

Headline Estimation with Multiple Research Designs*

Vod Vilfort[†]

September 1, 2026

Abstract

To study a scalar parameter, a researcher may consider multiple research designs. Based on the evidence across designs, the researcher may wish to formulate a headline estimate of the parameter. I examine how to choose this headline when it is unclear which design is most appropriate for studying the parameter. I model this setting by assuming that (i) exactly one of the designs is valid for the parameter and (ii) the researcher has ambiguity about which design is valid, represented by a class of priors over the candidate designs. To account for ambiguity, I propose reporting the headline estimate that minimizes the worst-case posterior risk over the class of priors. In three applications, I show cases where accounting for ambiguity materially affects the quantitative conclusion and cases where an existing headline is already close to optimal.

*I thank Isaiah Andrews, Anna Mikusheva, and Alberto Abadie for their guidance and support. I thank seminar participants from the MIT econometrics lunch for helpful comments and discussions. I gratefully acknowledge support from the Jerry A. Hausman Fellowship and the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. I thank ChatGPT and Codex for research assistance.

[†]Department of Economics, Massachusetts Institute of Technology, vod@mit.edu.

1 Introduction

To study a quantity of interest, a researcher may consider multiple research designs. One design might exploit a policy discontinuity, a second might use quasi-random assignment, and a third might leverage comparisons across groups or time periods. When different designs yield similar estimates, the researcher may view this as evidence for a common conclusion about the quantity of interest.¹ The researcher may then want a *headline estimate*—a single number to emphasize in the abstract, introduction, or policy discussion. I examine how to formulate this headline when it is unclear which design is most appropriate for studying the quantity of interest.

Researchers consider various approaches to the headline estimation problem. One approach is to headline a precision-weighted average of the design estimates, which gives more weight to designs with smaller standard errors (Garin et al., 2025; Jain, Mommaerts and Weaver, 2026). This can be efficient when every design is unbiased for one common effect, but that assumption is difficult to justify when different designs use distinct sources of identifying variation. A second approach is to headline the estimate from a favored design that is viewed as especially credible or policy relevant (Bhuller, Mogstad and Salvanes, 2017; Sager and Singer, 2025; Chen, Chen and Yang, 2026; Fang et al., 2026). This avoids the common-effect assumption, but gives no systematic role to other designs. A third approach is to report the range of the design estimates (Hastings and Shapiro, 2018; Alsan et al., 2023; Ganong et al., 2024; Ng and Riehl, 2024). This reveals the set of headlines that are attainable as convex averages of the design estimates, but does not determine what single number should summarize the quantitative conclusion.

In this paper, I propose a decision-theoretic solution to the headline estimation problem. I model the quantity of interest as a scalar target parameter. Each design produces an estimate of a design-specific estimand, but only one design is *target-valid* in the sense that its estimand identifies the target parameter—i.e., the target-valid design is “most appropriate” for studying the quantity of interest. The other estimands retain their design-specific interpretations and may differ from the target, although they can coincide with it numerically. The model therefore allows several “complementary” designs to illuminate a shared question without treating their estimates as unbiased measurements of a common effect.² The researcher does not know which design is target-valid, leading to ambiguity over the candidate designs. I model this ambiguity as a class of *design priors*. A design prior specifies how likely each design is to be target-valid. The class collects the set of design priors that the researcher is willing to entertain. Under this

¹Garin et al. (2025, page 504) say that “Using multiple research designs allows us to test the sensitivity of our results to empirical strategy.” Ganong et al. (2024, page 2907) say that “Each of these empirical exercises has distinct advantages and disadvantages, but they all lead to the same conclusion.” Chen, Chen and Yang (2026, page 1576) say that their results “remain remarkably consistent across both approaches.”

²The framing of designs as being “complementary” or as having “different strengths and weaknesses” is language that appears frequently in applied work: examples include Alsan et al. (2023, page 407), Dunn et al. (2024, page 217), Baran et al. (2025, page 14), Davis et al. (2025, page 1), Erten, Stiglitz and Verhoogen (2025, page 13), Han et al. (2025, page 3), and Agrawal et al. (2026, page 13).

structure, I formulate the headline estimation problem as a statistical decision problem under ambiguity (Berger, 1985; Gilboa and Schmeidler, 1989; Stoye, 2012). To estimate the target parameter while accounting for ambiguity, I propose a conditional Gamma-minimax procedure (DasGupta and Studden, 1989; Betrò and Ruggeri, 1992; Giacomini, Kitagawa and Read, 2021). The procedure conditions on the design estimates and chooses the scalar report that minimizes worst-case posterior risk (i.e., expected loss) over the class of design priors. I call this report the *ambiguity-optimal headline estimate*, or *optimal headline* for short.

I propose default specifications for the above decision problem. Under these specifications, the headline estimation procedure requires only the design estimates and their standard errors. The procedure accommodates any compact and convex class of design priors. For example, the *simplex class* is the set of all design priors. Under this class, the procedure chooses the headline that minimizes the largest standardized discrepancy from the design estimates. Equivalently, place an interval around each estimate and expand every interval by the same multiple of its own standard error. The optimal headline is the first point at which all the intervals overlap. With two designs, the optimal headline is an inverse-standard-deviation-weighted average of the two design estimates—which differs from the inverse-variance-weighted average that is used in conventional precision-weighting. With more designs, the solution is determined by the set of pairwise headline estimation problems across designs.

The simplex class represents the benchmark of unrestricted ambiguity. In some cases, the researcher may have more structured forms of ambiguity. For example, the researcher may wish to limit attention to design priors within a neighborhood of a baseline prior. To formalize this, the *baseline-prior perturbation class* begins from a baseline prior and considers the set of priors generated by ϵ -perturbations of the baseline. When the baseline puts equal probability on each design, $\epsilon = 0$ recovers the conventional inverse-variance-weighted average while $\epsilon = 1$ recovers the simplex-optimal headline. Intermediate perturbation levels trace a monotone path between those two endpoint headlines. As a second example, the researcher may have a favored design whose target-validity is deemed to be especially likely. To formalize this, the *favored-design ratio class* considers the set of design priors where the relative probability of the favored design is not too low—governed by a ratio parameter. This formalizes the practice of emphasizing one design by stating how much more plausible it must be than each alternative. As the ratio parameter grows, the optimal headline converges to the favored design estimate. These example structured ambiguity classes can serve as primary specifications when their restrictions are substantively justified or as sensitivity analyses relative to the unrestricted ambiguity benchmark.

In my framework, the performance of a candidate headline is measured by its worst-case posterior risk. Intuitively, the risk quantifies how well the design-specific evidence can support—or be reconciled with—a given headline. By construction, the optimal headline yields the minimum attainable risk. Comparing the risk of a paper’s existing headline with the minimum attainable

risk shows whether accounting for design ambiguity materially improves the headline report. The headline estimation procedure can therefore be used to support a paper’s existing headline or to rank candidate headlines while keeping the degree of reconciliation visible.

I illustrate my framework in three applications. In [Chen, Chen and Yang \(2026\)](#), the target is the effect of courtroom online broadcasting intensity on the gender gap in plaintiff win rate, and the designs are a difference-in-differences (DiD) specification and a Bartik instrumental variable (IV) strategy. Under the simplex class, the optimal headline supports the conclusion that broadcasting narrows the gap, but its high risk suggests that the DiD and IV designs do not tightly reconcile one exact magnitude. Under the favored-design ratio class for DiD—the paper’s primary design—I examine how favored DiD would have to be to obtain meaningful risk reductions. In [Garin et al. \(2025\)](#), the targets are the effects of a 12-month incarceration sentence on later incarceration and labor market outcomes, and the two designs use sentencing-guideline discontinuities in North Carolina and random judge assignments in Ohio. Under an equal-probability baseline-prior perturbation class, the paper’s precision-weighted averages for the main outcomes are already close to optimal, and the paper’s substantive conclusions are largely unchanged. In [Ganong et al. \(2024\)](#), the target is the marginal propensity to consume out of unemployment benefits, and the six designs exploit benefit-receipt delays, changes in pandemic benefit supplements, and cross-state variation in expiration timing. Under the simplex class, the optimal headline indicates a high spending response, but its high risk reveals that the exact magnitude cannot be tightly reconciled across the six designs.

I emphasize two points regarding the headline estimation framework. First, the framework takes as given that the researcher has specified a scalar target parameter and a set of designs that could plausibly supply the identifying link. It does not decide whether the designs *should* be viewed as speaking to one latent target parameter—that is a substantive judgment for the researcher to make. Second, some settings provide enough economic or structural information to combine different designs to directly identify a target parameter. In those settings, the headline estimation framework is not suitable.³

This paper relates to three literatures. First, robust Bayesian decision theory studies how to make optimal decisions under classes of prior beliefs ([Gilboa and Schmeidler, 1989](#); [Stoye, 2012](#)). I use this logic to represent ambiguity over the identity of the target-valid design and to choose a headline that minimizes worst-case posterior risk. The conditional Gamma-minimax criterion is applied conditional on the observed estimates, which distinguishes it from ex-ante Gamma-minimax criteria that evaluate decision rules based on worst-case Bayes risk. Advan-

³For example, [Laliberté \(2021\)](#) targets the share of neighborhood effects on long-term educational outcomes that operates through access to better schools, uses a spatial regression discontinuity design to identify the school effects and a movers design to identify the neighborhood effects, then combines them within an education production function decomposition to estimate the target share. The designs are different inputs to a structured identification argument rather than candidate estimates of the target.

tages of the conditional approach include analytical and numerical tractability (Christensen, Moon and Schorfheide, 2026) and favorable dynamic consistency properties (Lim and Park, 2026). A disadvantage is that the resulting decision rules are not guaranteed to be admissible (Giacomini, Kitagawa and Read, 2021).

Second, work on averaging and adaptation examines how to use a precise estimator that may be biased together with a robust estimator that is valid under weaker assumptions (Green and Strawderman, 1991; Cheng, Liao and Shi, 2019; Armstrong, Kline and Sun, 2025). Here the use of a risk-based performance criterion is related, but the present problem has no estimator known in advance to provide an unbiased anchor. Which design is target-valid is precisely what remains ambiguous.

Third, this paper contributes to recent work on how to combine evidence across models or research designs. Bhattacharya, Ocelli and Westling (2026) construct data-dependent weights for methodological triangulation using testable implications of candidate causal models. Park and Sasaki (2026) exploit an ordering among matching, DiD, and hybrid estimands to select the hybrid estimand under minimax regret. My framework instead chooses a general scalar report while design ambiguity remains unresolved.⁴ Vilfort (2026) develops inference procedures for settings where there is ambiguity over how to average causal effects from a given design—e.g., an event study design may generate many causal effects across different treatment cohorts. The present framework instead considers decision-making in settings where there is ambiguity over a collection of candidate designs—e.g., a researcher may consider a weighted average for an event study design and a weighted average for an IV design.

The remainder of this paper proceeds as follows. Section 2 illustrates the headline estimation framework in an example setting. Section 3 develops the general decision problem and imposes my default specifications. Section 4 characterizes the optimal headline under general ambiguity classes. Section 5 gives asymptotic results. Section 6 presents empirical applications. Section 7 concludes. The supplemental appendix contains proofs.

2 Example Setting

In this section, I illustrate the framework at a high level in a two-design setting, postponing the general formulation and additional sensitivity analysis to later sections. Chen, Chen and Yang (2026) study a judicial reform in China that mandated courts to broadcast legal proceedings on an online platform, and examine how changes in broadcasting intensity affect the female-male plaintiff win-rate gap—estimated to be -0.0356 in their all-litigants sample.⁵ Let $\theta \in \mathbb{R}$ denote

⁴There is also more classical work on frequentist and Bayesian model averaging (Hansen, 2007; Hoeting et al., 1999). With a singleton design-prior class, the optimal headline reduces to a Bayes estimator that aligns with the logic of Bayesian model averaging estimators.

⁵This is the estimated female coefficient from the authors’ pre-reform all-litigants regression of win rate on a female indicator, control variables, and fixed effects (Chen, Chen and Yang, 2026, Table 1).

the target parameter, defined as “the effect of broadcasting intensity on the gender gap in plaintiff win rate.” The authors’ primary design uses a DiD specification, while their alternative design uses a Bartik IV to instrument for broadcasting intensity. For the all-litigants sample, the reported estimates and standard errors are

$$Y_{\text{DiD}}=0.0394, \quad \sigma_{\text{DiD}}=0.00255, \quad Y_{\text{IV}}=0.0726, \quad \sigma_{\text{IV}}=0.00599. \quad (1)$$

This means that a 10 percentage-point increase in broadcasting intensity corresponds to a 0.394 percentage-point narrowing of the gender gap under the DiD design, and a 0.726 percentage-point narrowing under the IV design. Both designs imply that greater broadcasting intensity narrows the gender gap, although the IV estimate is larger. Suppose that a researcher wishes to formulate a headline estimate of θ , but is unsure which of the two designs provides the more appropriate link to θ .

Let design 1 denote DiD and design 2 denote IV. Let (Y_1, Y_2) denote the design estimates, (σ_1, σ_2) their standard errors, and (μ_1, μ_2) their unknown estimands. One design is *target-valid*: for some unknown $k \in \{1, 2\}$, the target parameter is identified as $\theta = \mu_k$. To estimate θ , the researcher must account for two unknowns: the value of the estimands (μ_1, μ_2) and the identity of the target-valid design k .

To account for the unknown estimands (μ_1, μ_2) , the researcher conditions on the observed estimates (Y_1, Y_2) and adopts a diffuse-limit normal posterior, which yields $\mu_1|Y \sim N(Y_1, \sigma_1^2)$ and $\mu_2|Y \sim N(Y_2, \sigma_2^2)$. Let $a \in \mathbb{R}$ denote a candidate headline for θ . If k is target-valid, then the posterior mean squared error (MSE) is

$$\underbrace{\mathbb{E}[(\theta - a)^2|Y, k]}_{\text{posterior MSE}} = \underbrace{(Y_k - a)^2}_{\text{posterior squared bias}} + \underbrace{\sigma_k^2}_{\text{posterior variance}}, \quad \theta|Y, k \stackrel{d}{=} \underbrace{\mu_k|Y, k \sim N(Y_k, \sigma_k^2)}_{\text{posterior distribution}}.$$

Consider an oracle that knows k and therefore headlines the design- k estimate $a^* = Y_k$, leading to posterior MSE σ_k^2 . Under k , the posterior MSE of a candidate headline $a \in \mathbb{R}$ relative to the oracle benchmark is given by the design- k posterior risk

$$R_k(a|Y) = \frac{\mathbb{E}[(\theta - a)^2|Y, k]}{\sigma_k^2} = \frac{(Y_k - a)^2}{\sigma_k^2} + 1, \quad k \in \{1, 2\}.$$

This risk quantity measures the posterior mean-squared discrepancy between θ and a relative to how precisely design k can be estimated—I show in Section 3 that $R_k(a|Y)$ arises naturally from a decision problem with loss function $L(a, \mu, k) = (\mu_k - a)^2 / \sigma_k^2$.

To account for the unknown target-valid design k , the researcher considers different prior beliefs over k , which in this setting correspond to different prior probabilities $\lambda \in [0, 1]$ that DiD

is the target-valid design. Under a given design prior λ , the posterior risk is

$$R_\lambda(a|Y) = \lambda R_1(a|Y) + (1 - \lambda) R_2(a|Y).$$

Each design prior λ reflects *uncertainty* over research designs. For example, $\lambda = 2/3$ says that DiD is twice as likely to be target-valid as IV, while $\lambda = 1/3$ says that DiD is half as likely. The researcher may entertain many such priors, giving rise to a class of design priors $\Lambda \subseteq [0, 1]$. The class Λ reflects *ambiguity* over research designs.⁶ The optimal headline is the action $a = \hat{\theta}_\Lambda^*(Y)$ that minimizes worst-case posterior risk over the class Λ . In particular, the optimal headline and the optimized risk are

$$\hat{\theta}_\Lambda^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} R_\lambda(a|Y), \quad R_\Lambda^*(Y) = \min_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} R_\lambda(a|Y).$$

Intuitively, the optimal headline $\hat{\theta}_\Lambda^*(Y)$ guards against the worst-case design priors in Λ . The worst-case posterior risk $\max_{\lambda \in \Lambda} R_\lambda(a|Y)$ measures the consequences of compressing design-specific evidence into a scalar headline a , and the optimized risk $R_\Lambda^*(Y) = \max_{\lambda \in \Lambda} R_\lambda(\hat{\theta}_\Lambda^*(Y)|Y)$ is the minimum attainable risk. The excess risk $R_\Lambda^*(Y) - 1$ gives the proportional increase in risk relative to the oracle that knows the target-valid design, which provides a natural benchmark for interpreting the magnitudes of $R_\Lambda^*(Y)$.

The simplex class $\Lambda_\Delta = [0, 1]$ represents the benchmark of unrestricted ambiguity, wherein the researcher entertains every possible design prior. Under the simplex class, the solution is

$$\hat{\theta}_\Delta^*(Y) = \frac{\sigma_1^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}} Y_1 + \frac{\sigma_2^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}} Y_2, \quad R_\Delta^*(Y) = \left(\frac{|Y_1 - Y_2|}{\sigma_1 + \sigma_2} \right)^2 + 1.$$

The optimal headline $\hat{\theta}_\Delta^*(Y)$ is an inverse-standard-deviation-weighted average. It places higher weight on the more precise design, but less aggressively than the conventional inverse-variance-weighted average considered in models where both designs are independent unbiased estimates of a common parameter. The optimized risk $R_\Delta^*(Y)$ quantifies how well the DiD and IV designs can support—or be reconciled with—the optimal headline. For the simplex class, the square-root excess risk $c_\Delta^*(Y) = \sqrt{R_\Delta^*(Y) - 1}$ provides a natural measure of reconciliation cost. In particular, $c_\Delta^*(Y)$ is the smallest standard-error multiplier needed for two-sided intervals around the design estimates to share a common point:

$$c_\Delta^*(Y) = \min\{c \geq 0 : I_1(c) \cap I_2(c) \neq \emptyset\}, \quad I_k(c) = [Y_k - c\sigma_k, Y_k + c\sigma_k], \quad k \in \{1, 2\}.$$

That first common point is the simplex-optimal headline: $I_1(c_\Delta^*(Y)) \cap I_2(c_\Delta^*(Y)) = \{\hat{\theta}_\Delta^*(Y)\}$.

⁶This distinction between uncertainty and ambiguity—a single belief versus a class of beliefs—is consistent with standard formulations in decision theory (Berger, 1985; Gilboa and Schmeidler, 1989; Stoye, 2012).

From the [Chen, Chen and Yang \(2026\)](#) DiD and IV design estimates and standard errors in (1), the optimal headline and optimized risk under the simplex class $\Lambda_\Delta = [0, 1]$ are

$$\hat{\theta}_\Delta^*(Y) = 0.0493, \quad R_\Delta^*(Y) = 16.11.$$

In particular, a 10 percentage-point increase in broadcasting intensity yields a 0.493 percentage-point narrowing of the gender gap in plaintiff win rate. Headlining the DiD estimate $a = Y_1$ yields worst-case risk $R_\Delta(Y_1|Y) = 31.72$, so the optimal headline lowers risk by $(1 - (16.11/31.72)) \times 100 \approx 49$ percent. Even so, the optimized risk $R_\Delta^*(Y) = 16.11$ yields a $(16.11 - 1) \times 100 \approx 1500$ percent increase relative to the oracle that knows which design is target-valid. In particular, for the design intervals $I_k(c) = [Y_k - c\sigma_k, Y_k + c\sigma_k]$ to share a common point, one requires a standard-error multiplier of at least $c_\Delta^*(Y) = \sqrt{16.11 - 1} \approx 3.89$. Thus, while the optimal headline supports the conclusion that broadcasting narrows the gender gap, the high risk suggests that the DiD and IV designs do not tightly reconcile one exact magnitude. When I return to this application in Section 6, I will consider a favored-design ratio class that privileges DiD—the paper’s primary design—and examine how favored DiD would have to be to obtain meaningful risk reductions.

The above analysis considered a two-design setting with an unrestricted ambiguity class. The general decision problem allows for more designs and other ambiguity classes. The exposition below will separately introduce the ingredients used in the above analysis, including the latent-design setup, the beliefs over estimands and designs, the loss function preferences, and the proposed optimality criterion.

3 Decision Problem

Consider a researcher who wishes to estimate a scalar target parameter $\theta \in \mathbb{R}$. The researcher observes estimates $Y_k \in \mathbb{R}$ from research designs $k \in [K] = \{1, \dots, K\}$, where $K \geq 2$. Letting $Y = (Y_1, \dots, Y_K)'$ denote the vector of design estimates, I assume that Y is normally distributed:

$$Y \sim N(\mu, \Sigma),$$

where $\mu = (\mu_1, \dots, \mu_K)' \in \mathcal{M} \subseteq \mathbb{R}^K$ is a vector of unknown design estimands, $\mu_k = \mathbb{E}_\mu[Y_k]$, and Σ is a known positive definite covariance matrix. This assumption is motivated by large-sample approximations formalized in Section 5. Let $\sigma_k^2 = \Sigma_{kk} = \text{Var}_\mu(Y_k) > 0$.

I represent the set of research designs with the set of standard unit vectors $\mathcal{V} = \{v_1, \dots, v_K\}$. In particular, a given design $v_k \in \mathcal{V}$ yields $(v_k' Y, v_k' \mu, v_k' \Sigma v_k) = (Y_k, \mu_k, \sigma_k^2)$. For some latent design $v \in \mathcal{V}$, the target parameter θ is identified by the corresponding design estimand: $\theta = v' \mu$. I refer to v as the *target-valid* design. Thus, the underlying state variable is the unknown estimand-design pair $(\mu, v) \in \mathcal{M} \times \mathcal{V}$, while the target is a state-dependent functional $\theta(\mu, v) = v' \mu$. For example, $v = v_k$ means that design k is the target-valid design: $\theta = \theta(\mu, v_k) = \mu_k$. No restriction

is imposed on design estimands μ_j where $v_j \neq v$. Those estimands may differ from θ by arbitrary amounts, although they may coincide with θ numerically at some realizations of $\mu \in \mathcal{M}$. Thus, the value of v concerns the identifying link between a design and the target parameter—not a numerical requirement that exactly one coordinate of μ equal θ .

The researcher has a loss function $L(a, \mu, v) \geq 0$ that quantifies the consequences of taking action $a \in \mathcal{A} \subseteq \mathbb{R}$ when the estimand is $\mu \in \mathcal{M}$ and the target-valid design is $v \in \mathcal{V}$. For example, the squared error loss function $L(a, \mu, v) = (v'\mu - a)^2$ measures how far the action a is from the parameter $v'\mu = \theta$ when the estimand-design pair is (μ, v) . In other words, it measures the error relative to whichever design is target-valid in the state. I consider a general class of such loss functions in Section 3.3.

3.1 Researcher Beliefs

Let $\Delta(\mathcal{X})$ represent the set of probability distributions over a given set $\mathcal{X}$. Let $\gamma \in \Delta(\mathcal{M} \times \mathcal{V})$ denote prior beliefs about the state (μ, v) . This joint prior can be decomposed as

$$\gamma(d\mu, v_k) = \pi_k(d\mu)\lambda_k, \quad \pi_k(\cdot) = \pi(\cdot | v = v_k), \quad \lambda_k = \mathbb{P}_\lambda\{v = v_k\},$$

where $\pi_k \in \Delta(\mathcal{M})$ is the conditional prior on $\mu \in \mathcal{M}$ given $v = v_k$ and λ_k is the marginal prior probability that $v = v_k$. I call the vector $\lambda = (\lambda_1, \dots, \lambda_K)' \in \Delta(\mathcal{V})$ a *design prior*; its components are the design probabilities. Here, $\Delta(\mathcal{V})$ denotes the standard simplex in $\mathbb{R}^K$:

$$\Delta(\mathcal{V}) = \left\{ \lambda \in \mathbb{R}^K : \lambda \geq 0, \sum_{k=1}^K \lambda_k = 1 \right\}.$$

The researcher has ambiguity over research designs, represented by a nonempty, compact, and convex class $\Lambda \subseteq \Delta(\mathcal{V})$ of design priors $\lambda \in \Lambda$. Below are three example classes.

Example (Simplex class). The unrestricted-ambiguity benchmark is the simplex class:

$$\Lambda_\Delta = \Delta(\mathcal{V}) = \left\{ \lambda \in \mathbb{R}^K : \lambda \geq 0, \sum_{k=1}^K \lambda_k = 1 \right\}. \quad (2)$$

Under the simplex class, the researcher entertains every belief about v , including dogmatic priors (the simplex vertices $\mathcal{V} = \{v_1, \dots, v_K\}$) that put probability one on a single design. $\triangle$

Example (Baseline-prior perturbation class). Consider a baseline design prior $\lambda^0 \in \Delta(\mathcal{V})$. The perturbation class is the set of ϵ -perturbations towards other priors $q \in \Delta(\mathcal{V})$ in the simplex:

$$\Lambda_\epsilon(\lambda^0) = (1 - \epsilon)\lambda^0 + \epsilon\Delta(\mathcal{V}) = \left\{ (1 - \epsilon)\lambda^0 + \epsilon q : q \in \Delta(\mathcal{V}) \right\}, \quad \epsilon \in [0, 1]. \quad (3)$$

The baseline λ^0 represents the researcher's best assessment about the uncertainty in v , while

the perturbation level ϵ measures the ambiguity in that assessment. The perturbation class reduces to $\Lambda_0(\lambda^0) = \{\lambda^0\}$ at $\epsilon=0$ and expands to $\Lambda_1(\lambda^0) = \Delta(\mathcal{V})$ at $\epsilon=1$. $\triangle$

Example (Favored-design ratio class). Suppose a given design is favored in the sense that it is believed to be at least $r \geq 1$ times as plausible as every other design. Without loss of generality, suppose that $k=1$ is the favored design. The ratio class is defined as

$$\Lambda_r = \{\lambda \in \Delta(\mathcal{V}) : \lambda_1 \geq r \lambda_k, \forall k \neq 1\}, \quad r \geq 1. \quad (4)$$

The ratio parameter r places a lower bound on the favored design's relative probability of being target-valid. If $K=2$, the ratio class can be expressed as $\Lambda_r = \{\lambda : \lambda_1 \geq r/(r+1)\}$, which says that the researcher entertains any design prior satisfying $\mathbb{P}_\lambda\{v=v_1\} \geq r/(r+1)$. $\triangle$

Combining the conditional μ -priors with a class of design priors yields a class of joint priors:

$$\Gamma = \{\gamma : \gamma(d\mu, v_k) = \pi_k(d\mu) \lambda_k, \forall k \in [K], \lambda \in \Lambda\}.$$

Let $\phi_\Sigma(\cdot|\mu)$ denote the density function of $N(\mu, \Sigma)$. The marginal likelihood of Y conditional on $v=v_k$ is

$$m_{\pi,k}(Y) = \int_{\mathcal{M}} \phi_\Sigma(Y|\mu) \pi_k(d\mu).$$

I assume that $m_{\pi,k}(Y) > 0$ for all $k \in [K]$. Bayes' rule gives the posterior beliefs

$$\gamma(d\mu, v_k|Y) = \pi_k(d\mu|Y) \lambda_{\pi,k}(Y), \quad \pi_k(d\mu|Y) = \frac{\phi_\Sigma(Y|\mu) \pi_k(d\mu)}{m_{\pi,k}(Y)}, \quad \lambda_{\pi,k}(Y) = \frac{m_{\pi,k}(Y) \lambda_k}{\sum_{j=1}^K m_{\pi,j}(Y) \lambda_j}.$$

That is, the prior beliefs (γ, π_k) are updated into posterior beliefs $(\gamma(\cdot|Y), \pi_k(\cdot|Y))$. Likewise, $\lambda_{\pi,k}(Y)$ is the posterior probability that $v=v_k$. Letting $\lambda_\pi(Y) = (\lambda_{\pi,1}(Y), \dots, \lambda_{\pi,K}(Y))'$ denote the *design posterior* associated with design prior λ , the *design posterior class* is

$$\Lambda_\pi(Y) = \{\lambda_\pi(Y) : \lambda \in \Lambda\}.$$

For $\Lambda_\Delta = \Delta(\mathcal{V})$, the posterior simplex class preserves unrestricted ambiguity: $\Lambda_{\pi,\Delta}(Y) = \Lambda_\Delta$. For $\Lambda_\epsilon(\lambda^0) = (1-\epsilon)\lambda^0 + \epsilon\Delta(\mathcal{V}) = \{\lambda : \lambda \geq (1-\epsilon)\lambda^0\}$, the posterior perturbation class is

$$\Lambda_{\pi,\epsilon}(Y; \lambda^0) = \left\{ \bar{\lambda} \in \Delta(\mathcal{V}) : \frac{\bar{\lambda}_k / m_{\pi,k}(Y)}{\sum_{j=1}^K \bar{\lambda}_j / m_{\pi,j}(Y)} \geq (1-\epsilon) \lambda_k^0, \forall k \right\}.$$

For $\Lambda_r = \{\lambda : \lambda_1 \geq r\lambda_k, \forall k \neq 1\}$, the posterior ratio class is

$$\Lambda_{\pi,r}(Y) = \left\{ \bar{\lambda} \in \Delta(\mathcal{V}) : \frac{\bar{\lambda}_1/m_{\pi,1}(Y)}{\sum_{j=1}^K \bar{\lambda}_j/m_{\pi,j}(Y)} \geq r \frac{\bar{\lambda}_k/m_{\pi,k}(Y)}{\sum_{j=1}^K \bar{\lambda}_j/m_{\pi,j}(Y)}, \forall k \neq 1 \right\}.$$

If the conditional μ -priors π_k are common across designs k , then marginal likelihoods $m_{\pi,k}(Y)$ are common and any design posterior class coincides with its design prior class: $\Lambda_{\pi}(Y) = \Lambda$.

3.2 Optimality Criterion

To estimate the target parameter θ while accounting for ambiguity over designs, I propose a conditional Gamma-minimax approach (DasGupta and Studden, 1989; Betrò and Ruggeri, 1992; Giacomini, Kitagawa and Read, 2021). For each $k \in [K]$, define the design-specific posterior risk of action $a \in \mathcal{A} \subseteq \mathbb{R}$ as the posterior expected loss conditional on $v = v_k$:

$$R_{\pi,k}(a|Y) = \int_{\mathcal{M}} L(a, \mu, v_k) \pi_k(d\mu|Y), \quad v = v_k.$$

Under design prior $\lambda \in \Lambda$, the posterior risk is

$$R_{\pi,\lambda}(a|Y) = \sum_{k=1}^K \lambda_{\pi,k}(Y) R_{\pi,k}(a|Y).$$

For candidate headline $\hat{\theta}(Y)$, evaluate the design-specific and posterior risks at action $a = \hat{\theta}(Y)$. The conditional Gamma-minimax action $\hat{\theta}_{\pi,\Lambda}^*(Y)$ minimizes the worst-case posterior risk over the class of joint priors Γ , and thus equivalently over the class of design priors Λ :

$$\hat{\theta}_{\pi,\Lambda}^*(Y) \in \operatorname{argmin}_{a \in \mathcal{A}} \max_{\lambda \in \Lambda} R_{\pi,\lambda}(a|Y) = \operatorname{argmin}_{a \in \mathcal{A}} \max_{\bar{\lambda} \in \Lambda_{\pi}(Y)} \sum_{k=1}^K \bar{\lambda}_k R_{\pi,k}(a|Y).$$

The minimized worst-case posterior risk is

$$R_{\pi,\Lambda}^*(Y) = \max_{\lambda \in \Lambda} R_{\pi,\lambda}(\hat{\theta}_{\pi,\Lambda}^*(Y)|Y) = \max_{\bar{\lambda} \in \Lambda_{\pi}(Y)} \sum_{k=1}^K \bar{\lambda}_k R_{\pi,k}(\hat{\theta}_{\pi,\Lambda}^*(Y)|Y).$$

Thus, the conditional Gamma-minimax criterion defines the optimal headline as an optimal action $a = \hat{\theta}_{\pi,\Lambda}^*(Y)$ under design ambiguity, where the maximization over $\lambda \in \Lambda$ (or $\bar{\lambda} \in \Lambda_{\pi}(Y)$) accounts for the design ambiguity. The optimized risk $R_{\pi,\Lambda}^*(Y)$ measures the consequences of the optimal action under worst-case design beliefs. I refer to $\hat{\theta}_{\pi,\Lambda}^*(Y)$ as the *ambiguity-optimal headline estimate*, or *optimal headline* for short.⁷

⁷For a singleton $\Lambda = \{\lambda\}$, the optimal headline is a Bayes estimator under the joint prior γ induced by (π, λ) .

3.3 Default Specifications

The above decision problem allows for general specifications of the researcher's loss function and beliefs. In this section, I propose default specifications which yield particularly tractable headline estimation procedures. I characterize the resulting procedures in Section 4.

3.3.1 Loss Function

Let $\mathcal{A} = \mathbb{R}$. I consider weighted squared error loss functions

$$L(a, \mu, v) = w(v)(v'\mu - a)^2, \quad w(v) > 0.$$

The squared error component $(v'\mu - a)^2$ measures the discrepancy between the action $a \in \mathbb{R}$ and the target parameter $\theta = v'\mu$ when the estimand-design pair is (μ, v) . The loss weights $w(v)$ allow the penalty from such discrepancies to depend on the latent design v . Letting $w_k = w(v_k)$, the design-specific posterior risk is a conditional MSE:

$$R_{\pi, k}(a|Y) = w_k \left\{ (\mathbb{E}_{\pi}[\mu_k | v = v_k, Y] - a)^2 + \text{Var}_{\pi}(\mu_k | v = v_k, Y) \right\}, \quad v = v_k.$$

For design prior $\lambda \in \Lambda$, the posterior risk is therefore an average conditional MSE:

$$R_{\pi, \lambda}(a|Y) = \sum_{k=1}^K \lambda_{\pi, k}(Y) w_k \left\{ (\mathbb{E}_{\pi}[\mu_k | v = v_k, Y] - a)^2 + \text{Var}_{\pi}(\mu_k | v = v_k, Y) \right\}.$$

The loss weights w_k play a different role from the class Λ . The loss weights describe the cost of errors given the target-valid design, whereas the class describes the ambiguity over which design is target-valid. The former captures preferences while the latter captures beliefs. Either component can be varied while holding the other fixed.

3.3.2 Beliefs over Estimands

Let $\mathcal{M} = \mathbb{R}^K$ and consider common normal conditional μ -priors:

$$\pi_k = \pi_{\varsigma} = N(\beta_0, \varsigma^{-1} \Omega_0), \quad \forall k \in [K],$$

where Ω_0 is positive definite and $\varsigma > 0$. Because the conditional μ -priors are common, so are the marginal likelihoods $m_{\pi, k}(Y) = m_{\pi}(Y)$, and hence any design posterior class coincides with its design prior class: $\Lambda_{\pi}(Y) = \Lambda$. The conditional μ -posteriors are also common:

$$\pi_k(\cdot | Y) = \pi_{\varsigma}(\cdot | Y) = N\left(\left(\varsigma \Omega_0^{-1} + \Sigma^{-1}\right)^{-1} \left(\varsigma \Omega_0^{-1} \beta_0 + \Sigma^{-1} Y\right), \left(\varsigma \Omega_0^{-1} + \Sigma^{-1}\right)^{-1}\right), \quad \forall k \in [K].$$

Letting $\varsigma \rightarrow 0$ yields a convenient diffuse-limit normal posterior

$$\pi_\varsigma(\cdot|Y) \xrightarrow{d} \pi_0(\cdot|Y) = N(Y, \Sigma).$$

This approximates settings where the researcher has μ -priors that are low precision relative to the information in Y . Here this takes the form of a diffuse normal prior sequence $\{\pi_\varsigma\}_{\varsigma \rightarrow 0}$. I use $\pi_0(\cdot|Y) = N(Y, \Sigma)$ as the default μ -posterior. The design-specific posterior risk becomes

$$R_{\pi_0, k}(a|Y) = R_k(a|Y) = w_k \left\{ (Y_k - a)^2 + \sigma_k^2 \right\}, \quad v = v_k.$$

The posterior risk under $\lambda \in \Lambda$ is therefore

$$R_{\pi_0, \lambda}(a|Y) = R_\lambda(a|Y) = \sum_{k=1}^K \lambda_k w_k \left\{ (Y_k - a)^2 + \sigma_k^2 \right\}, \quad \lambda_{\pi_0, k}(Y) = \lambda_k.$$

Because design posteriors $\lambda_{\pi_0}(Y)$ coincide with design priors λ under the default specifications, I will simply refer to the latter when speaking about design beliefs.

3.3.3 Loss Function Weights

I consider precision weights

$$w(v) = (v' \Sigma v)^{-1} > 0, \quad w_k = w(v_k) = 1/\sigma_k^2.$$

The resulting standardized squared error loss function is

$$L(a, \mu, v) = \frac{(v' \mu - a)^2}{v' \Sigma v}.$$

This places the design-specific errors on a common standard-error scale and prevents a highly imprecise design estimate from dominating the decision problem. The precision weights also yield the following design-neutrality property: an oracle that observes v before acting incurs the same posterior risk regardless of the value of v . Intuitively, the oracle is not worse off learning that design k is target-valid versus learning that design j is target-valid.

Proposition 1 (Design-neutrality). *The oracle procedure that reports Y_k if and only if design k is target-valid yields oracle posterior risk*

$$\sum_{k=1}^K \mathbf{1}\{v = v_k\} R_k(Y_k|Y) = \sum_{k=1}^K \mathbf{1}\{v = v_k\} w_k \sigma_k^2, \quad v \in \mathcal{V}.$$

The oracle risk is the same across realizations of v if and only if $w_k = c/\sigma_k^2$ for a constant $c > 0$. Normalizing to $c = 1$ yields the precision weights $w_k = 1/\sigma_k^2$.

Proof. See Appendix A.1. □

Under precision weights $w_k = 1/\sigma_k^2$, the design-specific posterior risk becomes

$$R_k(a|Y) = \frac{(Y_k - a)^2}{\sigma_k^2} + 1, \quad v = v_k.$$

In words, the design-specific posterior risk of an action $a \in \mathbb{R}$ is equal to one plus the squared standardized discrepancy between a and Y_k . Intuitively, the design- k penalty of headlining a grows with a 's distance from the design estimate relative to that design's standard error. The posterior risk under $\lambda \in \Lambda$ is the corresponding average squared standardized discrepancy (+1) across designs:

$$R_\lambda(a|Y) = \sum_{k=1}^K \lambda_k \frac{(Y_k - a)^2}{\sigma_k^2} + 1.$$

Thus, $R_\lambda(a|Y) - 1$ gives the mean-squared standardized discrepancy of a candidate headline a from the design estimates $(Y_1, \dots, Y_K)$, where the mean is taken under $\lambda \in \Lambda$.

Remark 1 (Relationship to unweighted squared error loss). For unweighted (i.e., $w(v) = 1$) squared error loss $L(a, \mu, v) = (v'\mu - a)^2$, the design-specific posterior risk at v_k is $(Y_k - a)^2 + \sigma_k^2$. Comparing this posterior risk to that of an oracle with knowledge of v yields posterior regret

$$\frac{(Y_k - a)^2 + \sigma_k^2}{\min_{a \in \mathbb{R}} \{(Y_k - a)^2 + \sigma_k^2\}} = \frac{(Y_k - a)^2}{\sigma_k^2} + 1, \quad v = v_k.$$

This posterior regret measure is analogous to the one considered in [Andrews and Shapiro \(2026\)](#), but defined here as a ratio of posterior risks instead of a level-difference. The ratio comparison is analogous to the ex-ante adaptation regret formulation considered in [Armstrong, Kline and Sun \(2025\)](#). Thus, the posterior risk under precision-weighted squared error loss coincides with a measure of posterior (ex-post) adaptation regret under unweighted squared error loss. $\triangle$

4 Proposed Procedures

Under the default specifications in Section 3.3, the optimal headline estimate is

$$\hat{\theta}_\Lambda^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(Y_k - a)^2}{\sigma_k^2} + 1.$$

For any candidate headline $a \in \mathbb{R}$, its worst-case posterior risk is

$$R_\Lambda(a|Y) = \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(Y_k - a)^2}{\sigma_k^2} + 1.$$

The +1 term is the irreducible oracle risk of the infeasible procedure that reports Y_k if and only if design k is target-valid. Thus, $(R_\Lambda(a|Y) - 1) \times 100$ is the percentage increase over oracle risk. The optimal headline and optimized risk are

$$\hat{\theta}_\Lambda^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} R_\Lambda(a|Y), \quad R_\Lambda^*(Y) = \min_{a \in \mathbb{R}} R_\Lambda(a|Y) = R_\Lambda(\hat{\theta}_\Lambda^*(Y)|Y).$$

If a paper has a baseline headline $a_B \in \mathbb{R}$, then it is useful to define

$$R_\Lambda^B = R_\Lambda(a_B|Y), \quad R_\Lambda^* = R_\Lambda(\hat{\theta}_\Lambda^*(Y)|Y).$$

Comparing the baseline risk R_Λ^B with the optimized risk R_Λ^* shows whether accounting for the design ambiguity represented by Λ materially improves the headline report. The reduction in worst-case risk is $R_\Lambda^B - R_\Lambda^*$, and the percentage reduction is

$$\left(1 - \frac{R_\Lambda^*}{R_\Lambda^B}\right) \times 100.$$

This provides a convenient bounded summary.

In Section 4.1, I provide a general characterization of the proposed procedure for nonempty, compact, and convex classes Λ . In Section 4.2, I specialize to the simplex class Λ_Δ defined in (2), which represents the benchmark of unrestricted ambiguity. In Section 4.3, I consider more structured forms of ambiguity, specializing first to the perturbation class $\Lambda_\epsilon(\lambda^0)$ defined in (3) and then to the ratio class Λ_r defined in (4).

4.1 General Characterization

Let $D_k(a|Y) = |Y_k - a|/\sigma_k$ denote the standardized discrepancy under design k , which measures the distance between headline a and estimate Y_k in units of the design's standard error. For any nonempty, compact, and convex class $\Lambda \subseteq \Delta(\mathcal{V})$, the optimal headline and optimized risk are

$$\hat{\theta}_\Lambda^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} R_\Lambda(a|Y), \quad R_\Lambda^*(Y) = R_\Lambda(\hat{\theta}_\Lambda^*(Y)|Y), \quad R_\Lambda(a|Y) = \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k D_k(a|Y)^2 + 1.$$

For any design prior $\lambda \in \Delta(\mathcal{V})$, define its Bayes headline and excess posterior risk as

$$a(\lambda|Y) = \frac{\sum_{k=1}^K \lambda_k Y_k / \sigma_k^2}{\sum_{k=1}^K \lambda_k / \sigma_k^2}, \quad \rho(\lambda|Y) = \sum_{k=1}^K \lambda_k D_k(a(\lambda|Y)|Y)^2 = \frac{\sum_{j < k} (Y_j - Y_k)^2 (\lambda_j / \sigma_j^2) (\lambda_k / \sigma_k^2)}{\sum_{\ell=1}^K \lambda_\ell / \sigma_\ell^2}.$$

The following result characterizes the headline estimation procedure.

Proposition 2 (Least favorable design priors). *The optimal headline $\hat{\theta}_\Lambda^*(Y)$ exists uniquely.*

Moreover, there exists $\lambda^* \in \Lambda$ such that

$$\lambda^* \in \arg\max_{\lambda \in \Lambda} \rho(\lambda|Y), \quad \hat{\theta}_\Lambda^*(Y) = a(\lambda^*|Y), \quad R_\Lambda^*(Y) = \rho(\lambda^*|Y) + 1.$$

That is, the optimal headline can be represented as an inverse-variance-weighted average under a least favorable design prior $\lambda^* \in \Lambda$. Moreover, λ^* is a worst-case design prior at $a = \hat{\theta}_\Lambda^*(Y)$:

$$R_\Lambda(\hat{\theta}_\Lambda^*(Y)|Y) = R_{\lambda^*}(\hat{\theta}_\Lambda^*(Y)|Y).$$

For $K=2$, define $p_\Delta = \sigma_1/(\sigma_1 + \sigma_2)$, which satisfies (uniquely when $Y_1 \neq Y_2$)

$$p_\Delta \in \arg\max_{p \in [0,1]} \rho(\lambda(p)|Y), \quad \rho(\lambda(p)|Y) = \frac{p(1-p)(Y_1 - Y_2)^2}{p\sigma_2^2 + (1-p)\sigma_1^2}, \quad \lambda(p) = (p, 1-p)'$$

Define the interval of Λ probabilities on design 1 and the projection of p_Δ onto this interval as

$$\mathcal{I}_\Lambda = \{p \in [0,1] : \lambda(p) \in \Lambda\} = [\underline{p}_\Lambda, \bar{p}_\Lambda], \quad p_\Lambda^* = \arg\min_{p \in \mathcal{I}_\Lambda} |p - p_\Delta| = \min\left\{\max\{p_\Delta, \underline{p}_\Lambda\}, \bar{p}_\Lambda\right\}.$$

In this case, the optimal headline and optimized risk are

$$\hat{\theta}_\Lambda^*(Y) = \frac{p_\Lambda^* Y_1 / \sigma_1^2 + (1 - p_\Lambda^*) Y_2 / \sigma_2^2}{p_\Lambda^* / \sigma_1^2 + (1 - p_\Lambda^*) / \sigma_2^2}, \quad R_\Lambda^*(Y) = \frac{p_\Lambda^* (1 - p_\Lambda^*) (Y_1 - Y_2)^2}{p_\Lambda^* \sigma_2^2 + (1 - p_\Lambda^*) \sigma_1^2} + 1.$$

If $Y_1 \neq Y_2$, then $\lambda(p_\Lambda^*)$ is the unique least favorable design prior. If $Y_1 = Y_2$, then every $\lambda \in \Lambda$ is least favorable and $\hat{\theta}_\Lambda^*(Y) = Y_1 = Y_2$ with $R_\Lambda^*(Y) = 1$.

Proof. See Appendix A.2. □

Remark 2 (Computation). The function $\lambda \mapsto \rho(\lambda|Y)$ is continuous and concave, where concavity follows because $\rho(\lambda|Y) = \min_{a \in \mathbb{R}} \sum_{k=1}^K \lambda_k D_k(a|Y)^2$ is the pointwise minimum of functions that are affine in λ . Consequently, computing a least favorable prior amounts to maximizing a concave objective over the compact convex set Λ , which yields a standard finite-dimensional convex-optimization problem when Λ is represented by convex constraints. Once any maximizing λ^* is obtained, the optimal headline and optimized risk follow directly from $a(\lambda^*|Y)$ and $\rho(\lambda^*|Y) + 1$. The least favorable prior need not be unique, but every maximizing prior yields the same unique optimal headline and the same optimized risk. △

Remark 3 (Equivariance under affine transformations). The optimal headline is equivariant under affine transformations in the sense that using $(\tilde{Y}_k, \tilde{\sigma}_k) = (\xi_1 + \xi_2 Y_k, |\xi_2| \sigma_k)$ with $\xi_1 \in \mathbb{R}$ and $\xi_2 \neq 0$ yields headline $\xi_1 + \xi_2 \hat{\theta}_\Lambda^*(Y)$. The optimized risk is unchanged. Thus, a reader interested in the affine transformation $\xi_1 + \xi_2 \theta$ can obtain the corresponding optimal headline by applying the same transformation to the given headline $\hat{\theta}_\Lambda^*(Y)$. This shortcut does not extend to a nonlinear

transformation $h(\theta)$, since $h(\hat{\theta}_\Lambda^*(Y))$ generally differs from the optimal headline obtained by formulating the decision problem for $h(\theta)$ on the transformed scale. $\triangle$

Remark 4 (Off-diagonal covariance). The optimal headline and optimized risk depend on Σ only through the design variances $\sigma_1^2, \dots, \sigma_K^2$. This is a general implication of squared error loss under μ -posterior $\pi_0(\cdot|Y) = N(Y, \Sigma)$. Off-diagonal covariance terms affect the sampling distribution of $(\hat{\theta}_\Lambda^*(Y), R_\Lambda^*(Y))$, but they do not enter the decision problem under the default specifications. $\triangle$

4.2 Unrestricted-Ambiguity Benchmark

Under the simplex class $\Lambda_\Delta = \Delta(\mathcal{V})$, for any candidate headline a , a worst-case design prior can be chosen to place probability one on a design that attains the largest standardized discrepancy $\max_{k \in [K]} D_k(a|Y)$. Therefore, the optimal headline and optimized risk reduce to

$$\hat{\theta}_\Delta^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} \max_{k \in [K]} D_k(a|Y), \quad R_\Delta^*(Y) = \left(\min_{a \in \mathbb{R}} \max_{k \in [K]} D_k(a|Y) \right)^2 + 1.$$

Define the *reconciliation cost* $c_\Delta^*(Y) = \sqrt{R_\Delta^*(Y) - 1}$.

Proposition 3 (Unrestricted ambiguity). *The optimal headline and optimized risk are*

$$\hat{\theta}_\Delta^*(Y) = \max_{k \in [K]} \{Y_k - \sigma_k c_\Delta^*(Y)\} = \min_{k \in [K]} \{Y_k + \sigma_k c_\Delta^*(Y)\}, \quad R_\Delta^*(Y) = \left(\max_{j < k} \frac{|Y_j - Y_k|}{\sigma_j + \sigma_k} \right)^2 + 1.$$

If (j^*, k^*) attains the pairwise maximum in $R_\Delta^*(Y)$, then

$$\hat{\theta}_\Delta^*(Y) = \frac{\sigma_{k^*} Y_{j^*} + \sigma_{j^*} Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}}, \quad c_\Delta^*(Y) = D_{j^*}(\hat{\theta}_\Delta^*(Y)|Y) = D_{k^*}(\hat{\theta}_\Delta^*(Y)|Y) = \frac{|Y_{j^*} - Y_{k^*}|}{\sigma_{j^*} + \sigma_{k^*}}.$$

The reconciliation cost $c_\Delta^*(Y)$ is the smallest standard-error multiplier $c \geq 0$ such that the design intervals $I_k(c) = [Y_k - c\sigma_k, Y_k + c\sigma_k]$ have nonempty intersection. In particular,

$$\bigcap_{k=1}^K I_k(c_\Delta^*(Y)) = \{\hat{\theta}_\Delta^*(Y)\}, \quad c_\Delta^*(Y) = \min \left\{ c \geq 0 : \bigcap_{k=1}^K I_k(c) \neq \emptyset \right\}.$$

For $K=2$, the solution is given by the above formulas with $(j^*, k^*) = (1, 2)$. If $Y_1 \neq Y_2$, the unique least favorable prior is $\lambda^* = (p_\Delta, 1 - p_\Delta)'$, where $p_\Delta = \sigma_1 / (\sigma_1 + \sigma_2)$ as in Proposition 2.

Proof. See Appendix A.3. $\square$

For $K=2$, the solution is

$$\hat{\theta}_\Delta^*(Y) = \frac{\sigma_1^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}} Y_1 + \frac{\sigma_2^{-1}}{\sigma_1^{-1} + \sigma_2^{-1}} Y_2, \quad R_\Delta^*(Y) = \left(\frac{|Y_1 - Y_2|}{\sigma_1 + \sigma_2} \right)^2 + 1.$$

In this case, the optimal headline $\hat{\theta}_\Delta^*(Y)$ is an inverse-standard-deviation-weighted average of the two design estimates. It places higher weight on the more precise design, but less aggressively than the conventional inverse-variance-weighted average used when estimates are independent and unbiased for a common parameter. The latter procedure chooses weights to minimize the variance of a weighted estimator, while the above headline estimation procedure equalizes the two standardized discrepancies to account for ambiguity over research designs. The excess risk $R_\Delta^*(Y) - 1$ is the squared standardized separation between the two estimates.

For $K > 2$, an analogous solution is determined by the active pair (Y_{j^*}, Y_{k^*}) attaining the largest standardized separation. The remaining estimates do not affect the headline unless they change which pair is active. A risk of one means exact numerical agreement across designs. A risk no greater than two means that one-standard-error intervals around each design estimate share a common point. Under the simplex class, the reconciliation cost $c_\Delta^*(Y)$ is the smallest common standard-error multiplier such that the design intervals share a common point, and $\hat{\theta}_\Delta^*(Y)$ is the unique intersection point at that multiplier.

Remark 5 (Sensitivity to outliers). The unrestricted-ambiguity benchmark has the feature of being sensitive to outlier design estimates. Because any one design may be target-valid with arbitrarily high probability under the simplex class, the simplex-optimal headline has to guard against dogmatic design priors. This means that a cluster of proximate design estimates does not necessarily outweigh a single discrepant estimate. $\triangle$

Remark 6 (Calibrating the reconciliation cost). Under the default posterior,

$$\theta|Y, v = v_k \sim N(Y_k, \sigma_k^2), \quad \forall k \in [K].$$

For $\gamma \in (0, 1)$, define the design-specific posterior credible interval as

$$C_k(\gamma|Y) = [Y_k - z_{1-\gamma/2}\sigma_k, Y_k + z_{1-\gamma/2}\sigma_k], \quad z_{1-\gamma/2} = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right).$$

Conditional on $v = v_k$, the credible interval $C_k(\gamma|Y)$ contains θ with posterior probability $1 - \gamma$:

$$\mathbb{P}\{\theta \in C_k(\gamma|Y) | Y, v = v_k\} = 1 - \gamma.$$

This equality continues to hold for data-dependent $\gamma(Y)$, since the posterior conditions on Y . Under the simplex class, the interval characterization in Proposition 3 implies

$$\gamma_\Delta^*(Y) = 2(1 - \Phi(c_\Delta^*(Y))) = \sup\left\{\gamma \in (0, 1) : \bigcap_{k=1}^K C_k(\gamma|Y) \neq \emptyset\right\}.$$

Thus, $\gamma_\Delta^*(Y)$ is the largest common posterior noncoverage level at which the design-specific

credible intervals still share a point. Equivalently, $1 - \gamma_\Delta^*(Y)$ is the minimum common credibility level required for mutual overlap. At this boundary, the first common point is $\hat{\theta}_\Delta^*(Y)$.

The 95% credibility level benchmark is given by $c_\Delta^*(Y) = z_{0.975} \approx 1.96$. For ease of reporting, I use the rounded $c = 2$ as a cost benchmark, which corresponds to risk $R = 2^2 + 1 = 5$ —a 400% increase over the oracle risk. Under the simplex, a cost above two therefore means that even the approximately 95% design-specific credible intervals do not share a common point. $\triangle$

Remark 7 (Reconciliation cost for other classes). More generally for the other classes Λ under consideration, define the reconciliation cost

$$c_\Lambda^*(Y) = \sqrt{R_\Lambda^*(Y) - 1}.$$

For $\Lambda \neq \Lambda_\Delta$, the reconciliation cost does not generally have an interval-overlap interpretation. I nevertheless use $c = 2$, or equivalently $R = 5$, as a *simplex-calibrated* benchmark for the other classes of interest.⁸ $\triangle$

4.3 Structured Ambiguity Classes

The simplex class accommodates cases where the researcher does not want to rule out any possible design prior. In other cases, the researcher’s ambiguity may have additional structure, as with the perturbation and ratio classes defined in (3) and (4). I now characterize these structured ambiguity classes. They can serve as primary specifications when their restrictions are substantively justified or as sensitivity analyses relative to the unrestricted-ambiguity benchmark.

4.3.1 Baseline-Prior Perturbation Class

Consider the perturbation class $\Lambda_\epsilon(\lambda^0) = (1 - \epsilon)\lambda^0 + \epsilon\Delta(\mathcal{V})$ for a baseline design prior $\lambda^0 \in \Delta(\mathcal{V})$.

Proposition 4 (Baseline-prior ambiguity). *For perturbation level $\epsilon \in [0, 1]$,*

$$\hat{\theta}_\epsilon^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} R_\epsilon(a|Y), \quad R_\epsilon(a|Y) = \left((1 - \epsilon) \sum_{k=1}^K \lambda_k^0 D_k(a|Y)^2 + \epsilon \max_{k \in [K]} D_k(a|Y)^2 \right) + 1.$$

Let $R_\epsilon^*(Y) = R_\epsilon(\hat{\theta}_\epsilon^*(Y)|Y)$. By Proposition 2, the solution can be computed as

$$\lambda_\epsilon^* \in \operatorname{argmax} \left\{ \rho(\lambda|Y) : \lambda \in \Delta(\mathcal{V}), \lambda \geq (1 - \epsilon)\lambda^0 \right\}, \quad \hat{\theta}_\epsilon^*(Y) = a(\lambda_\epsilon^*|Y), \quad R_\epsilon^*(Y) = \rho(\lambda_\epsilon^*|Y) + 1.$$

The optimal headline is continuous in ϵ and follows a monotone path between its endpoints: if $\hat{\theta}_0^*(Y) < \hat{\theta}_1^*(Y)$, then $\epsilon \mapsto \hat{\theta}_\epsilon^*(Y)$ is nondecreasing; if $\hat{\theta}_0^*(Y) > \hat{\theta}_1^*(Y)$, then it is nonincreasing; and

⁸ $c_\Lambda^*(Y) \leq c_\Delta^*(Y)$ because $\Lambda \subseteq \Delta(\mathcal{V}) = \Lambda_\Delta$. Thus, $c_\Lambda^*(Y) > 2$ implies that $c_\Delta^*(Y)$ exceeds the two-standard-error multiplier benchmark. However, $c_\Lambda^*(Y) \leq 2$ need not imply that all 95% design-specific credible intervals overlap.

if the endpoints coincide, then the path is constant. At the endpoints,

$$\hat{\theta}_0^*(Y) = a(\lambda^0|Y) = \frac{\sum_{k=1}^K \lambda_k^0 Y_k / \sigma_k^2}{\sum_{k=1}^K \lambda_k^0 / \sigma_k^2}, \quad \hat{\theta}_1^*(Y) = \hat{\theta}_\Delta^*(Y).$$

For $K=2$, $p^0 = \mathbb{P}_{\lambda^0}\{v=v_1\}$, and $p_\Delta = \sigma_1/(\sigma_1 + \sigma_2)$, Proposition 2 gives

$$p_\epsilon^* = \operatorname{argmin}_{p \in \mathcal{I}_\epsilon} |p - p_\Delta| = \min\left\{\max\left\{p_\Delta, (1-\epsilon)p^0\right\}, (1-\epsilon)p^0 + \epsilon\right\}, \quad \mathcal{I}_\epsilon = [(1-\epsilon)p^0, (1-\epsilon)p^0 + \epsilon].$$

In this case, the optimal headline and optimized risk are

$$\hat{\theta}_\epsilon^*(Y) = \frac{p_\epsilon^* Y_1 / \sigma_1^2 + (1-p_\epsilon^*) Y_2 / \sigma_2^2}{p_\epsilon^* / \sigma_1^2 + (1-p_\epsilon^*) / \sigma_2^2}, \quad R_\epsilon^*(Y) = \frac{p_\epsilon^* (1-p_\epsilon^*) (Y_1 - Y_2)^2}{p_\epsilon^* \sigma_2^2 + (1-p_\epsilon^*) \sigma_1^2} + 1. \quad (5)$$

If $p_\Delta \in \mathcal{I}_\epsilon$, then $\hat{\theta}_\epsilon^*(Y) = \hat{\theta}_\Delta^*(Y)$. In particular, when $p^0 \in (0,1)$, the perturbation-optimal headline coincides with the simplex-optimal headline if $\epsilon \geq \max\{1 - (p_\Delta/p^0), 1 - (1-p_\Delta)/(1-p^0)\}$.

Proof. See Appendix A.4. □

The excess risk criterion $R_\epsilon(a|Y) - 1$ is a convex combination of (i) the baseline design prior average of the squared standardized discrepancies and (ii) the maximum discrepancy, where the perturbation level ϵ controls how much the headline estimation procedure protects against the baseline design prior being incorrect. At $\epsilon=0$, the researcher fully commits to the baseline prior, leading to Bayes headline $a(\lambda^0|Y)$. At $\epsilon=1$, the perturbation class coincides with the simplex class, leading to the unrestricted-ambiguity headline $\hat{\theta}_\Delta^*(Y)$. The perturbation-optimal headline moves monotonically between these two endpoint headlines—or remains constant when they coincide. In particular, when $\lambda^0 = \mathbf{1}/K$, the endpoint at $\epsilon=0$ is the inverse-variance-weighted average; when the design estimates are independent, this is the conventional precision-weighted average. For sensitivity analysis relative to the unrestricted-ambiguity benchmark, a researcher can trace the perturbation-optimal headline and optimized risk over a range of ϵ or report the smallest perturbation level at which a qualitative conclusion changes.

For $K=2$ and an equal-probability baseline $\lambda^0 = \mathbf{1}/2 = (1/2, 1/2)'$, solution (5) yields

$$\hat{\theta}_\epsilon^*(Y) = \frac{p_\epsilon^* Y_1 / \sigma_1^2 + (1-p_\epsilon^*) Y_2 / \sigma_2^2}{p_\epsilon^* / \sigma_1^2 + (1-p_\epsilon^*) / \sigma_2^2}, \quad p_\epsilon^* = \min\left\{\max\left\{p_\Delta, (1-\epsilon)\frac{1}{2}\right\}, (1-\epsilon)\frac{1}{2} + \epsilon\right\}, \quad \epsilon \in [0,1].$$

The corresponding endpoints are

$$\hat{\theta}_0^*(Y) = a(\lambda^0|Y) = \frac{Y_1/\sigma_1^2 + Y_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \quad \hat{\theta}_1^*(Y) = \hat{\theta}_\Delta^*(Y) = \frac{Y_1/\sigma_1 + Y_2/\sigma_2}{1/\sigma_1 + 1/\sigma_2}.$$

Thus, the perturbation-optimal headline moves from the inverse-variance-weighted average

toward the inverse-standard-deviation-weighted average. Every point on this path lies between those two endpoint headlines and can therefore be written as a convex combination of them, although the convex-combination weight is generally not equal to ϵ .

4.3.2 Favored-Design Ratio Class

Consider the ratio class $\Lambda_r = \{\lambda: \lambda_1 \geq r\lambda_k, \forall k \neq 1\}$. Denote a decreasing rearrangement of the squared standardized discrepancies $\{D_k(a|Y)^2: k \neq 1\}$ as $D_{-1,(1)}(a|Y)^2 \geq \dots \geq D_{-1,(K-1)}(a|Y)^2$.

Proposition 5 (Favored-design ambiguity). *For ratio parameter $r \geq 1$,*

$$\hat{\theta}_r^*(Y) = \operatorname{argmin}_{a \in \mathbb{R}} R_r(a|Y), \quad R_r(a|Y) = \max_{m \in \{0, \dots, K-1\}} \frac{rD_1(a|Y)^2 + \sum_{j=1}^m D_{-1,(j)}(a|Y)^2}{r+m} + 1.$$

For each m , the corresponding term in the maximum is generated by an extreme design prior that places probability $r/(r+m)$ on the favored design and probability $1/(r+m)$ on each of the m alternatives with the largest squared standardized discrepancies. If $m=0$, the inner sum is set to zero. Let $R_r^(Y) = R_r(\hat{\theta}_r^*(Y)|Y)$. By Proposition 2, the solution can be computed as*

$$\lambda_r^* \in \operatorname{argmax}\{\rho(\lambda|Y): \lambda \in \Delta(\mathcal{V}), \lambda_1 \geq r\lambda_k, \forall k \neq 1\}, \quad \hat{\theta}_r^*(Y) = a(\lambda_r^*|Y), \quad R_r^*(Y) = \rho(\lambda_r^*|Y) + 1.$$

The optimized risk $R_r^(Y)$ is nonincreasing in r and, as $r \rightarrow \infty$,*

$$R_r^*(Y) \rightarrow 1, \quad \hat{\theta}_r^*(Y) \rightarrow Y_1.$$

For $K=2$, $p_r = r/(r+1)$, and $p_\Delta = \sigma_1/(\sigma_1 + \sigma_2)$, Proposition 2 gives

$$p_r^* = \operatorname{argmin}_{p \in \mathcal{I}_r} |p - p_\Delta| = \max\{p_\Delta, p_r\}, \quad \mathcal{I}_r = [p_r, 1].$$

In this case, the optimal headline and optimized risk are

$$\hat{\theta}_r^*(Y) = \frac{p_r^* Y_1 / \sigma_1^2 + (1 - p_r^*) Y_2 / \sigma_2^2}{p_r^* / \sigma_1^2 + (1 - p_r^*) / \sigma_2^2}, \quad R_r^*(Y) = \frac{p_r^* (1 - p_r^*) (Y_1 - Y_2)^2}{p_r^* \sigma_2^2 + (1 - p_r^*) \sigma_1^2} + 1.$$

If $p_\Delta \in \mathcal{I}_r$, then $\hat{\theta}_r^(Y) = \hat{\theta}_\Delta^*(Y)$. Equivalently, the ratio-optimal and simplex-optimal headlines coincide if $r \leq \sigma_1/\sigma_2$.*

Proof. See Appendix A.5. □

The extreme-prior formula gives a concrete interpretation of the criterion $R_r(a|Y)$. Think of the favored design as receiving r units of prior probability and each selected alternative as receiving one unit. For a candidate headline and a fixed number m of alternatives, a worst-case extreme prior selects the m designs with the largest squared standardized discrepancies. The criterion

then chooses the value of m that gives the largest weighted average discrepancy. If the maximizing extreme prior is unique at the optimal headline, it is least favorable. When several extreme priors tie, the least-favorable prior may instead be a mixture, as described in Appendix A.5.

The class Λ_r shrinks as r increases. Consequently, the optimized risk shrinks toward one and the optimal headline approaches the favored-design estimate. The ratio class lets a researcher quantify how strongly a favored design must be privileged before the optimal headline changes materially. A useful implementation reports the optimal headline and risk over a grid of r . This allows one to see where $\hat{\theta}_r^*(Y)$ changes sign or crosses a policy threshold, or to examine where $R_r^*(Y)$ is low relative to some benchmark—such as the simplex-optimized risk $R_\Delta^*(Y)$.

5 Asymptotic Results

The foregoing results are developed under normally distributed estimates $Y \sim N(\mu, \Sigma)$ and a known covariance matrix Σ . In practice, the researcher can implement the headline estimation procedure with non-normal estimates and an estimated covariance matrix. I study this plug-in procedure below, which I propose as the default practical implementation.

For sample size $n \rightarrow \infty$, let $Y_n = (Y_{1,n}, \dots, Y_{K,n})'$ denote the vector of design estimates and let $\tilde{\Sigma}_n$ denote an estimated positive definite covariance matrix for Y_n . The corresponding design standard errors and standardized discrepancies are

$$\tilde{\sigma}_{k,n} = \sqrt{v_k' \tilde{\Sigma}_n v_k}, \quad D_{k,n}(a|Y_n) = \frac{|Y_{k,n} - a|}{\tilde{\sigma}_{k,n}}.$$

Fix a nonempty, compact, and convex class of design priors $\Lambda \subseteq \Delta(\mathcal{V})$. The proposed plug-in implementation simply replaces (Y, Σ) in Section 4 with $(Y_n, \tilde{\Sigma}_n)$, which yields

$$\hat{\theta}_{n,\Lambda}^* = \operatorname{argmin}_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k D_{k,n}(a|Y_n)^2, \quad R_{n,\Lambda}^* = \min_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k D_{k,n}(a|Y_n)^2 + 1.$$

In words, the plug-in optimal headline and plug-in risk $(\hat{\theta}_{n,\Lambda}^*, R_{n,\Lambda}^*)$ are obtained by applying the headline estimation criterion directly to the estimates $Y_{k,n}$ and standard errors $\tilde{\sigma}_{k,n}$. The formulas developed previously for classes $(\Lambda_\Delta, \Lambda_\epsilon(\lambda^0), \Lambda_r)$ therefore apply directly.

Assumption 1 (Local design disagreement). *For target-valid design $k_0 \in [K]$, parameter $\theta \in \mathbb{R}$, and local discrepancy vector $b = (b_1, \dots, b_K)' \in \mathbb{R}^K$ satisfying $b_{k_0} = 0$,*

$$\sqrt{n}(Y_n - \mu_n) \xrightarrow{d} N(0, \Sigma), \quad \mu_n = \theta \mathbf{1} + \frac{b}{\sqrt{n}},$$

where Σ is positive definite and $n\tilde{\Sigma}_n \xrightarrow{p} \Sigma$.

Assumption 1 considers differences in estimands that are on the same order as sampling

uncertainty. The target-valid design estimand satisfies $\mu_{k_0,n} = \theta$, while the other estimands may differ from θ by $b_k/\sqrt{n}$. The purpose of this local sequence is to produce a nondegenerate limit decision problem. Define objects for the Gaussian limit decision problem as

$$\hat{\theta}_\Lambda^*(z) = \operatorname{argmin}_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(z_k - a)^2}{\sigma_k^2}, \quad R_\Lambda^*(z) = \min_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(z_k - a)^2}{\sigma_k^2} + 1, \quad z \in \mathbb{R}^K.$$

Proposition 6 (Local asymptotic behavior of the plug-in procedure). *Under Assumption 1,*

$$\sqrt{n}(\hat{\theta}_{n,\Lambda}^* - \theta) \xrightarrow{d} \hat{\theta}_\Lambda^*(Z), \quad R_{n,\Lambda}^* \xrightarrow{d} R_\Lambda^*(Z), \quad Z \sim N(b, \Sigma).$$

Proof. See Appendix A.6. □

Proposition 6 gives the first-order asymptotic behavior of the plug-in headline estimation procedure. The normalized headline has the same limit as the optimal headline in the Gaussian limit decision problem, while the plug-in risk converges to the corresponding limit risk.

I next consider asymptotics with persistent design disagreement, where differences among design estimands do not shrink with sampling uncertainty.

Proposition 7 (Persistent design disagreement). *Suppose $Y_n \xrightarrow{p} \mu^0$ and $n\tilde{\Sigma}_n \xrightarrow{p} \Sigma$, where Σ is positive definite. Define*

$$\theta_\Lambda^{\lim}(\mu^0) = \operatorname{argmin}_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(\mu_k^0 - a)^2}{\sigma_k^2}, \quad R_\Lambda^{\lim}(\mu^0) = \min_{a \in \mathbb{R}} \max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(\mu_k^0 - a)^2}{\sigma_k^2}.$$

Then

$$\hat{\theta}_{n,\Lambda}^* \xrightarrow{p} \theta_\Lambda^{\lim}(\mu^0), \quad \frac{R_{n,\Lambda}^*}{n} \xrightarrow{p} R_\Lambda^{\lim}(\mu^0).$$

If every design receives positive probability under at least one prior in Λ and the coordinates of μ^0 are not all equal, then $R_\Lambda^{\lim}(\mu^0) > 0$ so that $R_{n,\Lambda}^ \xrightarrow{p} \infty$.*

Proof. See Appendix A.7. □

Taken together, Propositions 6 and 7 characterize the plug-in procedure in two asymptotic regimes. Under local disagreement, the differences among design estimands remain on the same order as sampling uncertainty and the normalized plug-in headline converges in distribution to the optimal headline for the corresponding Gaussian limit problem. Under persistent disagreement with $R_\Lambda^{\lim}(\mu^0) > 0$, the headline converges in probability to a deterministic compromise that need not equal the target, while the diverging risk increasingly makes visible the difficulty of reconciling a common headline across designs.

6 Empirical Applications

This section illustrates the headline estimation procedure in empirical applications to three papers: [Chen, Chen and Yang \(2026\)](#), [Garin et al. \(2025\)](#), and [Ganong et al. \(2024\)](#). For each application, I define the target parameter, candidate designs, baseline headline a_B , and class of design priors Λ . The outputs of interest include the optimal headline $\hat{\theta}_\Lambda^*$, baseline risk R_Λ^B , optimized risk R_Λ^* , active designs or least favorable design priors, and percentage reduction in risk when moving from the baseline to the optimal headline:

$$\left(1 - \frac{R_\Lambda^*}{R_\Lambda^B}\right) \times 100.$$

The oracle risk is one, so a risk value of R means a $(R - 1) \times 100$ percentage increase above the oracle value. When interpreting magnitudes, I also report the reconciliation cost $c_\Lambda^* = \sqrt{R_\Lambda^* - 1}$. Following [Remarks 6 and 7](#), I use $c = 2$ and $R = 5$ as simplex-calibrated benchmarks for cost and risk. Under the simplex class, exceeding this benchmark means that the 95% design-specific credible intervals fail to share a common point. Although this interval-overlap characterization does not hold more generally for other classes, the cost and risk calibrations still serve as useful benchmarks. The calculations below are based on point estimates and standard errors from the papers' published tables. Because these inputs are rounded, the resulting outputs may differ slightly from calculations based on the papers' underlying data.

6.1 Returning to the Courtroom-Broadcasting Example

[Section 2](#) introduced the courtroom-broadcasting setting of [Chen, Chen and Yang \(2026\)](#) and gave results for their all-litigants sample and the simplex class Λ_Δ . I now return to that example and additionally consider the two-litigant sample and ratio classes Λ_r that privilege the paper's primary DiD design over their alternative Bartik IV design.

- *Target.* The effect of broadcasting intensity on the female-male plaintiff win-rate gap.
- *Candidate designs.* The DiD and Bartik IV designs.⁹
- *Baseline headline.* The DiD estimate.
- *Design-prior class.* The simplex class Λ_Δ , followed by the favored-DiD ratio class Λ_r .

[Table 1](#) reports the simplex results using the published estimates in [Chen, Chen and Yang \(2026, Tables 2 and 3\)](#). The all-litigants row reproduces the calculations from [Section 2](#). For the two-litigant sample, the optimal headline is $\hat{\theta}_\Delta^* = 0.0651$, so a 10 percentage-point increase in broadcasting intensity corresponds to a 0.651 percentage-point narrowing of the gender gap.

⁹[Chen, Chen and Yang \(2026, page 1576\)](#) say that “While the Bartik IV and DID strategies rely on distinct identifying assumptions, our results remain remarkably consistent across both approaches.”

Table 1: Simplex-optimal headline estimation: [Chen, Chen and Yang \(2026\)](#)

Sample	DiD estimate (SE)	Bartik IV estimate (SE)	Optimal headline $\hat{\theta}_\Delta^*$	Baseline risk R_Δ^B	Optimal risk R_Δ^*	Risk reduction
All litigants	0.0394 (0.00255)	0.0726 (0.00599)	0.0493	31.72	16.11	49.2%
Two litigants	0.0536 (0.00336)	0.0889 (0.00699)	0.0651	26.50	12.63	52.3%

Notes. The estimates and standard errors are reported in [Chen, Chen and Yang \(2026, Tables 2 and 3\)](#). The DiD estimate serves as the baseline headline a_B . Risk reduction is $(1 - (R_\Delta^*/R_\Delta^B)) \times 100$.

The optimal headlines reduce the risk corresponding to baseline DiD headlines by roughly half. Even after optimization, however, risk remains high because the differences between the DiD and Bartik IV estimates are large relative to their small standard errors. The optimized risks are 16.11 and 12.63 times the oracle risk, corresponding to costs $c_\Delta^* \approx 3.89$ and $c_\Delta^* \approx 3.41$. Both exceed the cost benchmark. Equivalently, the approximately 95% design-specific credible intervals fail to share a common point. Thus, while the simplex-optimal headlines preserve the conclusion that broadcasting narrows the gender gap, the DiD and IV designs do not tightly reconcile one exact magnitude under unrestricted design ambiguity.

I now consider the structured ambiguity given by favored-design ratio classes that privilege DiD. In particular, a favored-DiD ratio class Λ_r requires DiD to be at least $r \geq 1$ times as likely to be target-valid as Bartik IV. Since $\sigma_{\text{DiD}} < \sigma_{\text{IV}}$, Proposition 5 implies that the least favorable prior places probabilities $r/(r+1)$ and $1/(r+1)$ on DiD and IV, respectively. Table 2 reports the sensitivity path $r \mapsto (\hat{\theta}_r^*, R_r^*)$ for a range of r .

Table 2: Favored-DiD ratio sensitivity: [Chen, Chen and Yang \(2026\)](#)

Design-prior class	Minimum DiD probability	All litigants		Two litigants	
		Headline $\hat{\theta}_r^*$	Risk R_r^*	Headline $\hat{\theta}_r^*$	Risk R_r^*
Unrestricted simplex	—	0.0493	16.113	0.0651	12.632
$r = 1$	0.500	0.0445	14.003	0.0602	11.358
$r = 2$	0.667	0.0422	10.389	0.0573	8.621
$r = 5$	0.833	0.0406	5.941	0.0552	5.063
$r = 10$	0.909	0.0400	3.743	0.0544	3.266
$r = 25$	0.962	0.0396	2.173	0.0539	1.972
$r = 100$	0.990	0.0395	1.304	0.0537	1.252
$r \rightarrow \infty$	1.000	0.0394	1.000	0.0536	1.000

Notes. The favored-DiD ratio class requires $\lambda_{\text{DiD}} \geq r\lambda_{\text{IV}}$, so the minimum probability assigned to DiD is $r/(r+1)$. At every finite r , the least favorable design prior places probabilities $r/(r+1)$ and $1/(r+1)$ on DiD and IV. The unrestricted row reproduces the simplex class calculations from Table 1.

At $r = 5$, the reconciliation costs are approximately 2.22 and 2.02, which remain just above the simplex-calibrated benchmark $c = 2$. At $r = 10$, the costs fall to approximately 1.66 and 1.51, respectively. Thus, under a favored-DiD ratio of this strength, the reconciliation costs fall below the benchmark and the optimal headlines are close to the DiD estimates.

The $r = 10$ class requires the prior probability on DiD to be at least $10/11 \approx 0.909$, so no admissible prior places more than approximately 9.1% probability on IV. This restriction may be substantively reasonable if the Bartik IV design is viewed primarily as a robustness check for the paper’s favored DiD design.¹⁰ Overall, this sensitivity analysis suggests that, under modest design ambiguity that favors DiD, there is scope for the DiD and Bartik IV designs to tightly reconcile one headline for the effect of broadcasting intensity on the gender win-rate gap.

6.2 When the Published Headline Is Already Adequate

Garin et al. (2025) estimate the long-run effects of a 12-month incarceration sentence on labor-market outcomes and later incarceration exposure using two IV designs based on discontinuities in sentencing guidelines in North Carolina and random judge assignments in Ohio. The authors headline precision-weighted averages across the two states. The corresponding IVs need not identify the same local average treatment effect: the two designs operate in different populations and shift different sentencing margins. However, if one imagines a latent “effect of a 12-month incarceration sentence for marginal felony defendants,” then both IV designs plausibly speak to that target parameter, with ambiguity over which design provides the more appropriate link.

- *Target.* For each outcome of interest, the effect of a 12-month incarceration sentence for marginal felony defendants.
- *Candidate designs.* The North Carolina and Ohio IV designs.¹¹
- *Baseline headline.* The published precision-weighted average.
- *Design-prior class.* The simplex class Λ_Δ , followed by the perturbation class $\Lambda_\epsilon(\lambda^0)$ under an equal-probability baseline prior $\lambda^0 = (1/2, 1/2)'$.

Panel A of Table 3 compares the published precision-weighted averages a_B with the simplex-optimal headlines $\hat{\theta}_\Delta^*$. For annual W-2 earnings, the baseline and optimized risks are 1.057 and

¹⁰Chen, Chen and Yang (2026, page 1576) seem to cast the IV design as a robustness check: “The primary empirical specification is the generalized difference-in-differences (DID) with continuous treatment, controlling for court-area and year-quarter fixed effects . . . A potential threat to identification is that the aggregate reform intensity may be endogenous. For instance, it could be contaminated by unobserved confounding factors that vary across court-area and over time. To address this concern, we strengthen our analysis by implementing a Bartik-like instrumental variable (IV) strategy. . . The IV estimates are only slightly larger than the corresponding baseline DID estimates, suggesting that potential endogeneity concerns are limited. While the Bartik IV and DID strategies rely on distinct identifying assumptions, our results remain remarkably consistent across both approaches.”

¹¹Garin et al. (2025, page 504) note that “Using multiple research designs allows us to test the sensitivity of our results to empirical strategy.”

1.041; for cumulative earnings, they are 1.376 and 1.262. Adopting the simplex-optimal headline reduces risk by 1.5% and 8.3%, respectively. The risks remain close to the oracle benchmark, and the modest improvements show that the published precision-weighted averages are already close to optimal headlines for the target parameter.

Panel B presents a perturbation class sensitivity path $\epsilon \mapsto \hat{\theta}_\epsilon^*$ under the baseline prior that places equal probability on the North Carolina and Ohio IV designs. Treating the two estimates from the distinct state samples as independent, the paper’s precision-weighted average has the same inverse-variance form as the uniform-prior headline at $\epsilon = 0$. Calculations from the state-design estimates reproduce the published averages up to small rounding differences. As ϵ increases, the optimal headline moves toward the inverse-standard-deviation simplex-optimal headline. The column $\bar{\epsilon}$ reports the smallest perturbation level at which the sensitivity path reaches the simplex-optimal headline. The perturbation class results are consistent with those from the simplex class, so below I focus on the latter.

Table 3: Simplex-optimal headline estimation and perturbation sensitivity: [Garin et al. \(2025\)](#)

Panel A. Published and simplex-optimal headlines

Outcome	N. Carolina (SE)	Ohio (SE)	Published average a_B	Baseline risk R_Δ^B	Optimal headline $\hat{\theta}_\Delta^*$	Optimal risk R_Δ^*	Risk reduction
Annual W-2	113.45 (223.8) −2,675	233.97 (371.5) −3,881	145.54	1.057	158.76	1.041	1.5%
Cumulative W-2	(782) 0.024	(1,576) 0.004	−2,914	1.376	−3,075	1.262	8.3%
Any W-2	(0.010) 3.20	(0.013) 13.50	0.016	1.852	0.015	1.756	5.2%
Days incarcerated	(3.31)	(2.52)	9.72	4.880	9.05	4.121	15.5%

Panel B. Uniform-prior perturbation path

Outcome	$\hat{\theta}_0^*$	$\hat{\theta}_{0.10}^*$	$\hat{\theta}_{0.25}^*$	$\hat{\theta}_1^*$	$\bar{\epsilon}$	R_0^*	R_1^*
Annual W-2	145.54	150.48	158.76	158.76	0.248	1.039	1.041
Cumulative W-2	−2,913	−2,954	−3,026	−3,075	0.337	1.235	1.262
Any W-2	0.0166	0.0156	0.0153	0.0153	0.130	1.743	1.756
Days incarcerated	9.72	9.23	9.05	9.05	0.136	4.065	4.121

Notes. The estimates, standard errors, and published averages are reported in [Garin et al. \(2025, Table III\)](#). “Annual W-2” denotes annual W-2 earnings, “Cumulative W-2” denotes cumulative W-2 earnings, and standard errors appear below the corresponding state-design estimates. Annual W-2 earnings, any W-2 earnings, and days incarcerated average outcomes over years 5–9 post-case-filing. Cumulative W-2 earnings are measured as of five years post-case-filing. Panel A evaluates the published average under the simplex class and compares it with the simplex-optimal headline. The risk reduction is $(1 - (R_\Delta^*/R_\Delta^B)) \times 100$. Panel B uses a uniform prior $\lambda^0 = (1/2, 1/2)'$. Here, $\bar{\epsilon} = |\sigma_1 - \sigma_2|/(\sigma_1 + \sigma_2)$ is the smallest perturbation level at which the ϵ -perturbation-optimal $\hat{\theta}_\epsilon^*$ reaches the simplex-optimal $\hat{\theta}_\Delta^*$. Because R_Δ^B evaluates the published rounded average while R_0^* evaluates the recomputed uniform-prior headline, the two risks can differ slightly.

The implications of the headline estimation output closely mirror the paper’s conclusions.

The simplex-optimal headlines for annual W-2 earnings and the probability of having any W-2 earnings are small and positive, reinforcing the finding of no meaningful long-run reduction in earnings or employment. For cumulative W-2 earnings, the headline implies an earnings reduction of about \$3,075, close to the paper’s baseline \$2,914, and thus supports the conclusion that earnings lost during incarceration are not subsequently recovered. For long-run incarceration exposure, the headline implies roughly 9 additional days incarcerated per year. Its larger risk indicates weaker agreement about the exact amount of residual incarceration exposure across states, although both designs imply that most of the initial sentence effect has dissipated.

6.3 A Stable High-MPC Conclusion Across Policy Episodes

Ganong et al. (2024) estimate one-month marginal propensities to consume (MPCs) out of unemployment benefits for unemployed U.S. households using six designs based on different policy changes and comparison strategies in the context of the COVID-19 pandemic. A one-month MPC is the change in spending during the first month after a benefit change, divided by the benefit change. The first design compares unemployed households that receive benefits promptly with households that face processing delays. Four designs study the expiration or onset of \$600 and \$300 supplements by comparing unemployed households with matched employed households. The sixth design compares unemployed households in U.S. states that ended the \$300 supplement in June with those in states that ended it in September. The authors report the range of estimated MPCs across designs—0.27 to 0.42—and emphasize that spending responds sharply across all designs. They describe the waiting-for-benefits design as their sharpest identification strategy, while noting that it measures the response to total benefits during the unusually uncertain opening months of the pandemic; the other designs instead isolate supplement changes at other dates. I treat the six designs as candidate links to a latent target one-month MPC out of unemployment benefits.

- *Target.* The one-month MPC out of unemployment benefits.
- *Candidate designs.* The six designs summarized in Ganong et al. (2024, Table 1).¹²
- *Baseline headline.* The waiting-for-benefits design estimate—0.42—treated as a natural baseline because the paper calls that design its sharpest.
- *Design-prior class.* The simplex class Λ_{Δ} .

Table 4 reports the simplex-optimal headline output for all six designs and, as a narrower comparison, for the four matched-employed supplement-change designs in Ganong et al. (2024, Table 1, rows 2–5). Across all six designs, the optimal headline is $\hat{\theta}_{\Delta}^* = 0.345$ and the optimized

¹²Ganong et al. (2024, page 2907) say that “Each of these empirical exercises has distinct advantages and disadvantages, but they all lead to the same conclusion.”

risk is $R_{\Delta}^* = 57.25$, determined by the 0.42 waiting estimate and the 0.27 September-expiration estimate. Taking $a_B = 0.42$ as the baseline headline gives a baseline risk of $R_{\Delta}^B = 226.00$; adopting the optimal headline reduces risk by 74.7%. Restricting attention to rows 2–5 yields an optimal headline of $\hat{\theta}_{\Delta}^* = 0.297$ and a smaller optimized risk of $R_{\Delta}^* = 8.11$.

Table 4: Simplex-optimal headline estimation: [Ganong et al. \(2024\)](#)

Candidate design set	Estimate range	Optimal headline $\hat{\theta}_{\Delta}^*$	Optimal risk R_{Δ}^*	Active designs
All six designs	0.27–0.42	0.345	57.25	Waiting for benefits; \$300 expiration (September states)
Matched-employed supplement changes, rows 2–5	0.27–0.35	0.297	8.11	\$300 expiration (June states); \$300 expiration (September states)

Notes. The estimates and standard errors are reported in [Ganong et al. \(2024, Table 1, rows 1–6\)](#). The “All six designs” calculation uses the six published estimates and standard errors. The narrower calculation uses rows 2–5, which compare supplement changes using unemployed and matched-employed households.

The all-design headline of 0.345 implies that about 35 cents of an additional benefit dollar is spent in the first month, while the supplement-change headline of 0.297 implies about 30 cents. Both magnitudes support the conclusion that spending responses to expanded unemployment benefits are large. This evidence is an input to the paper’s broader conclusion that temporary supplements can provide substantial demand support during recessions. At the same time, the all-design risk indicates that the six designs do not tightly support one exact MPC magnitude. The same is true for the supplement-change design set, although this set yields a much lower risk. Overall, the design-specific evidence supports a stable high-MPC conclusion more tightly than it supports one exact cross-episode value.

7 Conclusion

Researchers often use multiple research designs to study a single question, but lack a dedicated framework for producing a headline estimate. This paper formalizes the headline estimation problem as a decision under ambiguity about which design supplies the valid identifying link to a scalar target parameter, and proposes a conditional Gamma-minimax procedure for headline estimation under ambiguity. The procedure yields an ambiguity-optimal headline together with a measure of worst-case posterior risk given the candidate designs. The empirical applications illustrate several uses of the headline estimation procedure across different ambiguity classes. The simplex class can separate agreement about the direction of an effect from reconciliation about its exact magnitude across designs. The perturbation class can show when an existing precision-weighted average is already close to optimal. Finally, the ratio class can quantify how strongly a favored design must be privileged before an optimal headline yields low risk.

References

- Agrawal, Keshav, Susan Athey, Ayush Kanodia, Shanjukta Nath, and Emil Palikot.** 2026. “The economics of algorithmic personalization: Evidence from an educational technology platform.” National Bureau of Economic Research.
- Alsan, Marcella, Luca Braghieri, Sarah Eichmeyer, Minjeong Joyce Kim, Stefanie Stantcheva, and David Y Yang.** 2023. “Civil liberties in times of crisis.” *American Economic Journal: Applied Economics*, 15(4): 389–421.
- Andrews, Isaiah, and Jesse M Shapiro.** 2026. “Communicating scientific uncertainty via approximate posteriors.” *Econometrica*, 94(3): 843–875.
- Armstrong, Timothy B, Patrick Kline, and Liyang Sun.** 2025. “Adapting to misspecification.” *Econometrica*, 93(6): 1981–2005.
- Baran, Cavit, Janet Currie, Bahadir Dursun, and Erdal Tekin.** 2025. “Clean rides, healthy lives: The impact of electric vehicle adoption on air quality and infant health.” National Bureau of Economic Research.
- Berger, James O.** 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. 2 ed., New York:Springer.
- Betrò, Bruno, and Fabrizio Ruggeri.** 1992. “Conditional Gamma-minimax actions under convex losses.” *Communications in Statistics-Theory and Methods*, 21(4): 1051–1066.
- Bhattacharya, Rohit, Ina Ocelli, and Ted Westling.** 2026. “Robust Weighted Triangulation of Causal Effects Under Model Uncertainty.” *arXiv preprint arXiv:2603.01119*.
- Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes.** 2017. “Life-cycle earnings, education premiums, and internal rates of return.” *Journal of Labor Economics*, 35(4): 993–1030.
- Cheng, Xu, Zhipeng Liao, and Ruoyao Shi.** 2019. “On uniform asymptotic risk of averaging GMM estimators.” *Quantitative Economics*, 10(3): 931–979.
- Chen, Heng, Yuyu Chen, and Qingxu Yang.** 2026. “Women in the courtroom: Technology and justice.” *Review of Economic Studies*, 93(3): 1574–1601.
- Christensen, Timothy, Hyungsik Roger Moon, and Frank Schorfheide.** 2026. “Optimal decision rules when payoffs are partially identified.” *Review of Economic Studies*, rda017.
- DasGupta, Anirban, and William J Studden.** 1989. “Frequentist behavior of robust Bayes estimates of normal means.” *Statistics & Risk Modeling*, 7(4): 333–362.

- Davis, Jonathan, Samuel Norris, Jadon Schmitt, Yotam Shem-Tov, and Chelsea Strickland.** 2025. “Mobile crisis response teams support better policing: Evidence from CAHOOTS.” National Bureau of Economic Research.
- Dunn, Abe, Joshua D Gottlieb, Adam Hale Shapiro, Daniel J Sonnenstuhl, and Pietro Tebaldi.** 2024. “A denial a day keeps the doctor away.” *The Quarterly Journal of Economics*, 139(1): 187–233.
- Erten, Bilge, Joseph E Stiglitz, and Eric Verhoogen.** 2025. “Employment impacts of the CHIPS act.” *Brookings Papers on Economic Activity*, 2025(1): 1–63.
- Fang, Hanming, Ming Li, Long Wang, and Yang Yang.** 2026. “High-speed rail and china’s electric vehicle adoption miracle.” *Journal of Public Economics*, 260: 105705.
- Ganong, Peter, Fiona Greig, Pascal Noel, Daniel M Sullivan, and Joseph Vavra.** 2024. “Spending and job-finding impacts of expanded unemployment benefits: Evidence from administrative micro data.” *American Economic Review*, 114(9): 2898–2939.
- Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pencenno, Evan K Rose, Yotam Shem-Tov, and Jeffrey Weaver.** 2025. “The impact of incarceration on employment, earnings, and tax filing.” *Econometrica*, 93(2): 503–538.
- Giacomini, Raffaella, Toru Kitagawa, and Matthew Read.** 2021. “Robust Bayesian analysis for econometrics.” *CEPR Discussion Paper No. DP16488*.
- Gilboa, Itzhak, and David Schmeidler.** 1989. “Maxmin expected utility with non-unique prior.” *Journal of mathematical economics*, 18(2): 141–153.
- Green, Edwin J, and William E Strawderman.** 1991. “A James-Stein type estimator for combining unbiased and possibly biased estimators.” *Journal of the American Statistical Association*, 86(416): 1001–1006.
- Han, Hee Seo, David Hirshleifer, Jinfei Sheng, and Zheng Sun.** 2025. “Trading in twilight: Sleep, mental alertness, and stock market trading.” National Bureau of Economic Research.
- Hansen, Bruce E.** 2007. “Least squares model averaging.” *Econometrica*, 75(4): 1175–1189.
- Hastings, Justine, and Jesse M Shapiro.** 2018. “How are SNAP benefits spent? Evidence from a retail panel.” *American Economic Review*, 108(12): 3493–3540.

- Hoeting, Jennifer A, David Madigan, Adrian E Raftery, and Chris T Volinsky.** 1999. “Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors).” *Statistical Science*, 14(4): 382–417.
- Jain, Manisha, Corina Mommaerts, and Jeffrey Weaver.** 2026. “The Limits of Targeted Hiring Subsidies: Evidence from the Work Opportunity Tax Credit.” National Bureau of Economic Research.
- Laliberté, Jean-William.** 2021. “Long-term contextual effects in education: Schools and neighborhoods.” *American Economic Journal: Economic Policy*, 13(2): 336–377.
- Lim, Cheaheon, and Yechan Park.** 2026. “Dynamically Consistent Statistical Decisions.” *arXiv preprint arXiv:2607.10519*.
- Ng, Kevin, and Evan Riehl.** 2024. “The returns to STEM programs for less-prepared students.” *American Economic Journal: Economic Policy*, 16(2): 37–77.
- Park, Yechan, and Yuya Sasaki.** 2026. “Choosing A Headline Estimand from Matching, DID, and Hybrid Designs: A Minimax-Regret Approach.” *arXiv preprint arXiv:2606.20435*.
- Sager, Lutz, and Gregor Singer.** 2025. “Clean identification? the effects of the clean air act on air pollution, exposure disparities, and house prices.” *American Economic Journal: Economic Policy*, 17(1): 1–36.
- Sion, Maurice.** 1958. “On General Minimax Theorems.” *Pacific Journal of Mathematics*, 8(1): 171–176.
- Stoye, Jörg.** 2012. “New perspectives on statistical decisions under ambiguity.” *Annu. Rev. Econ.*, 4(1): 257–282.
- Vilfort, Vod.** 2026. “Robust Inference for Weighted Estimands.” *arXiv preprint arXiv:2607.07524*.

Supplemental Appendix to “Headline Estimation with Multiple Research Designs”

Vod Vilfort

A Proofs

A.1 Proof of Proposition 1

The oracle posterior risk conditional on $v = v_k$ is

$$\int_{\mathcal{M}} w_k (\mu_k - Y_k)^2 \pi_0(d\mu|Y) = w_k \{ (Y_k - Y_k)^2 + \sigma_k^2 \} = w_k \sigma_k^2, \quad \forall k \in [K].$$

Because $w_k > 0$ for all $k \in [K]$, equalizing the above across realizations of v requires $w_k \sigma_k^2 = c$ for a constant $c > 0$, and thus $w_k = c / \sigma_k^2$.

A.2 Proof of Proposition 2

For each $\lambda \in \Lambda$, the function

$$a \mapsto g(a, \lambda) = \sum_{k=1}^K \lambda_k D_k(a|Y)^2$$

is strongly convex with second derivative at least $2 / \max_k \sigma_k^2$. Hence, $R_\Lambda(a|Y) - 1$ is strongly convex and therefore has a unique minimizer $\hat{\theta}_\Lambda^*(Y)$.

For the characterization, consider the compact interval $\mathcal{I} = [\min_k Y_k, \max_k Y_k]$. Projecting any $a \notin \mathcal{I}$ onto $\mathcal{I}$ weakly reduces every $D_k(a|Y)$ and strictly reduces at least one. The minimization can thus be restricted to $a \in \mathcal{I}$. The function $(a, \lambda) \mapsto g(a, \lambda)$ is continuous, convex in a , and affine in λ . Thus, since $\mathcal{I}$ and Λ are compact and convex, the minimax theorem of [Sion \(1958\)](#) gives

$$\min_{a \in \mathcal{I}} \max_{\lambda \in \Lambda} g(a, \lambda) = \max_{\lambda \in \Lambda} \min_{a \in \mathcal{I}} g(a, \lambda).$$

The left side is uniquely minimized at $a = \hat{\theta}_\Lambda^*(Y)$. Compactness, continuity, and Berge’s maximum theorem ensure that the right side attains its maximum V at any least favorable $\lambda^* \in \Lambda$. Thus,

$$V = \min_{a \in \mathcal{I}} g(a, \lambda^*) \leq g(\hat{\theta}_\Lambda^*(Y), \lambda^*) \leq \max_{\lambda \in \Lambda} g(\hat{\theta}_\Lambda^*(Y), \lambda) = V.$$

Both inequalities are therefore equalities. This means λ^* is a worst-case prior at $a = \hat{\theta}_\Lambda^*(Y)$ and

$\hat{\theta}_\Lambda^*(Y)$ minimizes $g(a, \lambda^*)$. The first-order condition for the strictly convex quadratic $g(a, \lambda)$ is

$$\sum_{k=1}^K \lambda_k \frac{a - Y_k}{\sigma_k^2} = 0 \iff a(\lambda|Y) = \frac{\sum_{k=1}^K \lambda_k Y_k / \sigma_k^2}{\sum_{k=1}^K \lambda_k / \sigma_k^2}, \quad \forall \lambda \in \Delta(\mathcal{V}).$$

In particular, $\hat{\theta}_\Lambda^*(Y) = a(\lambda^*|Y)$. The weighted-variance formula with weights λ_k / σ_k^2 yields

$$\rho(\lambda|Y) = \sum_{k=1}^K \frac{\lambda_k}{\sigma_k^2} \{Y_k - a(\lambda|Y)\}^2 = \frac{\sum_{j < k} (Y_j - Y_k)^2 (\lambda_j / \sigma_j^2) (\lambda_k / \sigma_k^2)}{\sum_{\ell=1}^K \lambda_\ell / \sigma_\ell^2}.$$

Evaluating $g(\hat{\theta}_\Lambda^*(Y), \lambda^*) + 1 = \rho(\lambda^*|Y) + 1$ yields the formula for $R_\Lambda^*(Y)$.

For $K=2$, let $\lambda(p) = (p, 1-p)'$. Since Λ is nonempty, compact, and convex, the corresponding set $\{p \in [0, 1] : \lambda(p) \in \Lambda\}$ of probabilities on design 1 is an interval $\mathcal{I}_\Lambda = [\underline{p}_\Lambda, \bar{p}_\Lambda]$. For fixed $p \in [0, 1]$, the first-order condition above gives

$$a(p) = a(\lambda(p)|Y) = \frac{pY_1/\sigma_1^2 + (1-p)Y_2/\sigma_2^2}{p/\sigma_1^2 + (1-p)/\sigma_2^2}, \quad \rho(p) = \rho(\lambda(p)|Y) = \frac{p(1-p)(Y_1 - Y_2)^2}{p\sigma_2^2 + (1-p)\sigma_1^2}.$$

When $Y_1 \neq Y_2$, differentiation yields

$$\rho'(p) = \frac{(Y_1 - Y_2)^2 \{\sigma_1^2(1-p)^2 - \sigma_2^2 p^2\}}{\{p\sigma_2^2 + (1-p)\sigma_1^2\}^2}.$$

The derivative is positive for $p < p_\Delta$, zero at $p = p_\Delta$, and negative for $p > p_\Delta$. Thus, the maximizer of $\rho(p)$ over $\mathcal{I}_\Lambda$ is the projection p_Λ^* of p_Δ onto this interval. Evaluating $a(p)$ and $\rho(p) + 1$ at p_Λ^* yields the displayed headline and risk formulas, and strict monotonicity on either side of p_Δ gives uniqueness of the least favorable prior. If $Y_1 = Y_2$, then $\rho(p) = 0$ for every p , so every admissible design prior $\lambda \in \Lambda$ is least favorable and the same formulas give headline $Y_1 = Y_2$ and risk one.

A.3 Proof of Proposition 3

For $c \geq 0$, there exists a satisfying $D_k(a|Y) \leq c$ for every k if and only if

$$a \in \bigcap_{k=1}^K [Y_k - c\sigma_k, Y_k + c\sigma_k].$$

Given this interval-containment condition, the intersection is nonempty if and only if

$$\max_{j \in [K]} \{Y_j - c\sigma_j\} \leq \min_{k \in [K]} \{Y_k + c\sigma_k\}.$$

Equivalently, for every ordered pair (j, k) ,

$$Y_j - Y_k \leq c(\sigma_j + \sigma_k).$$

The smallest feasible interval multiplier is

$$c_{\Delta}^*(Y) = \max_{j,k \in [K]} \frac{Y_j - Y_k}{\sigma_j + \sigma_k} = \max_{j < k} \frac{|Y_j - Y_k|}{\sigma_j + \sigma_k}.$$

If (j^*, k^*) is an active pair, relabel its members if necessary so that $Y_{j^*} \geq Y_{k^*}$. Then

$$\begin{aligned} \max_{j \in [K]} \{Y_j - c_{\Delta}^*(Y) \sigma_j\} &\geq \left\{ Y_{j^*} - \frac{Y_{j^*} - Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}} \sigma_{j^*} \right\} = \frac{\sigma_{k^*} Y_{j^*} + \sigma_{j^*} Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}} \\ \min_{k \in [K]} \{Y_k + c_{\Delta}^*(Y) \sigma_k\} &\leq \left\{ Y_{k^*} + \frac{Y_{j^*} - Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}} \sigma_{k^*} \right\} = \frac{\sigma_{k^*} Y_{j^*} + \sigma_{j^*} Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}}. \end{aligned}$$

Thus, since $c_{\Delta}^*(Y)$ is the smallest feasible multiplier, the interval-containment condition gives

$$\max_{k \in [K]} \{Y_k - c_{\Delta}^*(Y) \sigma_k\} = \hat{\theta}_{\Delta}^*(Y) = \min_{k \in [K]} \{Y_k + c_{\Delta}^*(Y) \sigma_k\} = \frac{\sigma_{k^*} Y_{j^*} + \sigma_{j^*} Y_{k^*}}{\sigma_{j^*} + \sigma_{k^*}}.$$

The condition also gives $\max_{k \in [K]} D_k(\hat{\theta}_{\Delta}^*(Y)|Y) \leq c_{\Delta}^*(Y)$. Note further that

$$\max_{k \in [K]} D_k(\hat{\theta}_{\Delta}^*(Y)|Y) \geq D_{j^*}(\hat{\theta}_{\Delta}^*(Y)|Y) = D_{k^*}(\hat{\theta}_{\Delta}^*(Y)|Y) = c_{\Delta}^*(Y).$$

It therefore follows that

$$c_{\Delta}^*(Y) = \max_{k \in [K]} D_k(\hat{\theta}_{\Delta}^*(Y)|Y) = \sqrt{R_{\Delta}^*(Y) - 1}, \quad R_{\Delta}^*(Y) = \left(\frac{|Y_{j^*} - Y_{k^*}|}{\sigma_{j^*} + \sigma_{k^*}} \right)^2 + 1.$$

For $K = 2$ and $\lambda(p) = (p, 1-p)'$, the simplex class generates the interval

$$\mathcal{I}_{\Lambda_{\Delta}} = \{p \in [0, 1] : \lambda(p) \in \Lambda_{\Delta}\} = [0, 1].$$

Proposition 2 therefore gives $p_{\Lambda_{\Delta}}^* = p_{\Delta}$, and hence unique $\lambda^* = (p_{\Delta}, 1 - p_{\Delta})'$ when $Y_1 \neq Y_2$.

A.4 Proof of Proposition 4

For any $q \in \mathbb{R}^K$, the definition of $\Lambda_{\epsilon}(\lambda^0)$ implies

$$\max_{\lambda \in \Lambda_{\epsilon}(\lambda^0)} \sum_{k=1}^K \lambda_k q_k = \max_{\eta \in \Delta(\mathcal{V})} \sum_{k=1}^K \left\{ (1 - \epsilon) \lambda_k^0 + \epsilon \eta_k \right\} q_k = (1 - \epsilon) \sum_{k=1}^K \lambda_k^0 q_k + \epsilon \max_{k \in [K]} q_k.$$

Substituting $q_k = D_k(a|Y)^2$ and adding one gives the formula for $R_{\epsilon}(a|Y)$. By Proposition 2, $\hat{\theta}_{\epsilon}^*(Y)$ exists uniquely and a least favorable prior λ_{ϵ}^* can be computed as stated.

I now show continuity and monotonicity of $\epsilon \mapsto \hat{\theta}_\epsilon^*(Y) = a_\epsilon = \operatorname{argmin}_{a \in \mathbb{R}} F_\epsilon(a)$, where

$$F_\epsilon(a) = (1-\epsilon)B(a) + \epsilon M(a), \quad B(a) = \sum_{k=1}^K \lambda_k^0 D_k(a|Y)^2, \quad M(a) = \max_{k \in [K]} D_k(a|Y)^2.$$

Note $a_\epsilon \in [\min_k Y_k, \max_k Y_k]$. On this compact interval, $(a, \epsilon) \mapsto F_\epsilon(a)$ is jointly continuous and F_ϵ has unique minimizer a_ϵ , so Berge's maximum theorem implies continuity of $\epsilon \mapsto a_\epsilon$. If $a_0 = a_1$, this common value uniquely minimizes both B and M , and hence F_ϵ for every ϵ . Suppose that $a_0 < a_1$. Since B and M are strongly convex with unique minimizers a_0 and a_1 , respectively, no F_ϵ can be minimized outside $[a_0, a_1]$: if $a < a_0$, replacing a with a_0 lowers both B and M , while if $a > a_1$, replacing a with a_1 lowers both. On $[a_0, a_1]$, B is strictly increasing and M is strictly decreasing, so $M - B$ is strictly decreasing. Hence,

$$F_\epsilon(a) = B(a) + \epsilon \{M(a) - B(a)\}$$

has decreasing differences in (a, ϵ) . Topkis's theorem for minimization, together with uniqueness, implies $\epsilon \mapsto a_\epsilon$ is nondecreasing. The case $a_0 > a_1$ is symmetric and implies nonincreasing $\epsilon \mapsto a_\epsilon$. The a_ϵ endpoints follow from Proposition 2 for $\Lambda_0(\lambda^0) = \{\lambda^0\}$ and Proposition 3 for $\Lambda_1(\lambda^0) = \Lambda_\Delta$.

For $K = 2$ and $\lambda(p) = (p, 1-p)'$, the perturbation class generates the interval

$$\mathcal{I}_\epsilon = \{p \in [0, 1] : \lambda(p) \in \Lambda_\epsilon(\lambda^0)\} = [(1-\epsilon)p^0, (1-\epsilon)p^0 + \epsilon].$$

Proposition 2 therefore gives p_ϵ^* as the projection of p_Δ onto $\mathcal{I}_\epsilon$ and yields the headline and risk formulas. If $p_\Delta \in \mathcal{I}_\epsilon$, then $p_\epsilon^* = p_\Delta$ so that $\hat{\theta}_\epsilon^*(Y) = \hat{\theta}_\Delta^*(Y)$.

A.5 Proof of Proposition 5

Every $\lambda \in \Lambda_r$ satisfies $\lambda_1 > 0$. For $k \neq 1$, define $p_k = r\lambda_k/\lambda_1 \in [0, 1]$. The simplex constraint gives

$$\lambda_1 = \frac{r}{r + \sum_{k \neq 1} p_k}, \quad \lambda_k = \frac{p_k}{r + \sum_{j \neq 1} p_j}, \quad k \neq 1.$$

Conversely, these formulas map every $p \in [0, 1]^{K-1}$ to a design prior in Λ_r . The support function for Λ_r is given by

$$\max_{\lambda \in \Lambda_r} \sum_{k=1}^K \lambda_k q_k = \max_{p \in [0, 1]^{K-1}} H(p), \quad H(p) = \frac{rq_1 + \sum_{k \neq 1} p_k q_k}{r + \sum_{k \neq 1} p_k}, \quad q \in \mathbb{R}^K.$$

For each $k \neq 1$, $H(p)$ can be expressed in the form $(A + p_k q_k)/(B + p_k)$, whose partial derivative with respect to p_k has a sign that does not depend on p_k . A maximizer of $H(p)$ can therefore be chosen with $p_k \in \{0, 1\}$ for each $k \neq 1$. That is, $\max_{p \in [0, 1]^{K-1}} H(p) = \max_{p \in \{0, 1\}^{K-1}} H(p)$.

Let $S = \{k \neq 1 : p_k = 1\}$ and $m = |S|$. The corresponding prior λ^S places probability $r/(r+m)$

on design 1, probability $1/(r+m)$ on $k \in S$, and probability zero on other designs, yielding

$$H(p) = \frac{rq_1 + \sum_{k \in S} q_k}{r+m}, \quad \lambda_1 = \frac{r}{r+m}, \quad \lambda_k = \frac{\mathbb{1}\{k \in S\}}{r+m}, \quad k \neq 1, \quad m = |S|.$$

For fixed m , this is maximized by choosing the m largest alternative coordinates. Hence,

$$\max_{\lambda \in \Lambda_r} \sum_{k=1}^K \lambda_k q_k = \max_{m \in \{0, \dots, K-1\}} \frac{rq_1 + \sum_{j=1}^m q_{-1,(j)}}{r+m},$$

where $q_{-1,(1)} \geq \dots \geq q_{-1,(K-1)}$ is a decreasing rearrangement of $\{q_k : k \neq 1\}$. For $q_k = D_k(a|Y)^2$, substituting into the above display and adding one gives the stated formula for $R_r(a|Y)$. By Proposition 2, $\hat{\theta}_r^*(Y)$ exists uniquely and a least favorable prior λ_r^* can be computed as stated.

For each $S \subseteq [K] \setminus \{1\}$, consider the priors λ^S described above. Because the above equality for $\max_{\lambda \in \Lambda_r} \lambda'q$ holds for every $q \in \mathbb{R}^K$, support function uniqueness yields the polytope

$$\Lambda_r = \text{conv} \left\{ \lambda^S : S \subseteq [K] \setminus \{1\} \right\}.$$

Consequently, since $\lambda \mapsto \lambda'q$ is affine, the maximizing design priors at $a = \hat{\theta}_r^*(Y)$ are the convex hull of the subset of maximizing extreme priors λ^S . Proposition 2 gives a least favorable prior in this convex hull. If the maximizing m^* and S^* are unique, the least favorable prior is λ^{S^*} .

For $r_2 \geq r_1$, $\Lambda_{r_2} \subseteq \Lambda_{r_1}$, so $R_r^*(Y)$ is nonincreasing. To establish its limit, evaluate the criterion at $a = Y_1$. The extreme-prior formula gives

$$R_r^*(Y) - 1 \leq \max_{m \in \{0, \dots, K-1\}} \frac{\sum_{j=1}^m D_{-1,(j)}(Y_1|Y)^2}{r+m} \leq \frac{K-1}{r} \max_{k \neq 1} D_k(Y_1|Y)^2 \rightarrow 0.$$

The dogmatic prior on the favored design belongs to Λ_r for every r , so

$$D_1(\hat{\theta}_r^*(Y)|Y)^2 \leq R_r^*(Y) - 1 \rightarrow 0.$$

Therefore, $\hat{\theta}_r^*(Y) \rightarrow Y_1$.

For $K = 2$ and $\lambda(p) = (p, 1-p)'$, the ratio class generates the interval

$$\mathcal{I}_r = \{p \in [0, 1] : \lambda(p) \in \Lambda_r\} = [p_r, 1].$$

Proposition 2 therefore gives p_r^* as the projection of p_Δ onto $\mathcal{I}_r$ and yields the headline and risk formulas. If $p_\Delta \in \mathcal{I}_r$, then $p_r^* = p_\Delta$ so that $\hat{\theta}_r^*(Y) = \hat{\theta}_\Delta^*(Y)$.

A.6 Proof of Proposition 6

I first show continuity of functions that will also be used in the proof of Proposition 7. For $\mathbb{K}_\Delta$ the set of nonempty compact subsets of $\Delta(\mathcal{V})$, $c \in \mathbb{R}^K$, $s \in (0, \infty)^K$, and $A \in \mathbb{K}_\Delta$, define

$$\mathcal{Q}(a; c, s, A) = \max_{\lambda \in A} \sum_{k=1}^K \lambda_k \frac{(c_k - a)^2}{s_k}, \quad \mathcal{T}(c, s, A) = \operatorname{argmin}_{a \in \mathbb{R}} \mathcal{Q}(a; c, s, A),$$

$$\mathcal{M}(c, s, A) = \min_{a \in \mathbb{R}} \mathcal{Q}(a; c, s, A).$$

For every $\lambda \in A$, the criterion inside the maximum is strongly convex in a , with second derivative

$$2 \sum_{k=1}^K \frac{\lambda_k}{s_k} \geq \frac{2}{\max_{k \in [K]} s_k}.$$

The maximum over $\lambda \in A$ therefore preserves strong convexity and so $\mathcal{T}(c, s, A)$ exists uniquely. Moreover, projecting any a outside of $[\min_k c_k, \max_k c_k]$ onto this interval reduces every squared discrepancy. Hence,

$$\mathcal{T}(c, s, A) \in \left[\min_{k \in [K]} c_k, \max_{k \in [K]} c_k \right].$$

To show continuity of $\mathcal{Q}$, define $q(a; c, s) \in \mathbb{R}^K$ with continuous components

$$(a, c, s) \mapsto q_k(a; c, s) = \frac{(c_k - a)^2}{s_k}.$$

For support function $h_A(q) = \max_{\lambda \in A} \lambda' q$, Hausdorff distance d_H , $A, B \in \mathbb{K}_\Delta$, and $q, \tilde{q} \in \mathbb{R}^K$,

$$|h_A(q) - h_B(\tilde{q})| \leq \max_{\lambda \in A} \|\lambda\|_2 \|q - \tilde{q}\|_2 + \|\tilde{q}\|_2 d_H(A, B) \leq \|q - \tilde{q}\|_2 + \|\tilde{q}\|_2 d_H(A, B).$$

Hence, $(a, c, s, A) \mapsto \mathcal{Q}(a; c, s, A) = h_A(q(a; c, s))$ is continuous.

To show continuity of $\mathcal{T}$ and $\mathcal{M}$, consider any sequence $(c_j, s_j, A_j) \rightarrow (c, s, A)$ and let

$$\mathcal{Q}_j(a) = \mathcal{Q}(a; c_j, s_j, A_j), \quad \mathcal{Q}_0(a) = \mathcal{Q}(a; c, s, A), \quad (c_j, s_j, A_j) \rightarrow (c, s, A), \quad j \rightarrow \infty.$$

Because $c_j \rightarrow c$, there is a compact interval $I \subseteq \mathbb{R}$ containing $[\min_k c_{j,k}, \max_k c_{j,k}]$ for large j and also containing $[\min_k c_k, \max_k c_k]$. Thus, all corresponding minimizers lie in I . Since $s \in (0, \infty)^K$ and $s_j \rightarrow s$, the denominators are bounded away from zero for large j . It follows that

$$\sup_{a \in I} \|q(a; c_j, s_j) - q(a; c, s)\|_2 \rightarrow 0, \quad \sup_{a \in I} \|q(a; c, s)\|_2 < \infty.$$

Combining these bounds with $d_H(A_j, A) \rightarrow 0$ and the support-function inequality gives

$$\sup_{a \in I} |\mathcal{Q}_j(a) - \mathcal{Q}_0(a)| \rightarrow 0.$$

Let $a_j = \mathcal{T}(c_j, s_j, A_j)$ and $a_0 = \mathcal{T}(c, s, A)$. Take an arbitrary subsequence of (c_j, s_j, A_j) , and hence of a_j . Compactness of I gives a further subsequence, again indexed by j , such that $a_j \rightarrow \bar{a} \in I$. By optimality of a_j , $\mathcal{Q}_j(a_j) \leq \mathcal{Q}_j(a) \forall a \in I$. Uniform convergence on I and continuity of $\mathcal{Q}_0$ imply

$$\mathcal{Q}_0(\bar{a}) \leq \mathcal{Q}_0(a), \quad \forall a \in I.$$

The global minimizer of $\mathcal{Q}_0$ is in I , so $\bar{a}$ is the global minimizer. Uniqueness gives $\bar{a} = a_0$. Since every subsequence has a further subsequence converging to a_0 , the full sequence satisfies $a_j \rightarrow a_0$. This proves continuity of $\mathcal{T}$. Finally,

$$|\mathcal{M}(c_j, s_j, A_j) - \mathcal{M}(c, s, A)| \leq |\mathcal{Q}_j(a_j) - \mathcal{Q}_0(a_j)| + |\mathcal{Q}_0(a_j) - \mathcal{Q}_0(a_0)| \rightarrow 0,$$

which proves continuity of $\mathcal{M}$.

Now for the main claim, consider

$$Z_n = \sqrt{n}(Y_n - \theta \mathbf{1}), \quad s_n = (n\tilde{\sigma}_{1,n}^2, \dots, n\tilde{\sigma}_{K,n}^2)', \quad s^0 = (\sigma_1^2, \dots, \sigma_K^2)'.$$

The definitions of the plug-in headline and plug-in risk imply

$$\sqrt{n}(\hat{\theta}_{n,\Lambda}^* - \theta) = \mathcal{T}(Z_n, s_n, \Lambda), \quad R_{n,\Lambda}^* = \mathcal{M}(Z_n, s_n, \Lambda) + 1.$$

Assumption 1 and Slutsky's theorem give

$$(Z_n, s_n) \xrightarrow{d} (Z, s^0), \quad Z \sim N(b, \Sigma).$$

Continuity of $\mathcal{T}$ and $\mathcal{M}$ and the continuous mapping theorem (CMT) therefore yield

$$\sqrt{n}(\hat{\theta}_{n,\Lambda}^* - \theta) \xrightarrow{d} \mathcal{T}(Z, s^0, \Lambda) = \hat{\theta}_\Lambda^*(Z), \quad R_{n,\Lambda}^* \xrightarrow{d} \mathcal{M}(Z, s^0, \Lambda) + 1 = R_\Lambda^*(Z).$$

A.7 Proof of Proposition 7

Use the continuous functions $\mathcal{T}$ and $\mathcal{M}$ defined in the proof of Proposition 6, and let

$$s_n = (n\tilde{\sigma}_{1,n}^2, \dots, n\tilde{\sigma}_{K,n}^2)', \quad s^0 = (\sigma_1^2, \dots, \sigma_K^2)'.$$

Multiplying the plug-in objective by $1/n$ does not change its minimizer. Consequently,

$$\hat{\theta}_{n,\Lambda}^* = \mathcal{T}(Y_n, s_n, \Lambda), \quad \frac{R_{n,\Lambda}^*}{n} = \mathcal{M}(Y_n, s_n, \Lambda) + \frac{1}{n}.$$

Continuity of $\mathcal{T}$ and $\mathcal{M}$, $(Y_n, s_n) \xrightarrow{p} (\mu^0, s^0)$, and CMT therefore imply

$$\hat{\theta}_{n,\Lambda}^* \xrightarrow{p} \mathcal{T}(\mu^0, s^0, \Lambda) = \theta_\Lambda^{\text{lim}}(\mu^0), \quad \frac{R_{n,\Lambda}^*}{n} \xrightarrow{p} \mathcal{M}(\mu^0, s^0, \Lambda) = R_\Lambda^{\text{lim}}(\mu^0).$$

If $R_\Lambda^{\text{lim}}(\mu^0) = 0$, there exists an $a \in \mathbb{R}$ for which

$$\max_{\lambda \in \Lambda} \sum_{k=1}^K \lambda_k \frac{(\mu_k^0 - a)^2}{\sigma_k^2} = 0.$$

Every summand is nonnegative. If every design receives positive probability under at least one prior in Λ , the above equality requires $\mu_k^0 = a$ for every $k \in [K]$. Therefore, if the coordinates of μ^0 are not all equal, then $R_\Lambda^{\text{lim}}(\mu^0) > 0$.