stats

MDPI

Article

Identifying Feasible Sample Sizes from Incomplete Design Information

Satoshi Nakamura

Department of Advanced Diagnostic Imaging, Mie University Graduate School of Medicine, 2-174 Edobashi, Tsu 514-8507, Mie, Japan; s-nakamura@med.mie-u.ac.jp

Abstract

Sample size calculations return the smallest size that meets a requirement once the model and inputs are specified. We ask which sample sizes satisfy the design requirements when the planning information is compatible with several data distributions. We calculate the feasible set for each specification under prespecified error constraints. If every compatible distribution yields the same feasible set, the planning information identifies that set. Otherwise, the intersection and union give sharp inner and outer bounds. A size is feasible when a prespecified rule controls the probabilities of missing a meaningful effect and making a positive decision for a negligible effect. We distinguish the prespecified rule from the performance available in a decision class. The feasible set is therefore a partially identified target. When all proper subset margins are known, the multivariate binary identified set has at most one free coordinate, permitting exact calculation under finite polynomial risk constraints. A normal example shows that a test with null boundary zero can fail at a large sample size while a minimum effect test remains feasible. Public paired vision data illustrate sharp bounds for a test with null boundary zero when the marginal probabilities do not identify the discordance probability. Sample sizes in the inner bound satisfy the constraints for every compatible distribution. For sizes between the bounds, additional planning information is needed to determine feasibility. An empty outer bound shows that the rule or design requirements must change because no candidate size works.

Keywords: comparison of experiments; minimum effect testing; partial identification; sample size planning; scientific relevance

1. Introduction

Sample size planning is usually presented as a calculation that returns the smallest number of observations needed to achieve a target power or precision. The calculation requires a complete planning model, including the effect size and variability. The dependence structure and success criterion must also be specified. Real planning often starts earlier. Investigators may know the marginal event rates or plausible ranges for nuisance quantities without knowing the full joint distribution. Completing the missing information with a working model, prior distribution, or an uncertainty set can produce a recommendation, but a different completion can produce another one. The question addressed in this paper is as follows: when the supplied planning information is consistent with several specifications of the data distribution, which sample sizes can be justified under the declared error requirements without adding an unreported assumption? Our solution is to calculate the feasible set under each specification. Agreement identifies the set. Disagreement yields

Check for updates

Academic Editor: Wei Zhu

Received: 31 August 2026
Revised: 21 September 2026
Accepted: 23 September 2026
Published: 26 September 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

sharp inner and outer bounds, given by the intersection and union, that show which sizes work under every specification or under at least one.

The question also depends on what the planned analysis is meant to claim. A conventional one-sided test controls rejection throughout the nonpositive region. We call this the zero-boundary test, with null hypothesis $H_0 : d \leq 0$ and alternative $d > 0$. As the sample size grows, the same test becomes increasingly likely to reject for a small positive effect that investigators still regard as negligible. A large sample can therefore fail a false-positive requirement defined over a null region of negligible effects, even though the data have become more informative. A minimum effect test, which includes negligible effects in the null hypothesis, may continue to satisfy the requirement. Any conclusion that a size is too small or too large must consequently be tied to the decision rule, the null and alternative regions, and the allowed error probabilities.

Incomplete planning information creates a second issue. Several complete specifications of the future data may be consistent with the available information. Each specification can imply a different set of sample sizes that meet the design requirements. If all specifications in the identified set give the same feasible set, the available information identifies the answer. If they do not, the appropriate result is not an arbitrarily completed calculation. It is a pair of sharp set bounds: an inner bound containing the sizes that work under every specification and an outer bound containing the sizes that work under at least one.

The required components already exist in statistical theory. The Blackwell order compares experiments through the information they provide for decision making [1]. Partial identification describes what can be learned when the available information restricts, but does not determine, the state of nature [2,3]. Minimum effect testing aligns the null hypothesis with a region of negligible effects [4]. For multivariate binary outcomes, compatibility results characterize the joint distributions that share specified lower dimensional margins [5–7]. We combine these components to answer a sample size question that none of them answers alone.

Formally, let ω denote a state of nature, meaning a complete specification of the features needed to calculate the operating probabilities. For a rule fixed before outcomes are observed, let FN_n be the worst case probability of missing an effect in the alternative region of meaningful effects and let FP_n be the worst case probability of making a positive decision in the null region of negligible effects. The feasible set for that rule is

$$A(\omega) = \{n : \mathrm{FN}_n(\omega) \leq \beta,\ \mathrm{FP}_n(\omega) \leq \gamma\}.$$

We distinguish this set from the sizes at which at least one rule in a stated decision class can satisfy both constraints. This distinction prevents the failure of one analysis rule from being interpreted as a failure of the larger experiment. It also makes the monotonicity statement precise: when the larger experiment Blackwell dominates the smaller one and the decision class contains the induced rule, feasibility over the decision class persists at larger sizes.

We give a necessary and sufficient condition for the planning information to identify the full feasible set. When identification fails, we report sharp inner and outer bounds. When all proper subset margins are supplied and no other unknown design component remains, the binary specialization reduces the uncertainty to at most one highest-order product moment coordinate and permits exact finite sample calculations under the stated polynomial conditions. In the normal example, the zero-boundary test has an upper feasible boundary; the minimum effect test remains feasible at larger sizes. A second illustration uses the public vision table distributed with the irr package for the R statistical computing environment. Its marginal rates do not determine the discordance needed for the paired calculation. For two candidate sample sizes, each candidate is feasible in some states in the

identified set, but neither is feasible in all of them. At the discordance probability estimated from the paired counts, the zero-boundary test fails the false-positive constraint at the full sample size. A deterministic sensitivity analysis evaluates this same zero-boundary rule at all candidate sizes under changes to the null and alternative boundaries, error tolerances, and discordance. Additional calculations examine unknown variance, independent pilot estimation, and the computational limits of exact root isolation.

These results lead to specific planning actions. Candidates in the sharp inner bound satisfy the constraints in every compatible state and can be compared on cost or utility. Candidates in the outer bound, but not the inner bound, require the refinement of missing dependence or nuisance inputs to determine whether they satisfy the constraints. If the sharp outer bound is empty, changing the sample size alone cannot meet the requirements under the prespecified rule, so the rule, null and alternative regions, or error limits must be reconsidered. An upper boundary under a zero-boundary test therefore need not justify a recruitment cap or discarding observations; it may instead indicate that the null hypothesis should be aligned with a minimum meaningful effect.

Figure 1 summarizes the framework and its main practical message. Panel A traces the path from the available planning information to state-specific feasible sets, sharp bounds, and a design action. Panel B uses the normal example to show that an upper feasible boundary can arise from the prespecified rule even when another rule remains feasible at larger sample sizes.

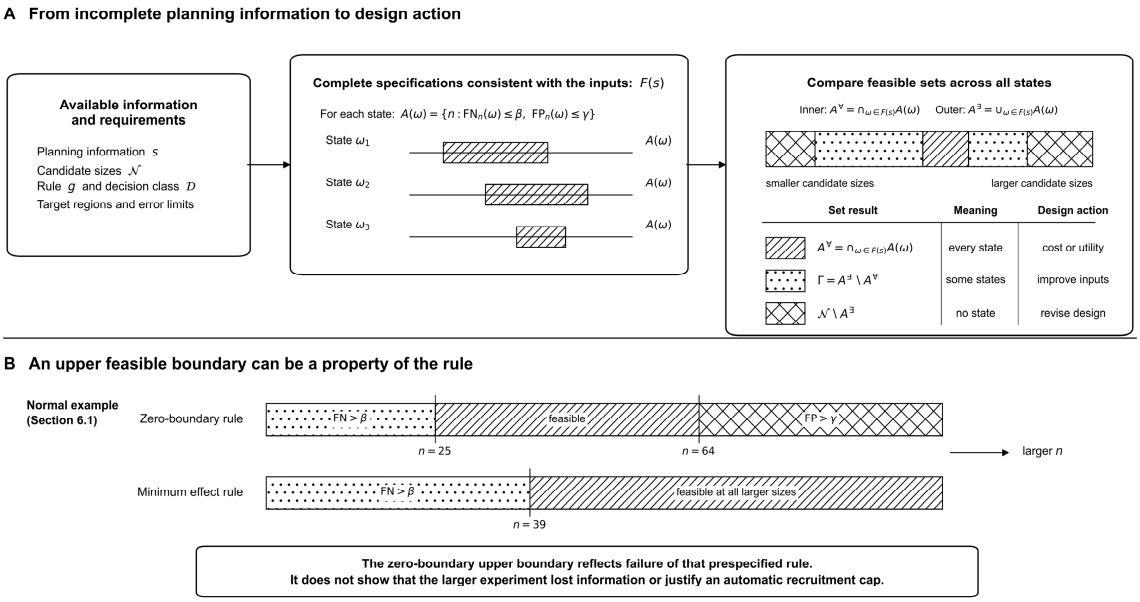

Figure 1. Overview of the framework and the rule comparison. (**A**) Planning information and design requirements define the compatible states F(s). Comparing feasible sets across those states yields sharp inner and outer bounds and directs the next design action. (**B**) In the normal example, the zero-boundary rule is feasible only for n from 25 through 64, whereas the minimum effect rule is feasible for every n of 39 or larger. The upper boundary concerns the prespecified rule; it does not imply that the larger experiment lost information.

The paper proceeds as follows. Section 2 relates the question to existing design methods, partial identification, and dependence bounds. Sections 3 and 4 define the feasible sets and establish what the planning information identifies. Section 5 gives the exact binary calculation. Section 6 presents the numerical illustrations and reporting guidance. Section 7 discusses the contribution, its limitations, and extensions. Section 8 states the main conclusion. Proofs and additional examples are provided in the Supplementary Materials.

2. Relation to Existing Design Methods and Partial Identification

2.1. From Selecting One Size to Identifying the Feasible Set

Conventional calculations select the smallest n that attains power or precision, often at a clinically relevant alternative. Existing methods align statistical success with substantive importance in several ways. A clinical threshold can define the powered alternative or supplement statistical significance [8,9], and discussions of large samples have likewise distinguished statistical from substantive importance [10]. A minimum effect test instead places negligible effects in the null hypothesis [4]. These methods usually evaluate the criterion under a complete design specification. More generally, sample size justification should connect the amount of data to the inferential goal and effects of interest [11]. Reliability acceptance sampling plans also determine sample sizes from specified operating characteristics and risk constraints under a lifetime model [12].

Assurance averages the probability of success over a prior distribution [13]. Robust Bayesian design assesses the sensitivity to classes of priors or other model specifications [14–17]. Decision theory can select a finite optimum by balancing information against cost or utility [18]. Recent work also studies decisions when parameters or payoffs are partially identified [19–21]. Our target lies earlier in the design process: whether the planning information identifies the sample sizes at which stated error constraints are feasible. Averaging or optimizing can select one size without identifying the full feasible set. Methods based on imprecise probabilities also represent uncertainty through sets of probability distributions [22]. Here, we ask whether each candidate sample size satisfies the risk constraints across the states compatible with the planning information.

Arguments for an upper sample size based on ethics or cost compare participant burden with the expected scientific value of additional observations [23,24]. Sample size methods have also been developed for particular incomplete data structures, including partially validated prevalence data [25], while ref. [26] studies Bayesian learning under nonidentified bias. Those calculations address specific completed models. We treat the feasible sample size set itself as the object of partial identification and ask what follows without completing the missing design information.

Formally, established selection procedures typically construct

$$\omega \mapsto n^*(\omega),$$

after completing ω through assumptions, priors, uncertainty classes, or utilities. We ask whether

$$S(\omega) \mapsto A(\omega)$$

exists without completing the unspecified parts of the design. This factorization exists exactly when A is constant over every identified set defined by S. The Blackwell order [1] describes what remains possible as sample size increases, while partial identification [2,3] describes what follows from the planning information. Their combination provides the design framework used here.

2.2. False Positive Probability over Negligible Effects

Let Θ_N denote the null region of clinically negligible effects, let Θ_M denote the alternative region of meaningful effects, and let $R_n(g_n)$ be the event that the planned rule makes a positive decision. Define

$$\mathrm{FN}_n(\omega, g_n) = \sup_{\theta \in \Theta_M} P_{\theta,\omega}\{R_n(g_n)^c\}, \qquad \mathrm{FP}_n(\omega, g_n) = \sup_{\theta \in \Theta_N} P_{\theta,\omega}\{R_n(g_n)\}.$$

These are worst case false-negative and false-positive probabilities over the prespecified alternative and null regions. A level-α test of $H_0 : \theta \leq 0$ controls rejection over nonpositive effects but not throughout a null region $0 < \theta \leq d_0$ that also contains negligible positive effects. For a consistent test, rejection at any fixed positive θ typically approaches one. The formal type I error can therefore remain controlled while the false-positive probability over the declared null region approaches one. A minimum effect test of $H_0 : \theta \leq d_0$ controls rejection throughout Θ_N [4]. Larger samples can then improve discrimination without violating the stated false-positive bound.

Accordingly, we call n too large for the prespecified rule only when it violates the false-positive constraint and a smaller size satisfies both constraints. The statement is relative to $(g, \Theta_N, \Theta_M, \beta, \gamma)$; it is not a statement about economic efficiency or the loss of information. A minimum effect test may satisfy both constraints at the same n. Section 3 distinguishes the failure of the prespecified rule from the failure of every rule in the decision class.

2.3. Established Dependence Bounds and Paired Binary Comparisons

For two binary variables with marginal probabilities p_1 and p_2, the joint probability

$$q = P(Y_1 = 1, Y_2 = 1)$$

is restricted by the Fréchet bounds

$$\max(0, p_1 + p_2 - 1) \leq q \leq \min(p_1, p_2).$$

These bounds are classical [27]. For more than two binary variables, the Bahadur representation and later compatibility results describe the remaining highest-order interaction when all lower dimensional margins are fixed [5–7]. Section 5 records this established representation in the notation needed here. The additional step is to carry its identified interval through a prespecified design criterion and classify candidate sample sizes by the sharp intersection and union sets.

For paired binary indicators, let R and L denote the right and left responses. Define

$$d = P(R = 1, L = 0) - P(R = 0, L = 1)$$

as the difference in marginal probabilities, and define

$$r = P(R \neq L)$$

as the discordance probability. Assume independent and identically distributed (i.i.d.) pairs, allowing dependence within each pair, and require $r > 0$ and $|d| \leq r$. Conditional on m discordant pairs, the number in the $(R, L) = (1, 0)$ cell has a binomial distribution with success probability $(r + d)/(2r)$. At $d = 0$, this gives the classical exact McNemar test [28]. Section 6.2 uses this conditional distribution to evaluate the false-negative and false-positive probabilities of the zero-boundary test.

The marginal probabilities determine d, but they do not determine r. The Fréchet bounds on the common positive cell imply a corresponding interval for the discordance probability. This is the part of the complete design specification that is missing when only the two margins are supplied.

Partial identification and sampling uncertainty are separate questions. The main results condition on the supplied planning information s, just as an ordinary sample size calculation conditions on its planning inputs. When entries of s are estimated, confidence procedures for partially identified parameters address sampling variations in the estimated bounds [29]. Section 4.5 gives a conservative construction using a confidence region and

Section 6.1.2 evaluates it with independent normal pilot data. Coverage of a parameter value and coverage of an entire identified set are distinct targets [30]; our construction starts with a confidence region for the true planning inputs and derives bounds on the feasible sets.

3. Feasible Sample Sizes and the Decision Class

3.1. States of Nature and Planning Information

Let $\mathcal{N} \subseteq \mathbb{N}$ be the candidate domain, the set of sample sizes under consideration, and let Ω be the state space. A state of nature ω is a complete design specification containing the features needed to evaluate operating behavior. These may include the sampling law, nuisance quantities, dependence, and the null and alternative regions. The state need not have finite dimension. For each n, ω induces the experiment

$$\mathcal{E}_n(\omega) = (\mathcal{X}_n, \mathcal{F}_n, \{P^{(n)}_{\theta,\omega} : \theta \in \Theta(\omega)\}),$$

where θ indexes the feature addressed by the decision while ω may contain additional design features. A randomized decision rule is a measurable map

$$\delta_n : \mathcal{X}_n \to [0,1],$$

giving the conditional probability of a positive decision; deterministic rules take values in $\{0,1\}$. The designer observes only

$$S : \Omega \to \mathcal{S}, \qquad S(\omega) = s.$$

The observed value s is the available planning information. The corresponding identified set is

$$F(s) = \{\omega \in \Omega : S(\omega) = s\}.$$

Thus, $F(s)$ contains the states consistent with the supplied planning information. Its members need only agree on that information, not on future data laws. For each n, let $\mathcal{D}_n(s)$ be the decision class, the decision rules available after observing s, and let

$$g_n(s) \in \mathcal{D}_n(s)$$

be the prespecified rule. Write $g = \{g_n\}$ and $\mathcal{D} = \{\mathcal{D}_n\}$. When s, β, γ are fixed, their arguments or subscripts may be suppressed; in particular, g_n and $\mathcal{D}_n$ abbreviate $g_n(s)$ and $\mathcal{D}_n(s)$. Rules in the decision class may depend on the observed planning information but not on the unknown member of $F(s)$.

The general identification results encode the sampling law and dependence in $P^{(n)}_{\theta,\omega}$; they do not impose a common independent sampling model. The normal and vision illustrations use independent observations or independent pairs and exact finite sample distributions. Formal conditions for the mathematical results are stated here and collected with proofs in the Supplementary Materials, Sections S4–S6.

3.2. Error Constraints and Two Feasible Sets

A canonical specification uses a null region of negligible effects $\Theta_N(\omega)$ and a disjoint alternative region of meaningful effects $\Theta_M(\omega)$:

$$\mathrm{FN}_n(\omega, \delta_n) = \sup_{\theta \in \Theta_M(\omega)} E^{(n)}_{\theta,\omega}\{1 - \delta_n(X_n)\},$$

$$\mathrm{FP}_n(\omega, \delta_n) = \sup_{\theta \in \Theta_N(\omega)} E^{(n)}_{\theta,\omega}\{\delta_n(X_n)\}.$$

Thus, FN_n and FP_n are worst case false-negative and false-positive probabilities over the alternative and null regions. The regions need not exhaust $\Theta(\omega)$; an intermediate region may remain undecided. The set results only require prespecified error measures valued in $[0, 1]$, so weighted or posterior predictive versions are also possible if their interpretation is stated.

Fix tolerances $\beta, \gamma \in [0, 1]$. The feasible set for the prespecified rule is

$$A_{\beta,\gamma}(\omega; g) = \{n \in \mathcal{N} : \mathrm{FN}_n(\omega, g_n) \leq \beta,\ \mathrm{FP}_n(\omega, g_n) \leq \gamma\}. \tag{1}$$

Equation (1) defines the feasible set for each state of nature. Its identification is studied below.

The feasible set over the decision class is

$$A_{\beta,\gamma}(\omega; \mathcal{D}) = \{n \in \mathcal{N} : \exists \delta_n \in \mathcal{D}_n \text{ such that } \mathrm{FN}_n(\omega, \delta_n) \leq \beta,\ \mathrm{FP}_n(\omega, \delta_n) \leq \gamma\}.$$

Since $g_n \in \mathcal{D}_n$,

$$A_{\beta,\gamma}(\omega; g) \subseteq A_{\beta,\gamma}(\omega; \mathcal{D}).$$

When the tolerances are fixed, write $A_g(\omega) = A_{\beta,\gamma}(\omega; g)$ and $A_{\mathcal{D}}(\omega) = A_{\beta,\gamma}(\omega; \mathcal{D})$. Thus, A_g records where the prespecified rule works. The difference $A_{\mathcal{D}} \setminus A_g$ contains sizes at which another rule in the decision class works, and exclusion from $A_{\mathcal{D}}$ means that no rule in the decision class works. Either set may be disconnected or empty. Throughout, "feasible" means satisfying the two stated error constraints.

3.3. Comparing Smaller and Larger Sample Sizes

The descriptions "too small" and "too large" are directional conclusions. Each requires a feasible comparator in addition to a failed constraint.

Definition 1 (too small or too large for a prespecified rule). *Fix ω, g, β, and γ. A sample size $n \in \mathcal{N}$ is too small for the prespecified rule if*

$$\mathrm{FN}_n(\omega, g_n) > \beta \quad \text{and} \quad \exists m \in A_{\beta,\gamma}(\omega; g) \text{ with } m > n.$$

It is too large for the prespecified rule if

$$\mathrm{FP}_n(\omega, g_n) > \gamma \quad \text{and} \quad \exists m \in A_{\beta,\gamma}(\omega; g) \text{ with } m < n.$$

No monotonicity is assumed, so a candidate may satisfy both definitions or neither. If A_g is empty, neither conclusion is drawn because no feasible comparator exists.

3.4. Monotonicity Under the Blackwell Order

For $n \leq m$, we say that the experiment at m Blackwell dominates the experiment at n if a known Markov kernel $K_{m \to n}$, independent of the unknown member of $F(s)$, satisfies

$$P^{(m)}_{\theta,\omega} K_{m \to n} = P^{(n)}_{\theta,\omega} \qquad \text{for every } \theta \in \Theta(\omega).$$

For independent observations, the kernel may retain the first n observations; generally, it reproduces the smaller experiment without knowing the state. The following statement specializes the Blackwell order to the decision class [1].

Theorem 1 (monotonicity under Blackwell dominance). *Fix s and $\omega \in F(s)$. Suppose the following for every $n \leq m$ in $\mathcal{N}$:*

1. *Nested experiments: A kernel $K_{m\to n}$ satisfying the preceding identity is available from the design architecture.*
2. *Closure under induced rules: For every $\delta_n \in \mathcal{D}_n(s)$, the lifted rule*

$$\delta_m^{\uparrow}(x_m) = \int_{\mathcal{X}_n} \delta_n(x_n) K_{m\to n}(dx_n \mid x_m)$$

belongs to $\mathcal{D}_m(s)$.

3. *Stable evaluation: The claim, target regions, tolerances, and error definitions do not change between n and m, and reproducing the claim distribution reproduces both error probabilities.*

Then,

$$n \in A_{\beta,\gamma}(\omega; \mathcal{D}) \quad \Rightarrow \quad m \in A_{\beta,\gamma}(\omega; \mathcal{D}).$$

Hence, $A_{\beta,\gamma}(\omega; \mathcal{D})$ is upward closed on $\mathcal{N}$.

Proof: See Supplementary Materials, Section S6.1.

Corollary 1 (alternative rules at larger sample sizes). *Under Theorem 1, if a rule in the decision class works at a smaller n, a rule in that class also works at every larger candidate m. The failure of g_m therefore concerns the prespecified rule rather than what the experiment can achieve over the decision class. Retaining only the information available at n proves existence; an application may instead recalibrate a rule that uses all observations.*

A finite upper boundary for $A_{\mathcal{D}}$ can still arise when enlargement changes the study population, estimand, decision class, evaluation criteria, or cost. In those cases, more than the amount of information has changed. Section S5.1 of the Supplementary Materials gives counterexamples showing why each condition in Theorem 1 is needed.

4. Identification from Limited Planning Information

4.1. Identification of the Feasible Set from Planning Information

The feasible set belongs to a state of nature, whereas a sample size procedure receives only $s = S(\omega)$. Let

$$A : \Omega \to 2^{\mathcal{N}}$$

denote either feasible set, with all design choices fixed. We say that A is identified from S if there is a map

$$C : \mathcal{S} \to 2^{\mathcal{N}}$$

such that

$$C\{S(\omega)\} = A(\omega) \qquad \text{for every } \omega \in \Omega.$$

Equivalently, $A = C \circ S$. The next result is the elementary factorization criterion: a map factors through an information map if and only if it is constant on the fibers of that map. We record it because it defines the identification of the feasible set; no new general factorization theorem is claimed.

Theorem 2 (identification from the planning information). *The following statements are equivalent:*

1. *There is a map C such that $C\{S(\omega)\} = A(\omega)$ for every $\omega \in \Omega$.*

2. *The feasible set is constant on every identified set:*

$$S(\omega_0) = S(\omega_1) \quad \Rightarrow \quad A(\omega_0) = A(\omega_1).$$

3. *For every s in the range of S, the collection of feasible sets induced by $F(s)$,*

$$\mathfrak{A}(s) = \{A(\omega) : \omega \in F(s)\}$$

is a singleton.

When these conditions hold, C is unique on the range of S.

Proof: See Supplementary Materials, Section S6.2.

Numerical values in every input field do not ensure identification. A disagreement between two states in the same identified set at one candidate is enough to show that no procedure using only those fields can recover the full set.

Corollary 2 (membership identification). *For a fixed $n \in \mathcal{N}$, the statement "n is feasible" is exactly identifiable from S if and only if*

$$\mathbf{1}\{n \in A(\omega_0)\} = \mathbf{1}\{n \in A(\omega_1)\}$$

whenever $S(\omega_0) = S(\omega_1)$. The entire set is identified if and only if this condition holds simultaneously for every $n \in \mathcal{N}$.

The identification of a minimum need not identify the set. Let

$$L(\omega) = \inf A(\omega), \qquad \inf \varnothing = +\infty.$$

For two states in one identified set, consider

$$A(\omega_0) = \{30, 50\}, \qquad A(\omega_1) = \{30, 100\}.$$

Both minima equal 30, although membership at 50 and 100 differs. A required minimum can therefore be identified even when the feasibility of larger sizes is not.

Section S5.3 of the Supplementary Materials gives examples covering the boundaries of this criterion. They show that an information map need not be injective to identify A, that a scalar feature may be identified when the full set is not, and that failure can be confined to one identified set. The uniqueness of C is asserted only on the range of S.

Randomization does not restore identification. Theorem S1 gives minimax error $1 - 1/M$ for the recovery of the full set when an identified set contains finite M distinct feasible sets and error one when it contains infinitely many. Corollary S1 gives minimax error $1/2$ for unresolved membership. These results concern the recovery of the design target, not the sampling error probabilities $\mathrm{FN}_n, \mathrm{FP}_n$.

4.2. Incomplete Binary Margins Can Prevent Identification

Theorem 2 does not itself show that the feasible set varies within an identified set in elementary models. The following finite construction shows failure even when every proper subset margin of a multivariate binary outcome is supplied.

Let $K \geq 2$, let $Y = (Y_1, \ldots, Y_K) \in \{0,1\}^K$, and define the parity function

$$\chi(y) = \prod_{j=1}^{K} (2y_j - 1) \in \{-1, +1\}.$$

For $t \in [-1, 1]$, consider the mass function

$$p_t(y) = 2^{-K}\{1 + t\chi(y)\}, \qquad y \in \{0,1\}^K.$$

For $|t| \leq 1$, this is a mass function because half the cells have each parity. Only the highest-order product moment changes with t.

Let a common negligible state θ_N satisfy $P^{(1)}_{\theta_N,\omega_t} = p_t$, with all other design components fixed. Then,

$$E_t\{\chi(Y)\} = t.$$

Theorem 3 (failure of identification with incomplete binary margins). *Let the planning information S contain the marginal distribution of $Y_J = (Y_j : j \in J)$ for every proper subset $J \subsetneq \{1, \ldots, K\}$, but not the full K variate distribution. Fix any candidate sample size $m \geq 1$ and any false-positive tolerance $\rho \in (0, 1)$. The remaining design components can be chosen in common, with the false-negative constraint made nonbinding by taking $\beta = 1$, so that there exist $t_-, t_+ \in [-1, 1]$ satisfying*

$$S(\omega_{t_-}) = S(\omega_{t_+}),$$

while m is feasible under the same prespecified rule in ω_{t_-} and infeasible under that rule in ω_{t_+}, using the constraint $FP_m \leq \rho$. Consequently, no procedure based only on all proper subset margins can identify the feasible set in this problem.

Proof: See Supplementary Materials, Section S6.3.

The construction applies to every $K \geq 2$, candidate m, and $\rho \in (0, 1)$. For $K = 2$, the omitted feature is bivariate association. For $K \geq 3$, identification may still fail after every $(K - 1)$ variate margin is supplied. The decision rule separates the two states for the proof; it is not proposed as a clinical analysis. The result states that proper subset margins cannot guarantee identification, not that every practical rule depends on parity.

The parity perturbation is an established algebraic device [5,6]. Here, it provides two states that agree on the planning information but give different conclusions at the same sample size. Section S5.4 of the Supplementary Materials gives a 2×2 example and examines endpoint choices, randomized rules, rules that do not depend on this product contrast, and the effect of supplying the missing coordinate.

4.3. Sharp Inner and Outer Bounds

When identification fails, partial identification uses intersections and unions to form sharp inner and outer bounds [2,3]. We apply that standard construction to feasible sample sizes. For $F(s) \neq \varnothing$, define

$$A^{\forall}(s) = \bigcap_{\omega \in F(s)} A(\omega), \qquad A^{\exists}(s) = \bigcup_{\omega \in F(s)} A(\omega). \tag{2}$$

In Equation (2), $A^{\forall}(s)$ contains the sizes feasible in every state in the identified set. The set $A^{\exists}(s)$ contains the sizes feasible in at least one such state.

Theorem 4 (sharp inner and outer set bounds). *For every s with a nonempty identified set, the following statements hold:*

1. *Sandwich property. For every $\omega \in F(s)$,*

$$A^{\forall}(s) \subseteq A(\omega) \subseteq A^{\exists}(s).$$

2. *Sharpness. If a set $B(s)$ satisfies*

$$B(s) \subseteq A(\omega) \qquad \text{for every } \omega \in F(s),$$

then $B(s) \subseteq A^{\forall}(s)$. *If a set* $D(s)$ *satisfies*

$$A(\omega) \subseteq D(s) \qquad \text{for every } \omega \in F(s),$$

then $A^{\exists}(s) \subseteq D(s)$. *Thus,* $A^{\forall}$ *is the unique largest universally valid inner set, and* $A^{\exists}$*is the unique smallest universally valid outer set.*

3. *Classification. With*

$$\Gamma(s) = A^{\exists}(s) \setminus A^{\forall}(s),$$

the candidate domain has the disjoint partition

$$\mathcal{N} = A^{\forall}(s) \,\dot\cup\, \Gamma(s) \,\dot\cup\, (\mathcal{N} \setminus A^{\exists}(s)).$$

The first region contains sizes that are feasible in every state in the identified set. The middle region contains sizes whose feasibility is not identified. The final region contains sizes that are feasible in no state in the identified set.

4. *Collapse criterion. The following are equivalent:*

$$A^{\forall}(s) = A^{\exists}(s), \qquad \Gamma(s) = \varnothing, \qquad A(\omega) \text{ is constant over } F(s).$$

Hence, exact identification at s occurs precisely when the inner and outer bounds are equal.

Proof: See Supplementary Materials, Section S6.4.

Enlarging the inner bound or shrinking the outer bound can give a false conclusion for a state in the identified set. These bounds answer questions about one candidate at a time. The fuller identified object $\mathfrak{A}(s) = \{A(\omega) : \omega \in F(s)\}$ is needed only when a decision depends on membership of several candidates together.

For the prespecified rule, define

$$\overline{\mathrm{FN}}_n(s; g) = \sup_{\omega \in F(s)} \mathrm{FN}_n(\omega, g_n), \qquad \overline{\mathrm{FP}}_n(s; g) = \sup_{\omega \in F(s)} \mathrm{FP}_n(\omega, g_n).$$

Then,

$$A^{\forall}_{\beta,\gamma}(s; g) = \{n : \overline{\mathrm{FN}}_n(s; g) \le \beta,\ \overline{\mathrm{FP}}_n(s; g) \le \gamma\}.$$

Possible feasibility under the prespecified rule requires both constraints to hold in one common state in the identified set; separate infima may be attained in different states. Allowing alternative rules introduces a distinct quantifier issue. Write $A^{\forall}_{g}(s)$ for the preceding set and let $A^{\forall}_{\mathcal{D}}(s)$ be the intersection of the pointwise sets $A_{\mathcal{D}}(\omega)$. The latter allows a different successful rule in each state. In contrast, feasibility under one common rule is represented by

$$A^{\mathrm{unif}}_{\beta,\gamma}(s; \mathcal{D}) = \{n : \exists \delta_n \in \mathcal{D}_n(s)\ \forall \omega \in F(s),\ \mathrm{FN}_n \le \beta,\ \mathrm{FP}_n \le \gamma\}.$$

This set requires one rule that can be selected from the planning information. Write it as $A^{\mathrm{unif}}_{\mathcal{D}}(s)$ when the arguments are fixed. Consequently,

$$A^{\forall}_{g} \subseteq A^{\mathrm{unif}}_{\mathcal{D}} \subseteq A^{\forall}_{\mathcal{D}},$$

and the second inclusion can be strict because $\forall\omega\,\exists\delta_n$ does not imply $\exists\delta_n\,\forall\omega$. This order of quantifiers parallels the distinction between static and adjustable robust optimization [31], but the state of nature is not observed before the rule is chosen. Thus, $A^{\forall}_{\mathcal{D}}$ permits a different successful rule for each state, whereas $A^{\text{unif}}_{\mathcal{D}}$ requires one rule to be selected from the planning information. Sections S4 and S5.5 in the Supplementary Materials give interval results and examples with disconnected outer bounds or empty inner bounds.

4.4. Additional Planning Information Cannot Widen the Gap Between Bounds

The inner and outer bounds describe the effect of additional planning information without assigning utility or recruitment cost. Let

$$S_i : \Omega \to \mathcal{S}_i, \qquad i \in \{1, 2\},$$

be two information maps. We say that S_2 refines S_1 if there is a map $r : \mathcal{S}_2 \to \mathcal{S}_1$ such that

$$S_1 = r \circ S_2.$$

Fix $\omega^\star$, set $s_i = S_i(\omega^\star)$, and write

$$F_i = \{\omega : S_i(\omega) = s_i\},$$

$$A_i^{\forall} = \bigcap_{\omega \in F_i} A(\omega), \qquad A_i^{\exists} = \bigcup_{\omega \in F_i} A(\omega), \qquad \Gamma_i = A_i^{\exists} \setminus A_i^{\forall}.$$

The same map A is used for both information maps.

Proposition 1 (monotonicity under information refinement). *If* S_2 refines S_1, then

$$A_1^{\forall} \subseteq A_2^{\forall} \subseteq A(\omega^\star) \subseteq A_2^{\exists} \subseteq A_1^{\exists},$$

and consequently,

$$\Gamma_2 \subseteq \Gamma_1.$$

Define the newly-resolved regions

$$R^{+}_{1\to 2} = A_2^{\forall} \setminus A_1^{\forall}, \qquad R^{-}_{1\to 2} = A_1^{\exists} \setminus A_2^{\exists}.$$

Then,

$$\Gamma_1 \setminus \Gamma_2 = R^{+}_{1\to 2} \,\dot\cup\, R^{-}_{1\to 2}.$$

The first component contains sample sizes newly shown to be feasible; the second contains sample sizes newly shown to be infeasible.

Proof: See Supplementary Materials, Section S6.5.

The first difference $A_2^{\forall} \setminus A_1^{\forall}$ contains candidates newly shown to be feasible in every state in the identified set. The second difference $A_1^{\exists} \setminus A_2^{\exists}$ contains candidates newly shown to be feasible in none. Refinement therefore cannot enlarge the set with unidentified membership, although it may leave the bounds unchanged when the excluded states do not alter membership. The comparison holds for a fixed map A; changing the design requirements defines a different problem.

https://doi.org/10.3390/stats9050105

4.5. Sampling Uncertainty in Planning Inputs

Let an external pilot sample Z, independent of the future study, yield a simultaneous confidence region $C_{1-\delta_p}(Z)$ for the true planning input $s_\star$, with coverage of at least $1-\delta_p$. Keep the rule and design requirements fixed while comparing states, and define

$$\widetilde{F}(Z) = \bigcup_{s' \in C_{1-\delta_p}(Z)} F(s'),$$
$$\widetilde{A}^{\forall}(Z) = \bigcap_{\omega \in \widetilde{F}(Z)} A(\omega),$$
$$\widetilde{A}^{\exists}(Z) = \bigcup_{\omega \in \widetilde{F}(Z)} A(\omega).$$

In the event $s_\star \in C_{1-\delta_p}(Z)$, the true identified set is contained in $\widetilde{F}(Z)$. Hence, simultaneously over the candidate domain,

$$\widetilde{A}^{\forall}(Z) \subseteq A^{\forall}(s_\star) \subseteq A^{\exists}(s_\star) \subseteq \widetilde{A}^{\exists}(Z).$$

These inclusions hold with probability at least the confidence level of the pilot region and give conservative confidence bounds for the true inner and outer feasible sets. The proof uses set inclusion and is given in the Supplementary Materials, Section S7. For a fixed realized confidence region, the intersection and union give sharp bounds relative to the enlarged state set. If the identified set obtained by substituting the pilot estimates is contained in the enlarged state set, its inner bound contains the confidence inner bound and its outer bound is contained in the confidence outer bound.

For paired binary margins, two exact binomial intervals with individual noncoverage of at most $\delta_p/2$ give simultaneous coverage of at least $1-\delta_p$ by Bonferroni, without assuming independence between the two margins. The enlarged set consists of nonnegative joint tables whose margins lie in those intervals, intersected with the restrictions imposed by the design. All nuisance coordinates in that set must be evaluated jointly; generally, the calculation with one free coordinate of Theorem 6 no longer suffices. If the enlarged state set is empty, the confidence region and design restrictions are incompatible, and no sample size is recommended. Section 6.1.2 gives a numerical example where only normal variance is estimated. The tolerances β, γ concern errors in the future study; δ_p sets the noncoverage probability of the pilot confidence region.

5. Binary Identified Sets and Exact Computation

5.1. The Remaining Highest-Order Interaction

Section 4.2 used a symmetric family to show failure of identification. We now characterize the full identified set for arbitrary compatible proper subset margins. Let $K \geq 2$, $Y \in \{0,1\}^K$, and

$$\chi(y) = \prod_{j=1}^{K} (2y_j - 1).$$

Suppose that the planning information s specifies the marginal mass function of $Y_J = (Y_j : j \in J)$ for every proper subset $J \subsetneq \{1,\ldots,K\}$. Let $\mathcal{P}(s)$ be the identified set of joint mass functions compatible with these margins, assume $\mathcal{P}(s) \neq \varnothing$, and choose any baseline $p_0 \in \mathcal{P}(s)$. The following affine representation is established in the theory of multivariate Bernoulli compatibility [5–7]. We state it in the notation used for the design calculation.

Theorem 5 (established affine representation of the binary identified set). *Every $p \in \mathcal{P}(s)$ has a unique representation*

$$p_\lambda(y) = p_0(y) + \lambda\chi(y), \qquad y \in \{0,1\}^K, \tag{3}$$

where

$$\lambda \in [\lambda_L, \lambda_U],$$

$$\lambda_L = \max_{\chi(y)=+1} \{-p_0(y)\}, \qquad \lambda_U = \min_{\chi(y)=-1} p_0(y).$$

Conversely, every $\lambda \in [\lambda_L, \lambda_U]$ produces a member of $\mathcal{P}(s)$. Hence, Equation (3) describes $\mathcal{P}(s)$ as a possibly degenerate closed affine interval.

Define the highest-order product moment:

$$\eta(p) = E_p\{\chi(Y)\} = \sum_{y\in\{0,1\}^K} \chi(y)p(y).$$

Over the identified set,

$$\eta(p_\lambda) = \eta(p_0) + 2^K\lambda.$$

Therefore, η is an affine coordinate on $\mathcal{P}(s)$, with identified interval

$$\eta\{\mathcal{P}(s)\} = \left[\eta(p_0) + 2^K\lambda_L,\ \eta(p_0) + 2^K\lambda_U\right].$$

In particular, all proper subset margins, together with one value of η, identify the full joint distribution.

Proof: See Supplementary Materials, Section S6.6.

Changing the baseline translates λ but not the identified set or the interval for the highest-order product moment. At each endpoint, one cell probability reaches zero. The identified set is a singleton exactly when $\lambda_L = \lambda_U$.

For $K = 2$, the coordinate describes dependence left unspecified by the two univariate margins. For $K = 3$, univariate and pairwise margins leave at most one free affine coordinate corresponding to the highest-order product moment. For any $K \geq 2$, the coordinate may remain unspecified even when every $(K-1)$ variate margin is known, but nonnegativity can reduce its interval to a singleton. Pairwise margins alone generally leave more than one dimension.

Under ± 1 coding, η is the highest-order product moment. It is distinct from the highest-order parameter in a log linear model, although both quantities describe dependence. The bounded coordinate lets us evaluate the two error probabilities over $[\lambda_L, \lambda_U]$. The family in Section 4.2 is the special case $p_0(y) = 2^{-K}$ and $t = 2^K\lambda$.

5.2. Polynomial Operating Probabilities

Theorem 5 gives a bounded scalar identified set. Holding all other state components fixed, let $Y^{(1)}, \ldots, Y^{(n)}$ be independent, with common mass function p_λ.

Let

$$\phi_n : (\{0,1\}^K)^n \to [0,1]$$

be a fixed randomized rule whose value is the conditional claim probability; deterministic rules take only zero and one. Expanding the product below gives polynomiality. Standard results for univariate semialgebraic sets then give the interval classification [32].

Theorem 6 (polynomial consequence for feasibility). *Assume the i.i.d. affine family and fixed rule just specified. For the set classifications below, assume additionally that $\lambda \in [\lambda_L, \lambda_U]$ pa-*

rameterizes the entire identified state set relevant to the risks: all other design components are fixed, and no additional unidentified nuisance parameter remains. The operating probability

$$\pi_n(\lambda) = \sum_{y_1,\ldots,y_n} \phi_n(y_1,\ldots,y_n) \prod_{i=1}^{n} (p_0(y_i) + \lambda\chi(y_i))$$

is a polynomial in λ of degree at most n.

Suppose, in addition, that the two error probabilities can be written on $[\lambda_L, \lambda_U]$ as finite polynomial envelopes,

$$\mathrm{FN}_n(\lambda) = \max_{a\in\mathcal{I}_{\mathrm{FN}}} f_{n,a}(\lambda), \qquad \mathrm{FP}_n(\lambda) = \max_{b\in\mathcal{I}_{\mathrm{FP}}} h_{n,b}(\lambda),$$

where $\mathcal{I}_{\mathrm{FN}}$ and $\mathcal{I}_{\mathrm{FP}}$ are finite and every component is a polynomial. Then, the set of coordinate values that satisfy both constraints,

$$I_n = \{\lambda \in [\lambda_L, \lambda_U] : \mathrm{FN}_n(\lambda) \le \beta,\ \mathrm{FP}_n(\lambda) \le \gamma\}$$

is a closed univariate semialgebraic set and therefore a finite union of closed intervals and isolated points. Moreover,

$$\begin{aligned} n \in A^{\forall}(s) \quad &\Leftrightarrow \quad I_n = [\lambda_L, \lambda_U], \\ n \in A^{\exists}(s) \quad &\Leftrightarrow \quad I_n \neq \varnothing, \\ n \in \Gamma(s) \quad &\Leftrightarrow \quad \varnothing \neq I_n \subsetneq [\lambda_L, \lambda_U]. \end{aligned}$$

Thus, $I_n = \varnothing$ means that the candidate is infeasible in every state in the identified set.

Proof: See Supplementary Materials, Section S6.7.

An exact calculation isolates the real roots of the constraint polynomials on $[\lambda_L, \lambda_U]$. It then evaluates the endpoints and the intervals on which the signs are constant. Sturm sequences allow exact verification for rational or algebraic coefficients. Interval methods may also be used if they resolve every root and sign. A continuous nuisance supremum requires a separate optimization method with an error bound suitable for the stated tolerances.

Corollary 3 (finite grids are insufficient for universal conclusions). *Fix $0 < \gamma < 1$. For every nondegenerate binary identified interval in Theorem 5 and every finite grid, there is a randomized rule based on two independent observations whose false-positive probability is strictly below γ at every grid point but exceeds γ at an unsampled interior point. Proof: See Supplementary Materials, Section S6.8. Grid evaluation is therefore exploratory unless additional shape information is available or the root analysis covers the full interval.*

Polynomiality and real root isolation are established methods [32]. Our computational step applies them to the identified interval supplied by the margins and the error polynomials supplied by the prespecified rule. This gives an exact sample size classification without individual records.

5.3. Computational Scope

Although the operating probabilities have polynomial representations, exact evaluation can require substantial computation. Let $D \le n$ be the actual polynomial degree, b the coefficient bit length after rational representation, and q the number of constraint polynomials. The cost of isolating roots depends on D, b, q and the distances between the roots; sample size n alone does not determine the computational cost. In an unrestricted binary model, direct enumeration of ordered samples has 2^{Kn} terms; a permutation-invariant rule can instead aggregate over $\binom{n+2^K-1}{2^K-1}$ count vectors. Polynomial construction can therefore

dominate the subsequent univariate calculation. Users must supply the risk polynomials to the package because deriving them depends on the decision rule.

The implementation uses Python arbitrary-precision rational arithmetic. This avoids floating-point underflow and overflow in its symbolic steps, but coefficient growth can exhaust time or memory. The division of dense polynomials takes at most $O(D^2)$ rational operations, and the naive Euclidean construction of a Sturm sequence has a conservative $O(D^3)$ bound on the number of arithmetic operations. Evaluating the number of sign changes requires at most $O(D^2)$ rational operations. The cost of each rational operation also grows with coefficient size. In addition, the implementation rebuilds Sturm sequences as it subdivides intervals, so these operation counts do not bound the total running time. Section S9 of the Supplementary Materials reports running times as the polynomial degree, coefficient size and distances between roots vary. The benchmarks also examine repeated constraints under fixed limits on time and memory.

The vision calculation uses binomial probabilities directly. For a fixed state, Equation (4) evaluates $n + 1$ binomial terms. Scanning all candidates through N evaluates $O(N^2)$ such terms and requires $O(N)$ working storage, excluding the internal cost of each binomial function evaluation. For each of the two candidate sizes in Section 6.2, specific values of the discordance probability establish feasibility in one state and infeasibility in another. These examples determine the inner and outer bounds for those two sizes. The point at which an alternative method becomes necessary depends on the polynomial and the available computational resources. If root isolation cannot finish within the specified limits, feasibility remains unresolved by that calculation. An interval method or an approximation with a uniform error bound can establish whether a risk constraint holds when the entire resulting risk interval lies on one side of its tolerance. Without such error bounds, normal or Poisson approximations provide only exploratory results. Failure to complete a calculation does not establish statistical nonidentification and therefore cannot assign a candidate to $\Gamma(s)$.

5.4. Paired Categorical Outcomes

For a paired outcome with C categories, write $p = (p_{ij})$, let H encode the known row and column sums, and choose a compatible table p_0. If the columns of B span the null space of H, then

$$\mathcal{P}(s) = \{p_0 + B\lambda : p_0 + B\lambda \geq 0\}.$$

With positive margins and no additional restrictions, this polytope has dimension $(C-1)^2$: four for a 3×3 table and nine for the original 4×4 vision table. Zero margins or restrictions imposing structural zeros can reduce the dimension. More generally, fixing every proper subset margin of variables with category counts $C_1, \ldots, C_K$ leaves an affine null space of dimension $\prod_j (C_j - 1)$, before nonnegativity is imposed. The binary representation with one free coordinate is a special case. Section S10 of the Supplementary Materials gives the parameterization and proof.

For i.i.d. observations and a fixed rule, operating probabilities are multivariate polynomials of a total degree of at most n. The inner and outer bounds retain their intersection and union definitions. Computing them requires multivariate optimization or methods from real algebraic geometry, which can be substantially more costly than univariate root isolation. Applying the Stuart–Maxwell test of marginal homogeneity also requires negligible and meaningful regions for the vector of marginal differences, for example using a prespecified norm. Its usual chi-square calibration provides an asymptotic approximation. Ordinal scores likewise specify a new estimand. A Dirichlet–multinomial analysis adds a prior distribution for the joint cell probabilities. A further extension could assess sensitivity to a class of such priors.

6. Exact Numerical Illustrations

The two illustrations evaluate finite sample operating probabilities directly. The normal example uses closed-form normal probabilities, and the vision example sums exact binomial probabilities over every discordance count for each candidate size. The nominal levels, null and alternative regions, and candidate domains are stated before the numerical conclusions.

6.1. Analytic Normal Example

A normal model with one parameter illustrates how the two tests yield different feasible sets. Let $X_1, \ldots, X_n$ be independent, with $X_i \sim N(d, 1)$ for $i = 1, \ldots, n$, where d is standardized and the variance is known. Treat $d \leq d_0 = 0.10$ as negligible and $d \geq d_1 = 0.50$ as meaningful, and treat (d_0, d_1) as an indifference region on which neither error guarantee is imposed. Set $(\alpha, \beta, \gamma) = (0.05, 0.20, 0.20)$.

First, consider the conventional zero-boundary test with decision rule

$$g_n^{(0)} = \mathbf{1}\{\sqrt{n}\,\bar{X}_n > z_{1-\alpha}\},$$

where one denotes a positive decision and $z_q = \Phi^{-1}(q)$. Monotonicity places the least favorable effects at the alternative and null boundaries d_1 and d_0, so

$$\begin{aligned} \mathrm{FN}_n^{(0)} &= \sup_{d \geq d_1} P_d\{g_n^{(0)} = 0\} = \Phi\left(z_{1-\alpha} - d_1\sqrt{n}\right), \\ \mathrm{FP}_n^{(0)} &= \sup_{d \leq d_0} P_d\{g_n^{(0)} = 1\} = \Phi\left(d_0\sqrt{n} - z_{1-\alpha}\right). \end{aligned}$$

The two constraints yield integer lower and upper boundaries,

$$\begin{aligned} L_\beta^{(0)} &= \left\lceil \left(\tfrac{z_{1-\alpha} + z_{1-\beta}}{d_1}\right)^2 \right\rceil = 25, \\ U_\gamma^{(0)} &= \left\lfloor \left(\tfrac{z_{1-\alpha} + z_\gamma}{d_0}\right)^2 \right\rfloor = 64. \end{aligned}$$

Consequently, the feasible set for this fixed rule is

$$A_{.20,.20}(g^{(0)}) = \{25, 26, \ldots, 64\}.$$

At $n = 30$, power at the meaningful boundary is 0.863, giving a false-negative probability of 0.137, while false-positive probability at the negligible boundary is 0.136. At $n = 1000$, power is effectively one but false-positive probability is

$$\mathrm{FP}_{1000}^{(0)} = \Phi(0.10\sqrt{1000} - z_{0.95}) \approx 0.935.$$

Thus, $n = 30$ is feasible, whereas $n = 1000$ violates only the false-positive constraint and is too large for the prespecified rule by Definition 1. The boundary occurs between 64 and 65 because $\mathrm{FP}_{64}^{(0)} \approx 0.1991$ and $\mathrm{FP}_{65}^{(0)} \approx 0.2008$.

Using the established minimum effect formulation makes the same 1000 observations feasible [4]. Denote this rule by $g_n^{(\mathrm{ME})}$, where ME means minimum effect:

$$g_n^{(\mathrm{ME})} = \mathbf{1}\{\sqrt{n}(\bar{X}_n - d_0) > z_{1-\alpha}\},$$

which tests the minimum effect null $H_0 : d \leq d_0$. Its error probabilities are

$$\mathrm{FP}_n^{(\mathrm{ME})} = \sup_{d \leq d_0} P_d\{g_n^{(\mathrm{ME})} = 1\} = \alpha = 0.05,$$
$$\mathrm{FN}_n^{(\mathrm{ME})} = \sup_{d \geq d_1} P_d\{g_n^{(\mathrm{ME})} = 0\} = \Phi\left(z_{1-\alpha} - (d_1 - d_0)\sqrt{n}\right).$$

Hence, under the same $(\beta, \gamma) = (0.20, 0.20)$,

$$A_{.20,.20}(g^{(\mathrm{ME})}) = \{39, 40, \ldots\}.$$

At $n = 30$, power for the minimum effect test at $d = 0.50$ is 0.707, first exceeding 0.80 at $n = 39$. At $n = 1000$, the false-negative probability is negligible and the false-positive probability remains at 0.05.

Table 1 shows the pattern: the zero-boundary test has a bounded feasible interval because its false-positive probability over the null region increases with sample size, whereas the minimum effect test has a lower boundary and remains feasible thereafter. If both tests are in the decision class, all observations can be retained by changing the test used for the decision. Statistical significance against a zero null boundary does not by itself establish a nonnegligible effect.

Table 1. Operating characteristics in the normal example under two prespecified rules.

Prespecified Rule	Feasible n	Power at d=0.50, n=30	False-Positive Probability at d=0.10, n=30	False-Positive Probability at d=0.10, n=1000
Zero-boundary test, $H_0 : d \leq 0$	25–64	0.863	0.136	0.935
Minimum effect test, $H_0 : d \leq 0.10$	$39, 40, \ldots$	0.707	0.050	0.050

The tolerances can also reveal an infeasible specification. Tightening the false-positive tolerance for the zero-boundary test to $\gamma = 0.10$ makes the upper boundary 13, while the power requirement still gives a lower boundary of 25. The feasible set is therefore empty. Definition 1 then supports neither directional conclusion because no feasible comparator exists.

6.1.1. Unknown Variance and a Variance Range

Now let $X_i \sim N(\mu, \sigma^2)$ independently, with unknown variance. To distinguish raw and standardized effects, hold the raw boundaries $a_0 = 0.10$ and $a_1 = 0.50$ fixed as σ varies. Write S_n for the sample standard deviation, $\nu = n - 1$, and $c_n = t_{\nu,1-\alpha}$. The zero-boundary and minimum effect statistics are $\sqrt{n}\bar{X}/S_n$ and $\sqrt{n}(\bar{X} - a_0)/S_n$, respectively. If $F_{\nu,\eta}$ is the noncentral t distribution function with noncentrality η, their error probabilities at the null and alternative boundaries are

$$\mathrm{FN}_n^{(0)}(\sigma) = F_{\nu,a_1\sqrt{n}/\sigma}(c_n),$$
$$\mathrm{FP}_n^{(0)}(\sigma) = 1 - F_{\nu,a_0\sqrt{n}/\sigma}(c_n),$$
$$\mathrm{FN}_n^{(\mathrm{ME})}(\sigma) = F_{\nu,(a_1-a_0)\sqrt{n}/\sigma}(c_n),$$
$$\mathrm{FP}_n^{(\mathrm{ME})}(\sigma) = \alpha.$$

The t statistics use the sample standard deviation, whereas their operating probabilities depend on the true σ. For $\sigma \in [\sigma_L, \sigma_U]$, the inner set is obtained by intersecting the feasible sets for each value of the standard deviation and the outer set by their union. The worst

false-negative probability occurs at σ_U; for the zero-boundary rule with $a_0 > 0$, the worst false-positive probability occurs at σ_L. Membership in the outer bound requires both constraints to hold at the same σ. Section S8 of the Supplementary Materials gives a method that solves monotone equations in one variable to check this condition over the full variance interval.

With the tolerances specified in Section 6.1 and candidate sample sizes $\{2, \ldots, 1000\}$, $\sigma = 1$ gives feasible sizes of 27–65 for the zero-boundary t rule and 41–1000 for the minimum effect t rule. For $\sigma \in [0.8, 1.2]$, the zero-boundary inner and outer sets are 38–42 and 18–94, whereas the minimum effect sets are 58–1000 and 27–1000. Widening the range to $[0.6, 1.4]$ makes the zero-boundary inner set empty, while the minimum effect inner set is 78–1000. The t calculations account for estimating the variance in the future study. The general set theorems and the Blackwell result continue to apply under their stated conditions. Theorem 6 is specific to the binary polynomial model. If σ is unbounded above, no finite sample size guarantees 80% power for every possible variance when the effect boundaries remain fixed in the original measurement units.

6.1.2. Uncertainty from an Independent Pilot

For each pilot size of 20, 40, 60, 100, and 200, we generated 10,000 sample variances using their exact sampling distribution under a normal model with true $\sigma = 1$. Equal-tailed 95% chi-square intervals for σ were used to calculate inner and outer feasible sets for the t tests over all 999 candidate sample sizes. The raw effect boundaries, significance level of the future test, and false-negative tolerance were fixed as above; false-positive tolerances 0.20 and 0.15 were examined. We evaluated the error probabilities in the future study using the noncentral t distribution.

At $\gamma = 0.20$, the zero-boundary inner set was empty in all replications with a 20-observation pilot. With a pilot of 40 observations, the inner set was nonempty in every replication but contained only 2.5 candidates on average, compared with 39 candidates at the true variance. At $\gamma = 0.15$, the zero-boundary inner set was empty in every replication for pilot sizes through 100, and in 73.54% at size 200 (Monte Carlo standard error 0.44 percentage points). The minimum effect inner set was nonempty in every replication of every pilot scenario among the candidate sample sizes considered. A wider confidence region can therefore leave no sample size feasible for every variance in that region, even when some sample sizes are feasible at the true variance.

Selecting the smallest plug-in feasible size frequently violated at least one exact risk constraint at the true variance. For the 40-observation pilot and $\gamma = 0.20$, this occurred in 51.92% of replications for the zero-boundary rule, compared with 3.51% when selecting from the confidence inner set. These percentages measure how often the selected sample size violates a design constraint at the true variance. The amount of the violation can be small. An empty inner set generated no recommendation and was recorded separately. Section S8 of the Supplementary Materials reports coverage of the feasible set bounds and the frequency of selecting each sample size for both rules in all scenarios, together with Monte Carlo standard errors. These results are conditional on normal sampling and an independent pilot; reusing pilot observations in the final test requires evaluating the joint adaptive protocol.

6.2. Public Vision Data

We use the published vision table, which records the unaided distance vision of the right and left eyes of 7477 female employees at Royal Ordnance factories from 1943 to 1946 [33]. The table is distributed with irr 0.84.1 [34], and the analysis used R 4.4.2 [35]. Each eye was assigned one of four grades. We reduce the grades to first grade versus

another grade to illustrate the established paired binary calculation in Section 2.3. The resulting estimand is the difference between the probabilities that each eye receives the first grade. Dichotomization discards information about the remaining grades. Section 5.4 describes the multivariate extension for the original categories. The analysis is descriptive and does not assess eye health in a current population. Table 2 presents the resulting paired binary counts.

Table 2. Paired first grade classifications in the vision data.

Right Eye	Left Eye: First Grade	Left Eye: Other Grade
First grade	1520	456
Other grade	387	5114

The empirical first grade proportions are $1976/7477 \approx 0.2643$ for the right eye and $1907/7477 \approx 0.2550$ for the left eye. Their difference is

$$\hat{d} = (456 - 387)/7477 \approx 0.00923.$$

There are 843 discordant pairs, giving $\hat{r} = 843/7477 \approx 0.11275$. All calculations for the empirical state use the unrounded fraction $843/7477$. The exact McNemar P-value for a difference in the positive direction is 0.0096. Thus, the zero-boundary null is rejected even though the observed difference is less than one percentage point.

To connect this result with sample size design, suppose, for illustration, that differences $0 < d \leq d_0 = 0.01$ are negligible and differences $d \geq d_1 = 0.05$ are meaningful. These values explain the method and are not proposed as ophthalmic standards. We set $(\alpha, \beta, \gamma) = (0.05, 0.20, 0.20)$ and first evaluate the fixed zero-boundary rule at the empirical discordance probability r = 843/7477. We then evaluate the same rule across the discordance interval compatible with the marginal planning information.

For a candidate sample of n i.i.d. pairs, let M be the number of discordant pairs and let B be the number with a first grade right eye and another grade left eye. Then,

$$M \sim \text{Binomial}(n, r), \qquad B \mid M = m \sim \text{Binomial}\{m, \pi(d)\}, \qquad \pi(d) = \frac{r + d}{2r}.$$

For the zero-boundary test, let $c_m^{(0)}$ be the smallest count in {0, . . ., m + 1} whose upper tail under a Binomial(m, 1/2) distribution is at most α. A critical count of m + 1 yields no rejection because B cannot exceed m. The rule rejects when the observed count B reaches or exceeds $c_m^{(0)}$, using the observed number M = m of discordant pairs. These critical counts do not depend on r, so the rule can be implemented when r is unknown. The distribution of (M, B), and hence, the rejection probability, still depends on r. Restrict the effect regions to $[-r, r]$; the model can include d_1 only if $r \geq d_1 > 0$. The case $r = 0$ permits only $d = 0$ and is excluded from this design. The exact unconditional rejection probability of this fixed rule when the discordance probability equals r is

$$\rho_n^{(0)}(d; r) = \sum_{m=0}^{n} \binom{n}{m} r^m (1 - r)^{n-m} P_{\pi(d)}(B \geq c_m^{(0)} \mid M = m). \tag{4}$$

For this fixed rule, the false-negative probability is $1 - \rho_n^{(0)}(d_1; r)$, and the false-positive probability is $\rho_n^{(0)}(d_0; r)$. Evaluating Equation (4) at every integer in the candidate domain $\mathcal{N} = \{1, \ldots, 7477\}$ gives the results in Table 3.

Table 3. Observed conclusion and exact design calculations for the zero-boundary test at the empirical discordance.

Prespecified Rule	Observed Tail Calculation	Feasible Sample Sizes	Result at $n = 7477$
Zero-boundary test, $H_0 : d \leq 0$	Exact McNemar $P = 0.0096$	300 to 920	False positive probability 0.816; infeasible

For the zero-boundary test, $n = 500$ supplies the smaller feasible comparator required by Definition 1. The full sample fails because the false-positive probability over the null region is too high, despite essentially having complete power at the alternative boundary.

The paired counts in Table 2 also permit a direct calculation of the sharp bounds. The two empirical marginal rates determine d but not the discordance probability r. If only those margins are retained, the Fréchet bounds permit $69/7477 \leq r \leq 3883/7477$. The working model with common discordance must also support the alternative boundary $d_1 = 0.05$, which restricts the identified interval used for design to

$$0.05 \leq r \leq \frac{3883}{7477}.$$

Consider the zero-boundary test, whose conditional critical values do not depend on the unknown r, at the two candidate sample sizes $\mathcal{N}_{\text{sens}} = \{500, 1321\}$. The first candidate is the comparator already used above. The second is the smallest candidate that attains 0.80 power at the upper discordance bound. Table 4 gives the states in the identified set that establish both membership and nonmembership for each candidate.

Table 4. States establishing the sharp bounds in the vision illustration.

n	State Establishing Feasibility	Power; False Positive Probability	State Establishing Infeasibility	Power; False Positive Probability
500	$r = 843/7477$	0.9504; 0.1342	$r = 3883/7477$	0.4385; 0.0812
1321	$r = 3883/7477$	0.8000; 0.1192	$r = 843/7477$	0.9999; 0.2595

It follows directly that

$$A_g^{\forall}(s) = \varnothing, \qquad A_g^{\exists}(s) = \{500, 1321\}, \qquad \Gamma_g(s) = \{500, 1321\}$$

on $\mathcal{N}_{\text{sens}}$. Thus, feasibility of neither candidate is identified from the marginal rates alone. The full paired table supplies $r = 843/7477$, which resolves the calculation for that empirical planning input. The sharp bound calculation uses the zero-boundary test because its critical values remain fixed over the identified interval. The conditional minimum effect test in Section S5.6.1 of the Supplementary Materials uses critical counts that depend on r. Such a rule cannot be implemented from the marginal planning information alone. Section S5.6.1 of the Supplementary Materials reports calculations assuming that r is known. Those calculations do not establish feasibility of a minimum effect rule when only the marginal probabilities are known.

Table 4 establishes the sharp bounds for two candidate sizes. We next examined sensitivity to the planning inputs by evaluating every $n \in \{1, \ldots, 7477\}$ under the seven scenarios in Table 5. The baseline is $(d_0, d_1, \beta, \gamma, r) = (0.010, 0.050, 0.20, 0.20, 843/7477)$. Each subsequent row changes one input or uses an endpoint of the discordance interval permitted by the design. The lower endpoint is $\max(69/7477, d_1) = 0.05$, after imposing the restriction required by the alternative boundary; the upper endpoint is exactly $3883/7477 \approx 0.519326$. In all seven scenarios, the zero-boundary critical counts remain the same. The endpoints of the feasible sample size intervals are obtained from the exact finite sample sum in Equation (4).

Table 5. Sensitivity of the feasible sample size intervals for the fixed zero-boundary test.

Setting	Input Changed from Baseline	Zero-Boundary Feasible Sizes
Baseline	None	300 to 920
Meaningful boundary reduced	$d_1 = 0.040$	469 to 920
Negligible boundary increased	$d_0 = 0.015$	300 to 441
False-negative tolerance reduced	$\beta = 0.10$	403 to 920
False-positive tolerance reduced	$\gamma = 0.10$	300 to 301
Lower discordance bound permitted by the design	$r = 0.05000$	134 to 491
Upper Fréchet bound for discordance	$r = 3883/7477$	1321 to 3546

The zero-boundary test has a finite upper boundary in every scenario, although its location depends on the null and alternative regions, tolerances, and discordance. These calculations show how the feasible interval of the fixed zero-boundary rule depends on the illustrative planning inputs. Clinical use would require thresholds justified for the intended ophthalmic application.

The baseline and four scenarios that change an effect boundary or error tolerance hold the empirical discordance fixed. The final two scenarios use the discordance endpoints permitted by the design. Sampling uncertainty in these historical inputs is not included in Tables 2–5. Section 4.5 describes how to propagate simultaneous confidence regions for marginal rates, and Section 6.1.2 evaluates pilot uncertainty in the normal model.

6.3. Reporting the Design Conclusion

A conventional calculation returns one recommended number but does not show whether the conclusion holds for every state in the identified set. It also does not show whether failure at that size concerns only the prespecified rule. A complete report should therefore state the planning information and its identified set, then give the null and alternative regions, error limits, prespecified rule, decision class, and candidate domain. It should report the sharp inner and outer bounds with the calculations that support them. A directional conclusion must include the violated constraint and a feasible comparator. If another rule is said to work, that rule must be reported. The formulas, numerical tolerances, and software versions needed for reproduction should also be included.

The sharp bounds give three conclusions for the prespecified rule. A candidate may be feasible in every state in the identified set, have feasibility that is not identified, or be feasible in none. Direction is a separate statement. “Too large” requires violation of the false-positive constraint and a smaller feasible comparator, while “too small” requires violation of the false-negative constraint and a larger comparator.

For the singleton identified set in Section 6.1, the report at $n = 1000$ gives $A_g = \{25, \ldots, 64\}$, $\mathrm{FP}^{(0)}_{1000} \approx 0.935 > 0.20$, and the smaller comparator $m = 30$. If the decision class contains the minimum effect test, the report also gives $\mathrm{FP}^{(\mathrm{ME})}_{1000} = 0.05$ and a negligible false-negative probability. Thus, 1000 observations are too large for the zero-boundary test, but another rule in the decision class meets the requirements without discarding data. Section S5.2 of the Supplementary Materials gives the complete report produced by the reference software.

6.4. Five Cases Induced by the Nested Feasible Sets

The following nested sets distinguish five conclusions about the prespecified rule and the decision class:

$$A^{\forall}_{g} \subseteq A^{\mathrm{unif}}_{\mathcal{D}} \subseteq A^{\forall}_{\mathcal{D}} \subseteq A^{\exists}_{\mathcal{D}}.$$

Define

$$T_1 = A^{\forall}_{g}, \qquad T_2 = A^{\mathrm{unif}}_{\mathcal{D}} \setminus A^{\forall}_{g},$$
$$T_3 = A^{\forall}_{\mathcal{D}} \setminus A^{\mathrm{unif}}_{\mathcal{D}}, \qquad T_4 = A^{\exists}_{\mathcal{D}} \setminus A^{\forall}_{\mathcal{D}},$$
$$T_5 = \mathcal{N} \setminus A^{\exists}_{\mathcal{D}}.$$

Proposition 2 (set partition used for reporting). *These successive set differences partition $\mathcal{N}$. In T_1, the prespecified rule works in every state in the identified set. In T_2, the prespecified rule fails in at least one compatible state, while one alternative rule works in every state. In T_3, a rule works in each state, but the successful rule may differ between states. In T_4, a rule in the decision class works in some states but not in others. In T_5, no rule in the decision class works in any state in the identified set. This set partition follows directly from the nesting. Proof: See Supplementary Materials, Section S6.9.*

Corollary 4 (stopping conditions after classification). *Algorithm 1 has the following consequences.*

1. *For $n \in T_1$, the feasibility of the prespecified rule is established throughout the identified set, so rule search may stop for that candidate.*
2. *For $n \in T_2$, one reported alternative rule is feasible throughout the identified set, so rule search may stop after the prespecified rule is replaced by that witness.*
3. *For $n \in T_3 \cup T_4$, the planning information does not support one implementable feasibility conclusion. Further selection from these candidates requires refinement of the planning information or revision of the decision class.*
4. *A failed candidate receives no directional conclusion unless the violated constraint and comparator required by Definition 1 have both been established.*
5. *If $T_5 = \mathcal{N}$, no rule in the decision class works at any candidate in any state in the identified set. Changing sample size within the declared candidate domain cannot satisfy the design specification.*

The proof is in the Supplementary Materials, Section S6.10. An external criterion may rank the retained combinations of candidates and rules after these conditions have been applied.

The reference implementation ssadmiss 0.5.0rc1 applies the calculation to a finite candidate domain and a finite decision class over the binary-identified interval. It uses rational arithmetic and Sturm sequences. Section S5.2 of the Supplementary Materials gives a worked example covering all five cases.

Algorithm 1. Classification over a finite candidate domain and decision class

Supply the null and alternative regions, error limits, nonempty identified set $F(s)$, candidate domain $\mathcal{N}$, prespecified rule g, and finite decision class $\mathcal{D}$ containing g. Then, proceed as follows.

1. For every $n \in \mathcal{N}$ and $\delta \in \mathcal{D}_n(s)$, calculate the set of states for which the rule is feasible:

$$I_{n,\delta} = \{\omega \in F(s) : \mathrm{FN}_n(\omega, \delta) \leq \beta,\ \mathrm{FP}_n(\omega, \delta) \leq \gamma\}.$$

   In the binary specialization, isolate all real roots of the constraint polynomials on the identified interval and evaluate their common sign cells.
2. Form

$$A_g^{\forall} = \{n : I_{n,g_n} = F(s)\},$$
$$A_{\mathcal{D}}^{\mathrm{unif}} = \{n : \exists \delta \in \mathcal{D}_n(s),\ I_{n,\delta} = F(s)\},$$
$$A_{\mathcal{D}}^{\forall} = \{n : \bigcup_{\delta \in \mathcal{D}_n(s)} I_{n,\delta} = F(s)\},$$
$$A_{\mathcal{D}}^{\exists} = \{n : \bigcup_{\delta \in \mathcal{D}_n(s)} I_{n,\delta} \neq \varnothing\}.$$

3. Apply Proposition 2 to assign each candidate to exactly one of $T_1, \ldots, T_5$. Retain the successful common rule as a witness for every T_2 assignment. For T_3 and T_4, retain states showing the relevant change in the rule or feasibility.
4. Draw a directional conclusion only when Definition 1 supplies both a violated constraint and a feasible comparator under the same prespecified rule and state.
5. If at least one implementable combination of a candidate and a rule remains, apply any cost or utility criterion only after the preceding classification. If $T_5 = \mathcal{N}$, revise the declared design components rather than selecting a sample size within the failed specification.

Software Use

The reference package uses Python 3.10 or later and its standard library for the rational polynomial calculations. After unpacking the supplied archive, enter the package directory and run the following command:
python3 -m pip install .
The vision script separately uses R and the irr package. The additional numerical analyses use NumPy, SciPy, Matplotlib, and psutil, with versions and commands recorded in the additional analysis archive.

The following example classifies two designs specified by the user. To keep the example simple, the false-negative probability is set to zero, so only the false-positive constraint affects feasibility. For an actual study, users must supply the error polynomials implied by the study model and decision rule.

```
from fractions import Fraction as F
from ssadmiss import (RationalPolynomial as P,
PolynomialRiskConstraint as Risk,
binary_compatibility_interval, uniform_binary_mass,
classify_finite_rule_family)

x = P.variable()
def risks(fp):
return (Risk(‘‘FN’’, ‘‘false_negative’’, P.zero(), F(1, 2)),
Risk(‘‘FP’’, ‘‘false_positive’’, fp, F(1, 2)))
designs = {1:  risks(P.constant(F(1, 4))),
2:  risks(F(1, 2) + x)}
answer = classify_finite_rule_family(
compatibility=binary_compatibility_interval(uniform_binary_mass(2)),
candidate_sizes=(1, 2), planned_rule=‘‘g’’,
rule_family={‘‘g’’:  designs})
print(answer.as_dict())
```

Candidate 1 is feasible throughout the identified interval; candidate 2 is feasible only on its nonpositive half. The output includes the feasible intervals and the roots used to determine their endpoints. classify_finite_rule_family evaluates only the finite collection of decision rules supplied by the user.

7. Discussion

This paper treats the feasible sample size set as the object to be identified when planning information is incomplete. The main results separate the feasibility of a prespecified rule from the feasibility over a decision class, give the exact condition under which the planning information identifies the full set, and provide sharp inner and outer bounds when it does not. The binary specialization supplies an exact calculation when proper subset margins leave one highest-order interaction coordinate. For the two candidate sizes in the vision example, incomplete marginal information gives an empty inner bound and a nonempty outer bound. The sensitivity analysis evaluates the same zero-boundary rule under different planning inputs and shows that its upper feasible boundary changes with those inputs. The normal example also shows that a minimum effect test can remain feasible at larger sizes. This result illustrates the conclusion from the comparison of experiments: an upper feasible boundary for a particular rule does not imply that additional observations reduce information.

The framework combines the established results. The Blackwell order provides the monotonicity statement, partial identification provides the intersection and union bounds, minimum effect testing provides the alternative decision rule in the normal example, and Bernoulli compatibility results provide the binary parameterization. Together they determine which sample sizes are feasible in every state in the identified set, feasible in some states, or infeasible in every state under the information available at planning. The Supplementary Materials provide proofs and boundary examples for the theorems stated in the main article.

The indifference region for the effect (d_0, d_1) is different from the set $\Gamma(s)$ of sample sizes with unidentified feasibility. The error constraints impose no guarantee for effects in that interval. In the known-variance example, the minimum effect rule has the probability of a positive decision $\Phi\{\sqrt{n}(d - d_0) - z_{1-\alpha}\}$, which tends to one for every fixed $d > d_0$, including $d_0 < d < d_1$. A positive result supports exceeding d_0; it does not establish $d \geq d_1$. The boundary d_1 is where power is guaranteed, and a negative result does not establish equivalence. The two one-sided tests procedure (TOST) tests a null hypothesis of nonequivalence against an alternative within prespecified equivalence margins [36]. If a positive or negative decision would have important consequences for effects in the indifference region, the design must include a corresponding loss function or additional risk constraints. Supplementary Figure S1 shows how the probability of a positive decision under the t test varies with the true effect.

Several limitations determine where the conclusions can be applied. Table 6 separates the general identification argument from the assumptions needed for the monotonicity result, binary calculation, and two illustrations. These conditions must be checked again when the outcome, dependence structure, or decision rule changes.

The sharp bounds are conditional on the declared state space and design requirements. The binary calculation assumes the stated margins, independent sampling units, an affine mass function, and a fixed finite sample rule. The vision analysis uses historical marginal rates to bound discordance and evaluates one fixed zero-boundary rule across compatible states. The supplementary minimum effect calculations assume a known discordance probability; the historical marginal rates alone do not identify that probability. The vision analysis does not estimate current population quantities or the frequency of the five cases.

Substantive justification of the target regions and error tolerances remains external to the statistical calculation. Feasibility alone does not rank several candidates. A separate cost or utility criterion is needed.

Table 6. Scope of the main results and conditions requiring new analysis.

Result	General Conclusion	Condition to Check in a New Application
Identification and sharp bounds	Theorem 2 records the factorization criterion and Theorem 4 applies the standard intersection and union construction.	Specify the state space, planning information, nonempty identified set, and meaning of feasibility.
Larger experiments	Theorem 1 specializes the Blackwell order to the decision class.	Verify the reduction, closure under induced rules, and stability of the decision and error definitions.
Classification into five cases	Proposition 2 is the set partition induced by the nested feasible sets.	Establish the inner and outer bounds and any rule that is claimed to work uniformly.
Binary calculation	Theorem 5 records an established Bernoulli compatibility representation; Theorem 6 applies polynomial sign analysis.	A different set of margins, outcome type, dependence structure, or nuisance optimization requires a new calculation.
Normal example	Section 6.1 compares known-variance z and unknown-variance t rules for independent normal observations.	Section 6.1.1 supplies the unknown-variance t calculation. Two-sided rules, nonnormal sampling, or different claims require separate evaluation.
Public data illustration	Section 6.2 gives sharp bounds for two candidate sample sizes under the fixed zero-boundary rule for independent paired binary observations. The sensitivity analysis evaluates every size from 1 to 7477 under seven planning scenarios.	Tables 2–5 condition on historical planning inputs. Population uncertainty, clustered pairs, other thresholds or domains, and other outcome definitions require separate evaluation.

The pilot extension gives conservative confidence bounds for feasible sets under its sampling assumptions, and the t calculations address unknown normal variance. Further work could seek less conservative confidence bounds and methods for multivariate optimization with rigorous error bounds. Extensions are also needed for clustered observations and resource constraints. Sequential and adaptive studies require the complete protocol to be treated as the candidate design. Optimization over nuisance parameters with rigorous numerical error bounds would extend the polynomial calculations.

Applied use begins with the sharp bounds. Sizes in the inner bound can enter a cost or utility comparison. For sample sizes between the bounds, additional information about dependence or nuisance parameters is needed to determine feasibility. An empty outer bound calls for revision of the rule or design requirements. A minimum effect test may be appropriate when the zero-boundary test fails an error constraint, provided that the minimum effect test uses available information and satisfies both error constraints. Failure of the zero-boundary test does not justify capping recruitment or discarding observations.

8. Conclusions

The question is which sample sizes can be justified when the planning information admits several specifications of the data distribution. We evaluate the feasible set under each specification. If the feasible set is the same for every specification, the planning information identifies it. Otherwise, the intersection and union give sharp inner and outer bounds. The framework also distinguishes the failure of a prespecified rule from limitations of the experiment. In the normal and vision examples, a zero-boundary test can impose an upper feasible boundary because negligible effects increasingly produce positive decisions. In the normal example, a minimum effect test remains feasible while retaining all observations. The vision example establishes sharp bounds for the fixed zero-boundary rule. The present calculations do not establish the feasibility of a minimum effect test when discordance is unknown. In practice, candidates in the inner bound can proceed to cost or utility assessment. For candidates between the bounds, additional information about dependence or nuisance parameters is needed to determine feasibility. An empty outer bound shows that changing sample size alone cannot meet the requirements, so the rule or design specification must be revised.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/stats9050105/s1, The Supplementary Materials contain proofs, boundary examples, and reproduction instructions. The file ssadmiss_0.5.0rc1.zip contains the reference implementation, tests, and example scripts. The script vision_real_data_example.R reproduces the public data illustration and its deterministic sensitivity analysis. The additional analysis archive reproduces the t calculations, simulation of independent pilots, computational benchmarks, and example of software use.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. The empirical illustration is a secondary methodological analysis of a public, anonymous, historical contingency table distributed with the R package irr. No new participants were recruited, no identifiable records were accessed, and no intervention was performed.

Informed Consent Statement: Not applicable. The study did not recruit participants or access individual-level or identifiable records.

Data Availability Statement: The vision data are distributed with the R package irr; the analysis uses the variables r.eye and l.eye. The script vision_real_data_example.R reproduces the public data illustration and all quantities reported in Tables 2–5, including the seven sensitivity scenarios. It records the R and irr versions used. The reference implementation, tests, example scripts, and reproduction instructions are provided with this submission as Supplementary Software (ssadmiss_0.5.0rc1.zip), released under the Berkeley Software Distribution 3-Clause License. The additional analysis archive contains the additional scripts, numerical outputs, and environment information for Sections S7–S10.

Acknowledgments: During the preparation of this manuscript, the author used OpenAI ChatGPT and Codex with the GPT-5.6 model to assist with manuscript editing and the development of the reference implementation. The author reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Blackwell, D. Equivalent comparisons of experiments. *Ann. Math. Stat.* **1953**, *24*, 265–272. [CrossRef]
2. Manski, C.F. *Partial Identification of Probability Distributions*; Springer: New York, NY, USA, 2003. [CrossRef]
3. Beresteanu, A.; Molchanov, I.; Molinari, F. Partial identification using random set theory. *J. Econ.* **2012**, *166*, 17–32. [CrossRef]
4. Murphy, K.R.; Myors, B. Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model. *J. Appl. Psychol.* **1999**, *84*, 234–248. [CrossRef]

5. Bahadur, R.R. A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction*; Solomon, H., Ed.; Stanford University Press: Stanford, CA, USA, 1961; pp. 158–168.
6. Butterley, P.; Sudbery, A.; Szulc, J. Compatibility of subsystem states. *Found. Phys.* **2006**, *36*, 83–101. [CrossRef]
7. Fontana, R.; Semeraro, P. Representation of multivariate Bernoulli distributions with a given set of specified moments. *J. Multivar. Anal.* **2018**, *168*, 290–303. [CrossRef]
8. Chuang-Stein, C.; Kirby, S.; Hirsch, I.; Atkinson, G. The role of the minimum clinically important difference and its impact on designing a trial. *Pharm. Stat.* **2011**, *10*, 250–256. [CrossRef] [PubMed]
9. Roychoudhury, S.; Scheuer, N.; Neuenschwander, B. Beyond p-values: A phase II dual-criterion design with statistical significance and clinical relevance. *Clin. Trials* **2018**, *15*, 452–461. [CrossRef] [PubMed]
10. Armstrong, R.A. Is there a large sample size problem? *Ophthalmic Physiol. Opt.* **2019**, *39*, 129–130. [CrossRef] [PubMed]
11. Lakens, D. Sample size justification. *Collabra Psychol.* **2022**, *8*, 33267. [CrossRef]
12. Gillariose, J.; Abdelwahab, M.M.; Venkatesan, R.; Joseph, J.; Abdelkawy, M.A.; Hasaballah, M.M. Design and analysis of reliability sampling plans based on the Topp–Leone generated Weibull distribution. *Symmetry* **2025**, *17*, 1439. [CrossRef]
13. O'Hagan, A.; Stevens, J.W.; Campbell, M.J. Assurance in clinical trial design. *Pharm. Stat.* **2005**, *4*, 187–201. [CrossRef]
14. Ríos Insua, D.; Ruggeri, F. (Eds.) *Robust Bayesian Analysis*; Springer: New York, NY, USA, 2000. [CrossRef]
15. Brutti, P.; De Santis, F.; Gubbiotti, S. Robust Bayesian sample size determination in clinical trials. *Stat. Med.* **2008**, *27*, 2290–2306. [CrossRef] [PubMed]
16. Kunzmann, K.; Grayling, M.J.; Lee, K.M.; Robertson, D.S.; Rufibach, K.; Wason, J.M.S. A review of Bayesian perspectives on sample size derivation for confirmatory trials. *Am. Stat.* **2021**, *75*, 424–432. [CrossRef] [PubMed]
17. Wilson, K.J.; Williamson, S.F.; Allen, A.J.; Williams, C.J.; Hellyer, T.P.; Lendrem, B.C. Bayesian sample size determination for diagnostic accuracy studies. *Stat. Med.* **2022**, *41*, 2908–2922. [CrossRef] [PubMed]
18. Willan, A.R. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clin. Trials* **2008**, *5*, 587–594. [CrossRef] [PubMed]
19. Manski, C.F. Identification and statistical decision theory. *Econom. Theory* **2025**, *41*, 977–993. [CrossRef]
20. Christensen, T.; Moon, H.R.; Schorfheide, F. Optimal decision rules when payoffs are partially identified. *Rev. Econ. Stud.* **2026**, rdag017. [CrossRef]
21. Montiel Olea, J.L.; Qiu, C.; Stoye, J. Decision theory for treatment choice problems with partial identification. *Rev. Econ. Stud.* **2026**, rdag015. [CrossRef]
22. Walley, P. *Statistical Reasoning with Imprecise Probabilities*; Chapman and Hall: London, UK, 1991.
23. Bacchetti, P.; Wolf, L.E.; Segal, M.R.; McCulloch, C.E. Ethics and sample size. *Am. J. Epidemiol.* **2005**, *161*, 105–110. [CrossRef] [PubMed]
24. Bacchetti, P.; McCulloch, C.E.; Segal, M.R. Simple, defensible sample sizes based on cost efficiency. *Biometrics* **2008**, *64*, 577–585. [CrossRef] [PubMed]
25. Qiu, S.-F.; Poon, W.-Y.; Tang, M.-L. Sample size determination for disease prevalence studies with partially validated data. *Stat. Methods Med. Res.* **2016**, *25*, 37–63. [CrossRef] [PubMed]
26. Gustafson, P. Sample size implications when biases are modelled rather than ignored. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2006**, *169*, 865–881. [CrossRef]
27. Fréchet, M. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon 3E Ser. Sci. Sect. A* **1951**, *14*, 53–77.
28. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [CrossRef] [PubMed]
29. Imbens, G.W.; Manski, C.F. Confidence intervals for partially identified parameters. *Econometrica* **2004**, *72*, 1845–1857. [CrossRef]
30. Stoye, J. More on confidence intervals for partially identified parameters. *Econometrica* **2009**, *77*, 1299–1315. [CrossRef]
31. Ben-Tal, A.; Goryashko, A.; Guslitzer, E.; Nemirovski, A. Adjustable robust solutions of uncertain linear programs. *Math. Program.* **2004**, *99*, 351–376. [CrossRef]
32. Basu, S.; Pollack, R.; Roy, M.-F. *Algorithms in Real Algebraic Geometry*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006. [CrossRef]
33. Stuart, A. The estimation and comparison of strengths of association in contingency tables. *Biometrika* **1953**, *40*, 105–110. [CrossRef]
34. Gamer, M.; Lemon, J.; Fellows, I.; Singh, P. *irr: Various Coefficients of Interrater Reliability and Agreement*, version 0.84.1; Comprehensive R Archive Network: Vienna, Austria, 2019. Available online: https://CRAN.R-project.org/package=irr (accessed on 31 August 2026).

35. R Core Team. *R: A Language and Environment for Statistical Computing*, version 4.4.2; R Foundation for Statistical Computing: Vienna, Austria, 2024. Available online: https://www.R-project.org/ (accessed on 31 August 2026).
36. Wellek, S. *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2010. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.