UK Academic Publishing

Greenwich Journal of Mathematical and Statistical Analytics (GJMSA)

Vol. 01 | No. 01 | 2026 |pp. 23-32 Website: https://www.ukapub.com/journals/gjmsa

Received: 2 November 2026 Revised: 13 November 2026 Accepted: 18 November 2026 Published: 20 November 2026

Original Research Article

Asymmetric Finite-Sample Confidence Intervals for Inverse-Probability Estimation under Informative Missingness

Abdul Waheed [1]

[1] School of Mathematics and Statistics, Southwest University, Beibei, Chongqing, 400715, China

* Corresponding author: waheedmathswu@qq.com

Abstract: Unequal response rates can bias mean estimates when units respond infrequently. We consider bounded outcomes observed through independent Bernoulli indicators with known response probabilities. For a clipped inverse-weighted mean, we get exact bias and variance and a sharp worst-case exponential-moment bound. This gives asymmetric finite sample confidence intervals for the mean that directly adjust for clipping bias. We show that a clipping level chosen based solely on the response design maintains coverage without a selection penalty. We also permit each response chance to be in a fixed range and give a rule for its worst case. These intervals maintain coverage in simulations of three designs at two sample sizes but can be wide. Severe imbalance can cause normal weighted intervals to undercover, while complete-case intervals can miss the target. The results require fixed bounded outcomes, independent responses, and known or bounded response probabilities. These bounds can serve as an orientation in cases where normal approximations might not be valid.

Keywords: Missing data; inverse-probability weighting; concentration inequalities; interval estimation; response bias.

1. Introduction:

Missing outcomes are common in surveys, electronic records and observational studies, sometimes at rates that vary systematically between units. Inverse-probability weighting (based on the Horvitz–Thompson estimator [5]) adjusts for unequal probabilities of inclusion. Observational inference focuses on propensity-score methods [8]. The weights may be very variable when the probability of observing an event is small. Overlap restrictions and stabilized weights were studied [3]. Another alternative when outcome modeling is plausible are doubly robust procedures [1].

In moderate samples, conventional Wald intervals may be fragile due to variability in estimated variances with rare responses, and a normal approximation may mask asymmetric inverse weights. Clipping reduces dispersion but introduces bias. Concentration methods [2,4] give finite-sample versions, and inverse-propensity risk bounds have been employed in learning

© 2026 The Author(s). Published by UK Academic Publishing. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) License.

from logged data [9]. Here the target is a bounded finite-population mean so explicit design-based bias control is possible.

Here we investigate a family of inverse-probability estimators based on the response design itself that are clipped. Instead of bounding both tails by one symmetric constant, we get two different Chernoff radii, by using the exact moment-generating function of the Bernoulli response. The unknown outcomes are handled using a sharp envelope: the worst-case centered moment-generating function for each exponential tilt is when all bounded outcomes are at their upper endpoint. The resulting interval is asymmetric, and its upper endpoint explicitly includes the maximum possible downward clipping bias. We also consider an uncertain propensity specification with closed-form characterization of the least-favourable Bernoulli probability.

The contributions are thus tailored to the specified design-based context: (i) a uniform exponential-moment certificate and a finite-sample interval, (ii) a validity-preserving design-only choice of the clipping level and a comparison with symmetric bounds, and (iii) a propensity-uncertainty extension and a fully reproducible numerical assessment. We do not claim optimality among all possible confidence procedures, and the construction is not doubly robust. The model is presented in Section 2, the mathematical results in Sections 3-5, the implementation in Section 6, the simulations in Section 7 and limitations and conclusions in Section 8.

2. Response design and estimation target:

Consider a finite population of $n \geq 1$ labelled units. The values

$$y = (y_1, \dots, y_n) \in [0,1]^n, \qquad \theta_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

are fixed, and $R_i \in \{0,1\}$ indicates whether y_i is observed. Throughout Sections 2–4, conditionally on the population,

$$R_1, \dots, R_n \text{ are independent}, \qquad R_i \sim \text{Bernoulli}(p_i), \quad 0 < p_i < 1, \qquad (2.1)$$

where p_i is known for each unit. The probabilities need not be the same, or independent of y_i. The response design may therefore be informative; inference is supported by the correctness of the specified p_i, not by a model for the outcome distribution. Probabilities either must be known or determined independently of the observed $R_i y_i$; estimated propensity scores require separate treatment.

For a clipping level $0 < \tau \leq 1$, define

$$d_i(\tau) = \max\{p_i, \tau\}, \qquad \hat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n} \frac{R_i y_i}{d_i(\tau)}. \qquad (2.2)$$

Taking $\tau \leq p_{\min} = \min_i p_i$ recovers the usual inverse-probability estimator. We use $\tau \in [p_{\min}, 1]$ for tuning: there is no benefit to allowing smaller values, which yield identical weights.

Proposition 2.1 (Exact risk decomposition). Under (2.1), write

$$B_\tau(y) = \theta_n - \mathbb{E}\hat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{p_i}{d_i(\tau)}\right) y_i, \qquad B_\tau^{\max} = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{p_i}{d_i(\tau)}\right). \qquad (2.3)$$

Then $0 \leq B_\tau(y) \leq B_\tau^{\max}$ and

$$\mathrm{Var}(\hat{\theta}_\tau) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{p_i(1-p_i)y_i^2}{d_i(\tau)^2}, \qquad \mathbb{E}(\hat{\theta}_\tau - \theta_n)^2 = \mathrm{Var}(\hat{\theta}_\tau) + B_\tau(y)^2. \qquad (2.4)$$

Proof. Independence and $\mathbb{E}R_i = p_i$, $\mathrm{Var}(R_i) = p_i(1-p_i)$ imply

$$\mathbb{E}\hat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n}\frac{p_i y_i}{d_i(\tau)}, \quad \mathrm{Var}\hat{\theta}_\tau = \frac{1}{n^2}\sum_{i=1}^{n}\frac{y_i^2\mathrm{Var}(R_i)}{d_i(\tau)^2}.$$

Since $0 \le y_i \le 1$ and $p_i \le d_i(\tau)$, every bias term is between 0 and its coefficient. Adding the squared bias to the variance gives the mean squared error identity. □

Equation (2.4) shows the usual bias–variance tension. Instead we will minimize a guarantyd width of the interval that does not depend on the unobserved y_i or an estimated variance as our tuning criterion.

3. A least-favourable exponential-moment

Set $c_i(\tau) = 1/[nd_i(\tau)]$. Define the centered estimator

$$S_\tau = \hat{\theta}_\tau - \mathbb{E}\hat{\theta}_\tau = \sum_{i=1}^{n} c_i(\tau)y_i(R_i - p_i). \qquad (3.1)$$

For $t \in \mathbb{R}$ let

$$K_\tau(t) = \sum_{i=1}^{n}\{\log(1 - p_i + p_i e^{tc_i(\tau)}) - tp_i c_i(\tau)\}. \qquad (3.2)$$

Unlike a sub-Gaussian variance proxy, K_τ retains the skewness of the heterogeneous Bernoulli design.

Lemma 3.1 (Uniform and sharp exponential-moment bound). For every $t \in \mathbb{R}$,

$$\sup_{y\in[0,1]^n} \log\mathbb{E}\exp(tS_\tau) = K_\tau(t). \qquad (3.3)$$

The supremum is attained at $y_1 = \cdots = y_n = 1$; hence the envelope is exact for each fixed exponential tilt.

Proof. Independence gives

$$\mathbb{E}e^{tS_\tau} = \prod_{i=1}^{n}\{(1-p_i)e^{-tc_i p_i y_i} + p_i e^{tc_i(1-p_i)y_i}\}.$$

For fixed i and t, put

$$h_i(u) = (1-p_i)e^{-tc_i p_i u} + p_i e^{tc_i(1-p_i)u}, \qquad 0 \le u \le 1.$$

Direct differentiation yields

$$h_i'(0) = 0, \qquad h_i''(u) = (tc_i)^2\{(1-p_i)p_i^2 e^{-tc_i p_i u} + p_i(1-p_i)^2 e^{tc_i(1-p_i)u}\} \ge 0.$$

Thus h_i is nondecreasing on $[0,1]$ and $h_i(y_i) \le h_i(1)$. Equality holds at $y_i = 1$ for all i. Multiplying and taking logarithms yields (3.3). □

For a confidence budget $a \in (0,1)$, define the one-sided radii

$$q_\tau^+(a) = \inf_{t>0}\frac{K_\tau(t) + \log(1/a)}{t}, \qquad q_\tau^-(a) = \inf_{t>0}\frac{K_\tau(-t) + \log(1/a)}{t}. \qquad (3.4)$$

They are both deterministic functions of the design probabilities and τ. The infimum in (3.4) is interpreted via deterministic approximation: any finite positive t is a valid certificate, and

any sequence converging to the infimum preserves the probability constraint by continuity from above.

Theorem 3.2 (Asymmetric finite-sample confidence interval). Assume (2.1) and fix $\alpha \in (0,1)$. For any τ determined solely from $(p_1, \dots, p_n, \alpha)$, let

$$I_\tau = \left[\hat{\theta}_\tau - q_\tau^+(\alpha/2),\ \hat{\theta}_\tau + q_\tau^-(\alpha/2) + B_\tau^{\max}\right] \cap [0,1]. \qquad (3.5)$$

Then, uniformly for every fixed $y \in [0,1]^n$,

$$\mathbb{P}_p\{\theta_n \in I_\tau \mid y\} \geq 1 - \alpha. \qquad (3.6)$$

Proof. For every $t > 0$, Markov's inequality and Lemma 3.1 imply

$$\mathbb{P}_p(S_\tau \geq x \mid y) \leq e^{-tx}\mathbb{E}e^{tS_\tau} \leq \exp\{-tx + K_\tau(t)\}.$$

Choosing $x = \{K_\tau(t) + \log(2/\alpha)\}/t$ bounds this probability by $\alpha/2$; optimization gives

$$\mathbb{P}_p\{S_\tau > q_\tau^+(\alpha/2) \mid y\} \leq \alpha/2.$$

Applying the same argument to $-S_\tau$ gives

$$\mathbb{P}_p\{-S_\tau > q_\tau^-(\alpha/2) \mid y\} \leq \alpha/2.$$

Outside the union of these two events, with $m_\tau = \mathbb{E}\hat{\theta}_\tau$,

$$\hat{\theta}_\tau - q_\tau^+(\alpha/2) \leq m_\tau \leq \hat{\theta}_\tau + q_\tau^-(\alpha/2).$$

Proposition 2.1 yields

$$m_\tau \leq \theta_n = m_\tau + B_\tau(y) \leq m_\tau + B_\tau^{\max}.$$

Combining the inequalities and using $\theta_n \in [0,1]$ proves (3.6) by the union bound. □

Remark (Direction of the corrections). The lower bound is given by q^+, the threshold of positive variations of the estimator. The upper bound uses q^- and the positive worst-case bias, since clipping can only decrease the expectation for non-negative outcomes. Turning the two tails over is usually a mistake.

4. Width optimization, comparison, and ratesenvelope

Before intersecting with $[0,1]$, the interval (3.5) has deterministic width

$$W_\tau = q_\tau^+(\alpha/2) + q_\tau^-(\alpha/2) + B_\tau^{\max}. \qquad (4.1)$$

Let $\mathcal{T}(p) \subset [p_{\min}, 1]$ be a nonempty finite set computed only from the probabilities. Define

$$\hat{\tau}_{\text{des}} = \arg\min_{\tau \in \mathcal{T}(p)} W_\tau, \qquad (4.2)$$

using a predetermined tie-breaking criteria. A continuum minimizer can also be used: the functions in (4.1) are continuous on $[p_{\min}, 1]$ (as one-sided Chernoff boundary values) and so there is a minimum; a finite grid is computationally easier.

Proposition 4.1 (Valid design-only tuning). The interval $I_{\hat{\tau}_{\text{des}}}$ satisfies (3.6), with no multiplicity correction. Moreover, if

$$M_\tau = \max\{q_\tau^+(\alpha/2), q_\tau^-(\alpha/2)\}, \qquad W_\tau^{\text{sym}} = 2M_\tau + B_\tau^{\max},$$

then

$$\min_{\tau \in \mathcal{T}(p)} W_\tau \leq \min_{\tau \in \mathcal{T}(p)} W_\tau^{\text{sym}}. \qquad (4.3)$$

Proof. The value $\hat{\tau}_{\text{des}}$ that has been chosen is invariant for the design p set. Therefore, Theorem 3.2 is directly applicable at that point and no outcome or response metric has been used to select it. for all τ,

$$q_\tau^+(\alpha/2) + q_\tau^-(\alpha/2) \leq 2\max\{q_\tau^+(\alpha/2), q_\tau^-(\alpha/2)\}.$$

Adding $B_\tau^{\max}$ and minimizing proves (4.3). □

Corollary 4.2 (A universal finite-sample rate). For every $\tau > 0$ and $a \in (0,1)$,

$$q_\tau^{\pm}(a) \le \sqrt{\frac{\log(1/a)}{2n^2}\sum_{i=1}^{n}\frac{1}{d_i(\tau)^2}} \le \frac{1}{\tau}\sqrt{\frac{\log(1/a)}{2n}}. \qquad (4.4)$$

If $\inf_{n,i} p_i \ge p_0 > 0$, the untruncated interval has width at most $2p_0^{-1}\sqrt{\log(2/\alpha)/(2n)}$, and the design-tuned interval is no wider on the untruncated scale.

Proof. For $U_i = c_i(\tau)(R_i - p_i)$, the range length is $c_i(\tau)$. Hoeffding's lemma gives

$$\log \mathbb{E}\exp(tU_i) \le \frac{t^2 c_i(\tau)^2}{8}.$$

Summing and using (3.2) yields $K_\tau(\pm t) \le t^2 \sum_i c_i^2 /8$. Hence

$$q_\tau^{\pm}(a) \le \inf_{t>0}\left\{\frac{t}{8}\sum_i c_i^2 + \frac{\log(1/a)}{t}\right\} = \sqrt{\frac{\log(1/a)}{2}\sum_i c_i^2}.$$

The second inequality in (4.4) follows from $d_i \ge \tau$. When $\tau \le p_{\min}$ the bias bound vanishes, and $p_{\min} \ge p_0$ gives the asserted order. The tuned width is bounded above by the width of that grid candidate whenever the grid contains $p_{\min}$. □

For nonuniform designs with $p_{\min} \to 0$, the rate in Corollary 4.2 need not hold uniformly. If truncation remains fixed while a positive fraction of $p_i < \tau$, the worst-case bias need not vanish. This is a genuine identification–precision trade-off, not a defect that a confidence algorithm can remove without more assumptions.

5. Uncertain inclusion probabilities

In practice, probabilities may be supplied as externally validated ranges rather than exact values. Suppose known, nonrandom numbers satisfy

$$0 < \ell_i \le p_i \le u_i < 1, \qquad 0 < q_i < 1, \qquad (5.1)$$

where q_i is a nominal probability used to construct $d_i = \max\{q_i, \tau\}$ and $c_i = 1/(nd_i)$. Unlike an ordinary plug-in interval, the following construction guarantees coverage simultaneously for every vector p in the rectangle $\prod_i [\ell_i, u_i]$.

Write for $v \in \mathbb{R}$,

$$g(p,v) = \log\big(1 + p(e^v - 1)\big) - pv, \qquad \overline{K}_\tau(t) = \sum_{i=1}^{n} \max_{p\in[\ell_i,u_i]} g(p, tc_i). \qquad (5.2)$$

Define $\overline{q}_\tau^{\pm}(a)$ as in (3.4) with K_τ replaced by $\overline{K}_\tau$. Further put

$$B_\tau^- = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{u_i}{d_i} - 1\right)_+, \qquad B_\tau^+ = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\ell_i}{d_i}\right)_+, \qquad x_+ = \max\{x, 0\}. \qquad (5.3)$$

Theorem 5.1 (Uniform coverage under propensity uncertainty). Under (5.1) and independent Bernoulli responses, define

$$\overline{I}_\tau = \left[\hat{\theta}_\tau - \overline{q}_\tau^{+}(\alpha/2) - B_\tau^-,\ \ \hat{\theta}_\tau + \overline{q}_\tau^{-}(\alpha/2) + B_\tau^+\right] \cap [0,1]. \qquad (5.4)$$

For any τ selected solely from (q, ℓ, u, α),

$$\inf_{p_i\in[\ell_i,u_i]} \inf_{y\in[0,1]^n} \mathbb{P}_p\{\theta_n \in \overline{I}_\tau \mid y\} \ge 1 - \alpha.$$

For any $v \neq 0$, the maximizer of $g(p, v)$ over $[\ell_i, u_i]$ is

$$p_i^*(v) = \text{clip}\left(\frac{1}{v} - \frac{1}{e^v - 1},\ \ell_i,\ u_i\right), \qquad (5.5)$$

with the continuous value $1/2$ for the unconstrained maximizer at $v = 0$. Thus the robust radii require only two scalar optimizations and no multidimensional search.

Proof. For an arbitrary admissible p, Lemma 3.1, with c_i defined using q_i, still gives

$$\log \mathbb{E}_p e^{t(\hat{\theta}_\tau - \mathbb{E}_p \hat{\theta}_\tau)} \leq \sum_i g\,(p_i, tc_i) \leq \overline{K}_\tau(t).$$

Applying Chernoff's argument to t and $-t$ yields one-sided deviations bounded by $\overline{q}_\tau^+$ and $\overline{q}_\tau^-$, respectively. Moreover,

$$\theta_n - \mathbb{E}_p \hat{\theta}_\tau = \frac{1}{n} \sum_i \left(1 - \frac{p_i}{d_i}\right) y_i, \qquad -B_\tau^- \leq \theta_n - \mathbb{E}_p \hat{\theta}_\tau \leq B_\tau^+.$$

The last inequality follows by bounding separately the negative and positive parts of each summand using $p_i \in [\ell_i, u_i]$ and $y_i \in [0,1]$. Combining these statements proves the confidence bound uniformly in p and y.

To compute the envelope, differentiation in p gives

$$\frac{\partial g}{\partial p}(p, v) = \frac{e^v - 1}{1 + p(e^v - 1)} - v, \qquad \frac{\partial^2 g}{\partial p^2}(p, v) = -\frac{(e^v - 1)^2}{[1 + p(e^v - 1)]^2} \leq 0.$$

The unique unconstrained critical point for $v \neq 0$ is $1/v - 1/(e^v - 1)$. Concavity implies that the constrained maximum is its projection onto $[\ell_i, u_i]$. The limit $1/2$ follows from $e^v - 1 = v + v^2/2 + O(v^3)$. □

Remark (Interpretation and restrictions). If $q_i = \ell_i = u_i = p_i$, Theorem 5.1 reduces to Theorem 3.2. The intervals $[\ell_i, u_i]$ must be fixed independently of the response data or guaranteed simultaneously by another procedure. Marginal 95% intervals for each p_i do *not* automatically furnish a simultaneous 95% rectangle; if a simultaneous probability statement has error γ, the unconditional coverage bound becomes $1 - \alpha - \gamma$ by another union bound.

6. Numerical implementation

For numerical stability, optimize the positive parameter $s = t/n$ rather than t itself. Since $tc_i = s/d_i$, write

$$K_\tau(\pm ns) = \sum_{i=1}^{n} \left[\log\{(1 - p_i) + p_i \exp(\pm s/d_i)\} \mp \frac{p_i s}{d_i}\right], \quad q_\tau^\pm(a)$$
$$= \inf_{s>0} \frac{K_\tau(\pm ns) + \log(1/a)}{ns}. \qquad (6.1)$$

The expression inside logarithms is evaluated by a log-sum-exp routine to avoid overflow, and the optimization is performed on $\log s$. The population probabilities are known before the response vector is drawn, so (6.1) and tuning need only be computed once for a given design. The algorithm is:

1. Fix α and construct a deterministic grid $\mathcal{T}(p)$ containing $p_{\min}$.
2. For each τ compute d_i, $B_\tau^{\max}$, and $q_\tau^\pm(\alpha/2)$ from (6.1).
3. Select the grid value minimizing W_τ in (4.1).
4. After observing $R_i y_i$, calculate (2.2) and return (3.5).

The same procedure with $\overline{K}$, B^-, and B^+ implements Theorem 5.1. For narrow intervals one may also allocate tail budgets $a_+ + a_- \le \alpha$ unequally, minimizing $q^+(a_+) + q^-(a_-) + B^{\max}$ *using the design only*. Theorem 3.2 then holds with $q^+(a_+)$ and $q^-(a_-)$ by the identical union-bound proof. We use the equal split $a_+ = a_- = \alpha/2$ below so comparisons share the same error allocation.

7. Reproducible Monte Carlo experiments

7.1 Designs and benchmarks

We conducted $B = 3000$ independent response replications per setting, with the NumPy PCG64 generator seeded at 20261114 and finite population sizes $n \in \{400,1600\}$. To eliminate a second layer of outcome-generation variability, each outcome vector is deterministic. For $x_i = (i - 1/2)/n$ $(i = 1, \dots, n)$, define

Moderate overlap: $p_i = 0.45 + 0.40x_i$, $y_i = 0.20 + 0.60x_i + 0.05\sin(10\pi x_i)$,

Informative response: $p_i = 0.07 + 0.78x_i$, $y_i = 0.90 - 0.70x_i + 0.04\sin(12\pi x_i)$,

Severe imbalance: $p_i = 0.025 + 0.75x_i^2$, $y_i = 0.95 - 0.75x_i^2 + 0.03\cos(14\pi x_i)$.

All values lie in $[0,1]$. In the latter two designs the units with larger outcomes have smaller response probabilities. In all settings the target is the *finite-population* $\theta_n = n^{-1}\sum_i y_i$, not the respondent mean.

The clipping grid contains the minimum probability, 23 equally spaced empirical probability-quantile levels from 0.04 to 0.96, and 1 (25 candidates in total). The compared methods are: (A) the tuned asymmetric interval (3.5); (S) a symmetrized interval with common radius $\max(q^+, q^-)$ and separately optimized grid level; (U) the asymmetric bound with no clipping, $\tau = p_{\min}$; (W) the ordinary 95% normal interval about the untruncated inverse-probability estimate with the unbiased design variance estimator

$$\hat{V} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{R_i(1-p_i)y_i^2}{p_i^2}; \qquad (7.1)$$

and (C) the complete-case normal interval about $\sum_i R_i y_i / \sum_i R_i$, using the usual respondent sample standard error. The complete-case interval is $[0,1]$ if there are no respondents. Every interval is intersected with $[0,1]$; widths below are therefore the *actual reported widths*, not the deterministic upper certificates.

7.2 Coverage and interval width

Table 1. Empirical coverage / mean interval width at nominal 95% confidence (3,000 replications). A: tuned asymmetric; S: tuned symmetric; U: asymmetric untruncated; W: inverse-probability Wald; C: complete-case Wald.

Design	n	A	S	U	W	C
Moderate	400	1.000 / .208	1.000 / .210	1.000 / .208	.945 / .067	.068 / .041
Moderate	1600	1.000 / .104	1.000 / .105	1.000 / .104	.949 / .034	.000 / .021
Informative	400	1.000 / .387	1.000 / .398	1.000 / .403	.944 / .221	.000 / .053
Informative	1600	.999 / .201	1.000 / .207	.999 / .201	.942 / .111	.000 / .026
Severe	400	1.000 / .633	1.000 / .650	.997 / .742	.917 / .461	.000 / .078
Severe	1600	1.000 / .391	1.000 / .407	.998 / .409	.934 / .265	.000 / .039

Table 2. Population means, mean and minimum response probabilities, and tuned clipping levels across the three response designs.

Design	n	θn	Mean p	pmin	τA	τS
Moderate	400	.500	.650	.451	.451	.451
Moderate	1600	.500	.650	.450	.450	.450
Informative	400	.550	.460	.071	.200	.200
Informative	1600	.550	.460	.070	.070	.101
Severe	400	.700	.275	.025	.183	.213
Severe	1600	.700	.275	.025	.057	.072

Table 1 shows a clear distinction between finite sample guaranties and empirical accuracy. The proposed intervals have nominal coverage or better in all simulation settings and often cover nearly all replications. Such is the nature of true conservatism, since the intervals are uniformly valid across all outcome vectors in $[0,1]^n$. The traditional inverse-probability intervals, hidden in the strict design, are much narrower (0.917 and 0.934, roughly). The complete-case approach is severely biased when the response depends on the unit outcomes. The intervals are narrow, but often do not hit the target in informative designs. These insights are specific to the designs cited, not assertions of universal superiority.

In the extreme case at $n = 400$, the mean width is reduced from 0.742 to 0.633 compared with U by the design choice of τ while still having a high observed coverage. Proposition 4.1. The asymmetric certificate is always tighter than the symmetric one for the same clipping threshold. The different grid optimizations in Table 1 follow that pre-intersection bound. In the moderate design we do not use clipping as the variance advantage does not outweigh the worst case bias. The nominal 95% coverage corresponds to a binomial Monte Carlo standard error of approximately 0.004 for a method with true coverage of 0.95 based on 3000 replications.

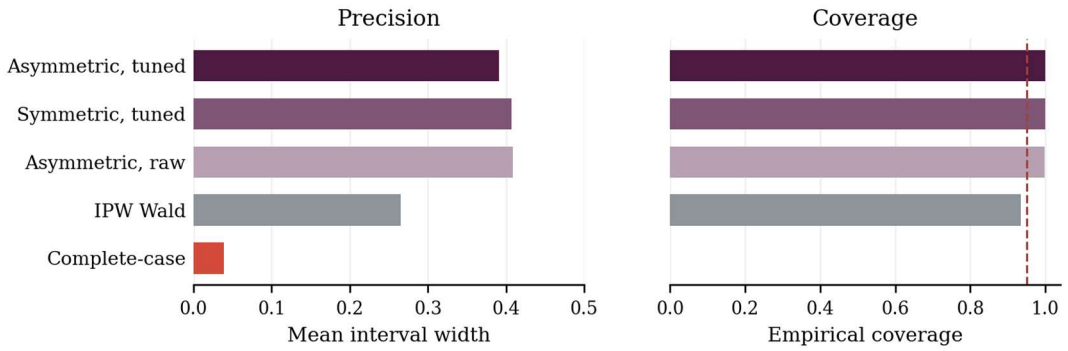


Figure 1. Mean width and empirical coverage for the severe response design (n = 1,600). Dashed line: nominal 95% coverage.

7.3 What the numerical evidence does and does not establish

The tightness of the bound for typical populations is not vindicated by the experiments. However, the finite-sample coverage statement and the consequent deterioration in precision relative to approximate methods are vindicated. Indeed, the moment generating function envelope in Lemma 3.1 is attained at the extreme outcome vector $y_i = 1$. Our three experiments use more moderate values, typically below this threshold. This discrepancy explains much of the behavior of conservatives. Practical implementation is most suitable if a guarantyd limit is of utmost importance, if the range of outcomes is inherently limited, and if the probabilities of inclusion can be reliably determined. If valid, unit-specific upper bounds

$0 \le y_i \le u_i^{(y)} \le 1$ are externally available, the proof of Lemma 3.1 also permits replacing c_i by $c_i u_i^{(y)}$ in the envelope and replacing $B_\tau^{\max}$ by $n^{-1}\sum_i (1 - p_i/d_i)\, u_i^{(y)}$, with a potentially smaller certificate.

8. Discussion and Conclusion

We derived finite-sample, distribution-free confidence intervals for a finite-population mean observed via heterogeneous Bernoulli response. The key mathematical step is an exact least favorable exponential-moment envelope for the centered clipped inverse-probability estimator. The two unilateral Chernoff radii address skewness, and coincide with a fixed, one-sided clipping-bias constraint. A clipping threshold chosen just by the design probabilities preserves conditional coverage. The same argument extends to externally validated uncertainty intervals for response probabilities; concavity implies a unique least-favorable probability for each exponential tilt.

Many constraints are important. First, simultaneous coverage requires independent response indicators, correct knowledge of the design, or valid propensity uncertainty set, and bounded outcomes. These theorems do not deal with correlated or clustered responses. Second, the intervals can be much wider than model based alternatives, especially when many response probabilities are small. Third, a propensity model built from the same response data cannot stand in for the known p_i; validation or other inferential adjustment would be required. Fourth, worst-case coverage is not equivalent to short expected length. Although these limitations exist, the findings offer a powerful and readily replicable method for sensitive inference in informative nonresponse and point to conservative calibration and propensity uncertainty as clear targets for future methodological development.

Data availability. All experimental populations are generated by the deterministic equations shown. The complete Python replication script and formatted results are provided alongside the manuscript. No human-related or confidential records were utilized.

Conflict of interest. The author declares no competing interests.

References

[1] Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics, 61*(4), 962–973. https://doi.org/10.1111/j.1541-0420.2005.00377.x

[2] Boucheron, S., Lugosi, G., & Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[3] Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika, 96*(1), 187–199. https://doi.org/10.1093/biomet/asn055

[4] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association, 58*(301), 13–30. https://doi.org/10.1080/01621459.1963.10500830

[5] Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47*(260), 663–685. https://doi.org/10.1080/01621459.1952.10483446

[6] Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association, 113*(521), 390–400. https://doi.org/10.1080/01621459.2016.1260466

[7] Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association, 89*(427), 846–866. https://doi.org/10.1080/01621459.1994.10476818

[8] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

[9] Swaminathan, A., & Joachims, T. (2015). Counterfactual risk minimization: Learning from logged bandit feedback. *Proceedings of Machine Learning Research, 37*, 814–823. https://proceedings.mlr.press/v37/swaminathan15.html