases from 100 to 200 under both priors and correlation regimes.
- P2 and P3 widen substantially at ρ = .7; EB and HS widths are nearly identical within each condition.
- HS intervals are consistently narrower than EB intervals for P1, with the largest differences under ρ = .7.
- P4 and P5 widths are small and relatively insensitive to correlation; HS provides additional interval shortening.

Solutions from Bayesian Perspective for Proportional Effects

Proportional effects were computed using MLE and both EB and Hierarchical Bayes based on Models 1 , 2, and 3.

$$abs(P_1) = |P_1|$$

$$abs(P_2 * P_4) = |P_2 P_4|$$

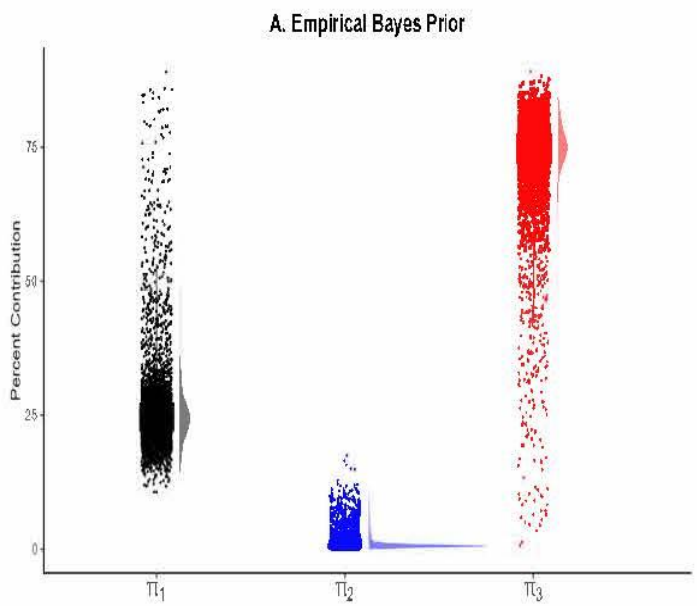
$$abs(P_3 * P_5) = |P_3 P_5|$$

% Direct Effect by P_1 : π1 : $F_1^{(j)} = \dfrac{abs(P_1^{(j)})}{abs(P_1^{(j)}) + abs(P_2^{(j)} P_4^{(j)}) + abs(P_3^{(j)} P_5^{(j)})}$

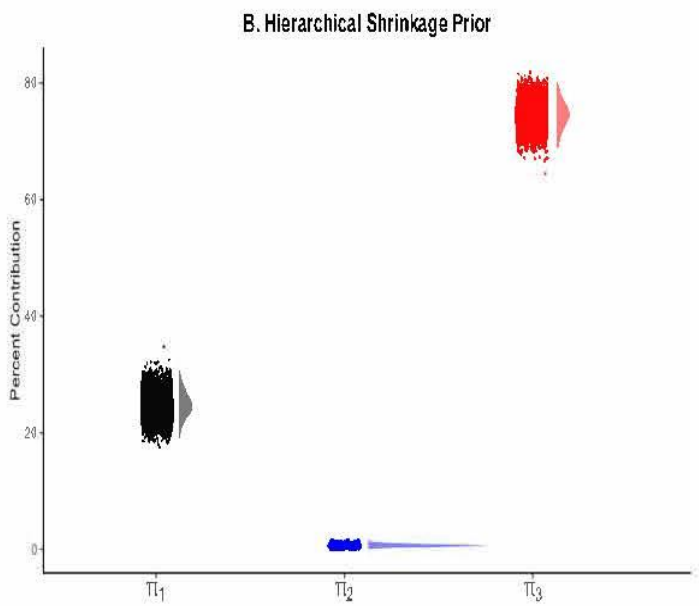
*% Indirect Effect by P_2*P_4 : π2::* $F_2^{(j)} = \dfrac{abs(P_2^{(j)} P_4^{(j)})}{abs(P_1^{(j)}) + abs(P_2^{(j)} P_4^{(j)}) + abs(P_3^{(j)} P_5^{(j)})}$

*% Indirect Effect by P_3*P_5 : π3:* $F_3^{(j)} = \dfrac{abs(P_3^{(j)} P_5^{(j)})}{abs(P_1^{(j)}) + abs(P_2^{(j)} P_4^{(j)}) + abs(P_3^{(j)} P_5^{(j)})}$

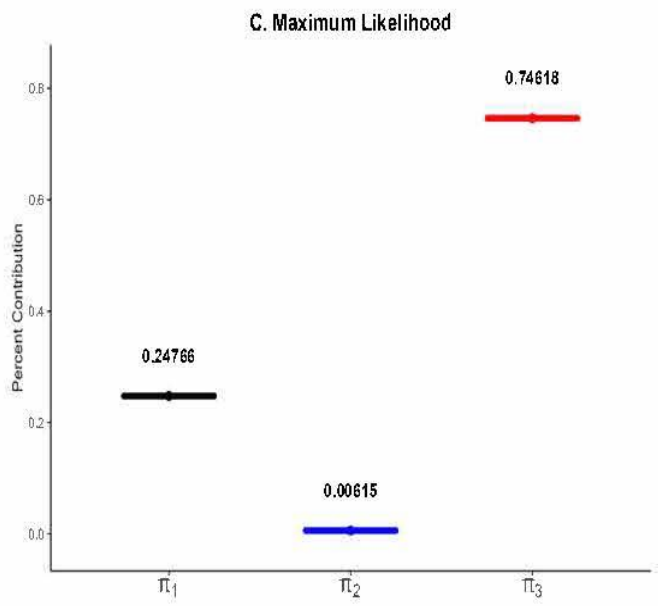
Solutions – Local Trial



Group	Median	Mean	P25	P75
π1	24.58	26.14	22.19	27.14
π2	0.65	1.09	0.46	0.91
π3	74.78	72.78	72.12	77.20



Group	Median	Mean	P25	P75
π1	24.70	24.77	23.30	26.20
π2	0.63	0.65	0.48	0.79
π3	74.65	74.59	73.13	76.10



Frequentist Perspective

Discussion

- There was NO universal winner. Presenting the totality of evidence within a regulatory context can offer a more comprehensive and informed response, especially when navigating complex datasets and topics.

- A Bayesian analysis framework is conceptually intuitive, as it allows for **a natural distribution** around indirect effects and quantifies the percentage contribution of various factors. This approach is particularly powerful because it integrates prior information, leading to improved statistical power and more robust conclusions.

- In cases of multilevel mediation analysis, the Bayesian method proves especially advantageous. Unlike traditional frequentist approaches, which struggle with multilevel structures, Bayesian analysis handles this complexity with ease.

- Moreover, one of the key strengths of Bayesian mediation analysis is its ability to provide exact inference, even with small sample sizes. This contrasts with frequentist methods, which rely on large-sample approximations and may falter when sample sizes are limited.

- Use Empirical Bayes
  - Many parameters
  - Little prior information
  - Routine estimation

Use Hierarchical Bayes
- Uncertainty matters
- Adaptive trials
- Subgroup estimation

Thank you!

Backup

Aspect	Uninformative Prior	Weakly Informative Prior	Robust Prior
Definition	Provides little to no prior information.	Incorporates some prior knowledge, but vague.	Prior designed to be less sensitive to model assumptions and outliers.
Purpose	Allows data to dominate the inference.	Offers guidance but still lets data influence heavily.	Ensures valid inferences even with model misspecification or outliers.
Examples	Uniform distribution over a wide range.	Normal distribution with a large standard deviation.	T-distribution or Laplace distribution, which have heavier tails.
Impact on Posterior	Primarily reflects the likelihood of the data.	Moderates influence of the data while incorporating prior knowledge.	Results in a posterior that is robust against outliers and model assumptions.
Flexibility	Highly flexible; no preconceived notions.	Somewhat flexible, but with a guiding structure.	Flexible, but more conservative regarding the impact of extreme values.
Convergence	May lead to slower convergence in certain cases.	Generally helps in quicker convergence due to guidance.	Can achieve convergence reliably even in the presence of outliers.
Interpretability	Can be difficult to justify if prior information is needed.	More interpretable as it suggests plausible values for parameters.	Can be interpreted as cautious or protective against uncertain behavior in data.
Risk of Non-Identification	May lead to non-identifiable models if not careful.	Reduces the risk of non-identifiability by constraining the parameter space slightly.	Minimizes the risk of non-identifiability in presence of outliers or extreme values.
Use Cases	Useful in exploratory analyses or when little is known.	Useful when some prior information exists but one wants to remain flexible.	Useful when data may contain outliers or when model assumptions are uncertain.
Subjectivity	Minimizes the introduction of subjective beliefs.	Introduces some subjective beliefs while aiming to remain mild.	Can still be somewhat subjective, but aims to reduce impact of this.
Data Dependence	Strongly data-dependent, allowing for pure data-driven insights.	Balances data dependence with prior knowledge for more robust estimates.	Less sensitive to specific data points and more focused on overall trends.
Computation	Simpler computationally in some models.	Can be slightly more complex due to constraints on the parameter space.	May require more complex computations due to heavier-tailed distributions.