Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(9): 6957–6998.
DOI: 10.3934/era.2026302
Received: 27 May 2026
Revised: 05 August 2026
Accepted: 06 August 2026
Published: 14 August 2026

Research article

Bayesian credible and prediction intervals for the two-parameter negative binomial (NB2) model

Md Mahadi Hasan[1,*], **Md Monzur Murshed**[2] **and Md Nahid Hasan**[3,4]

[1] Department of Mathematics and Statistics, Murray State University, Murray, KY 42071, USA
[2] Department of Mathematics and Statistics, Minnesota State University, Mankato, MN 56001, USA
[3] Department of Mathematics, East Texas A&M University, Commerce, TX 75428, USA
[4] Department of Mathematics, Augusta University, Augusta, GA 30912, USA

* **Correspondence:** Email: mhasan18@murraystate.edu.

Abstract: Overdispersed count data are common in many applications, where the variance exceeds the mean, and the Poisson model is inadequate. The two-parameter negative binomial (NB2) model addresses this limitation through a dispersion parameter and provides a standard framework for modeling overdispersed counts. In this paper, we evaluated Bayesian credible intervals for the NB2 mean and Bayesian posterior predictive intervals for a single future count and for the mean of a future sample in a controlled one-sample, intercept-only setting. Posterior inference was obtained using weakly informative priors and Hamiltonian Monte Carlo with the No-U-Turn Sampler. The Bayesian intervals were assessed through simulation using operational coverage probability and expected width, and were compared with Wald, MLE-based normal, and parametric bootstrap confidence and prediction intervals. Prior-family sensitivity was examined using calibrated lognormal, Gamma, and log-Student-t priors. Robustness was also evaluated under zero-inflated NB2, structured-zero, and Poisson-lognormal data-generating mechanisms while fitting the NB2 model. Two real-data examples, involving Quine school-absence counts and epilepsy baseline seizure counts, illustrated the practical implementation of the methods.

Keywords: Bayesian inference; coverage probability; credible interval; Markov chain Monte Carlo (MCMC); overdispersed count data; posterior predictive interval; prediction of future counts; two-parameter negative binomial distribution (NB2)

1. Introduction

Count data are commonly modeled using the Poisson distribution, which assumes that the mean and variance are equal. In many applications, however, the empirical variance substantially exceeds

the mean, a phenomenon known as overdispersion. Examples include lesion counts in multiple sclerosis studies, crash counts in transportation safety, ecological counts, and internet packet arrivals [1–7]. In such settings, the Poisson model is inadequate, and more flexible count models are required. The two-parameter negative binomial (NB2) distribution, also known as the Poisson-gamma mixture model, provides a widely used alternative by introducing a dispersion parameter to accommodate extra-Poisson variation.

The NB2 parametrization separates the population mean μ from the dispersion parameter θ, with

$$\mathrm{Var}(Y) = \mu + \frac{\mu^2}{\theta}.$$

The one-sample model considered here is the intercept-only special case of NB2 regression. It is directly applicable when counts arise from comparable observational units measured over a common exposure period and no covariate adjustment is required. Examples include cross-sectional counts and baseline counts recorded over a fixed observation window. The simplified setting also provides a controlled benchmark for studying how n, μ, θ, and the future sample size m affect interval calibration and width without introducing covariate, clustering, or longitudinal effects. This setting is not presented as a new probability model. Its value lies in isolating the finite-sample behavior of interval procedures when the mean and dispersion are unknown. This remains a nontrivial problem because estimation of the negative-binomial dispersion parameter can be biased or numerically unstable in small and moderately sized samples, particularly under strong overdispersion [8, 9].

Frequentist inference for the negative-binomial model includes likelihood, score, adjusted-score, bootstrap, and large-sample procedures for the mean and dispersion parameters [10–16]. In the two-parameter NB2 setting, Hasan and Krishnamoorthy [8] studied confidence intervals for the mean and prediction intervals for a future sample mean. Kenne Pagui et al. [9] showed that ordinary maximum-likelihood estimation of the dispersion parameter can exhibit substantial finite-sample bias and developed mean and median bias-reduced procedures for negative-binomial regression.

Bayesian analysis and posterior prediction for negative-binomial models are also well established. Existing work includes Bayesian model comparison for overdispersed counts, hierarchical negative-binomial regression, zero-inflated negative-binomial models, robust Bayesian count modeling, and posterior predictive forecasting [17–20]. Bayesian prediction intervals for the negative-binomial distribution have also been studied by Hamura et al. [21], further indicating that the present contribution is not the introduction of posterior prediction, but a finite-sample comparison of Bayesian and frequentist interval procedures for the NB2 mean and future-count targets. He and Huang [19], for example, developed a Bayesian zero-inflated negative-binomial regression framework for spatiotemporal count data. Hamura et al. [20] studied robust Bayesian count modeling in the presence of zero inflation and outliers. Moreover Huang et al. [22] combined NB2 and ZINB2 likelihoods with temporal regression architectures and used posterior predictive simulation for forecast quantiles and anomaly probabilities.

Accordingly, we do not claim that Bayesian NB2 inference, MCMC estimation, or posterior predictive intervals are new; we address a narrower evaluation question. Here, we examine, under a controlled intercept-only NB2 design, the repeated-sampling calibration and width of credible and predictive intervals for μ, a single future count, and a future sample mean. We also evaluate how these results change across prior families and under specified departures from the fitted NB2 likelihood.

The Bayesian model, posterior distribution, and posterior predictive construction used in this paper follow standard Bayesian analysis for the NB2 distribution. The contribution of this study is therefore comparative and computational rather than the derivation of a new Bayesian model or sampling algorithm. The study provides a common finite-sample evaluation of parameter and prediction intervals while explicitly examining prior sensitivity, likelihood misspecification, computational reliability, and interval-construction failures.

The contributions are as follows:

i. We evaluate equal-tailed Bayesian credible intervals for the NB2 mean and posterior predictive intervals for a single future count and the mean of m future observations under a common simulation design.

ii. We compare the Bayesian procedures with Wald, MLE-based normal, and parametric bootstrap confidence and prediction intervals using matched data-generating streams.

iii. We assess prior-family sensitivity using calibrated lognormal, Gamma, and log-Student-t priors. This comparison evaluates distributional changes in the prior rather than only changes in the hyperparameters of one prior family.

iv. We examine likelihood misspecification by fitting the NB2 model to data generated from zero-inflated NB2, structured-zero, and Poisson-lognormal mechanisms. The Poisson-lognormal scenarios match the NB2 mean and variance while changing the distributional form.

v. We report operational coverage that counts failed interval constructions as noncoverage, coverage conditional on successful computation, interval width among successful fits, MCMC diagnostics, and method-specific failure rates.

vi. We illustrate the interval procedures using the Quine school-absence counts and one baseline eight-week seizure count per subject from the epilepsy data. These examples are treated as intercept-only analyses of counts observed over common measurement windows.

These contributions should be interpreted as an evaluation of Bayesian and frequentist interval procedures in a focused NB2 setting. They do not constitute a new Bayesian model, a new posterior distribution, or a new MCMC algorithm.

The remainder of this paper is organized as follows; In Section 2, we present the NB2 model, Bayesian interval procedures, and frequentist comparison methods. In Section 3, we report the simulation study based on coverage probability and expected width. In Section 4, we illustrate the methods using two real datasets. In Section 5, conclude with the major findings and future directions.

2. Methodology

Let $Y_1, \ldots, Y_n$ be i.i.d. count data from the two-parameter negative binomial distribution (NB2) with mean $\mu > 0$ and dispersion parameter $\theta > 0$. The probability mass function is

$$\Pr(Y = y \mid \mu, \theta) = \frac{\Gamma(y+\theta)}{\Gamma(\theta)\Gamma(y+1)} \left(\frac{\mu}{\mu+\theta}\right)^y \left(\frac{\theta}{\mu+\theta}\right)^\theta, \qquad y = 0, 1, 2, \ldots. \tag{2.1}$$

This parameterization satisfies

$$\mathbb{E}(Y) = \mu, \qquad \mathrm{Var}(Y) = \mu + \frac{\mu^2}{\theta}.$$

The likelihood for (μ, θ) given $\mathbf{y} = (y_1, \ldots, y_n)$ is

$$L(\mu, \theta \mid \mathbf{y}) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \theta)}{\Gamma(\theta)\Gamma(y_i + 1)} \left(\frac{\mu}{\mu + \theta}\right)^{y_i} \left(\frac{\theta}{\mu + \theta}\right)^{\theta}. \tag{2.2}$$

Up to an additive constant, the log-likelihood is

$$\begin{aligned} \ell(\mu, \theta) = &\sum_{i=1}^{n} \log \Gamma(y_i + \theta) - n \log \Gamma(\theta) + \left(\sum_{i=1}^{n} y_i\right) \log \mu + n\theta \log \theta \\ &- \left(\sum_{i=1}^{n} y_i + n\theta\right) \log(\mu + \theta) + \text{const.} \end{aligned} \tag{2.3}$$

2.1. Baseline prior and prior-family sensitivity

For the main simulation study, positivity is enforced by sampling the parameters on the log scale. Let

$$\eta_\mu = \log \mu, \qquad \eta_\theta = \log \theta.$$

The baseline prior specification is

$$\eta_\mu \sim N(0, 2^2), \qquad \eta_\theta \sim N(0, 2.5^2), \qquad \eta_\mu \perp \eta_\theta.$$

Equivalently,

$$\mu \sim \text{Lognormal}(0, 2^2), \qquad \theta \sim \text{Lognormal}(0, 2.5^2).$$

These are proper priors with median one and broad right tails. Their central 95% prior intervals are

$$\mu \in [\exp\{-2z_{0.975}\}, \exp\{2z_{0.975}\}] = [0.0198, 50.3968]$$

and

$$\theta \in [\exp\{-2.5z_{0.975}\}, \exp\{2.5z_{0.975}\}] = [0.0074, 134.2777],$$

where $z_{0.975} = 1.959964$. Thus, the baseline priors enable substantial uncertainty on the original parameter scale. Because their implications are most transparent on the outcome scale, we examine the induced prior predictive distributions before fitting the model [23–26].

Sensitivity to the prior distribution is assessed using three distinct prior families:

$$\begin{aligned} p_1: &\quad \log \mu \sim N(0, 2^2), && \log \theta \sim N(0, 2.5^2), \\ p_2: &\quad \mu \sim \text{Gamma}(0.2271, 0.03226), && \theta \sim \text{Gamma}(0.1669, 0.01011), \\ p_3: &\quad \log \mu \sim t_3(0, 1.232), && \log \theta \sim t_3(0, 1.540). \end{aligned}$$

Parameters μ and θ are assigned independent priors within each prior family. The Gamma distributions use the shape-rate parametrization. For the Student-t distributions, the second parameter denotes the scale.

The Gamma and log-Student-t hyperparameters are calibrated so that each alternative prior has median approximately one and approximately the same 0.975 quantile as the corresponding baseline lognormal prior. Under the baseline prior p_1, the relevant quantiles are

$$Q_{0.50}(\mu) = Q_{0.50}(\theta) = 1,$$

$$Q_{0.975}(\mu) = \exp\{2z_{0.975}\} = 50.3968,$$

and

$$Q_{0.975}(\theta) = \exp\{2.5z_{0.975}\} = 134.2777.$$

For a Gamma distribution with shape a and rate b, the calibration is obtained by solving

$$Q_{\text{Gamma}}(0.50; a, b) = 1$$

and

$$Q_{\text{Gamma}}(0.975; a, b) = Q_{0.975}(p_1).$$

Solving these two equations numerically gives

$$(a_\mu, b_\mu) = (0.2271336, 0.0322591)$$

for μ, and

$$(a_\theta, b_\theta) = (0.1669128, 0.0101115)$$

for θ. These numerical values are calibration results rather than separately selected scientific constants. The resulting Gamma priors have medians approximately equal to 1 and 0.975 quantiles approximately equal to 50.40 for μ and 134.30 for θ.

For the log-Student-t priors, let

$$t_{3,0.975} = 3.182446$$

denote the 0.975 quantile of a standard Student-t distribution with three degrees of freedom. The scale parameters are chosen as

$$s_\mu = \frac{2z_{0.975}}{t_{3,0.975}} = 1.2317568$$

and

$$s_\theta = \frac{2.5z_{0.975}}{t_{3,0.975}} = 1.5396960.$$

Therefore,

$$\exp\{s_\mu t_{3,0.975}\} \approx 50.40$$

and

$$\exp\{s_\theta t_{3,0.975}\} \approx 134.29.$$

Thus, the log-Student-t priors also have median 1 and approximately the same 0.975 quantiles as the baseline lognormal priors.

This calibration changes the prior shape while holding the median and one upper quantile approximately fixed. It does not make the three prior families identical or equally informative. The Gamma shape parameters are less than 1, so the Gamma alternatives place more prior mass near 0

than the baseline lognormal priors. The log-Student-t alternatives have heavier extreme tails than the lognormal priors, even though their 0.975 quantiles are matched. Consequently, p_2 examines sensitivity to greater prior concentration near zero, whereas p_3 examines sensitivity to heavier tail behavior. The three families therefore provide a prior-sensitivity comparison in which the central location and a selected upper quantile remain comparable while the distributional shape changes.

For the Gamma family, Stan samples

$$\eta_\mu = \log \mu, \qquad \eta_\theta = \log \theta.$$

The transformed densities therefore include the required Jacobian terms

$$\log \pi(\eta_\mu) = \log \pi_\Gamma(e^{\eta_\mu}) + \eta_\mu, \qquad \log \pi(\eta_\theta) = \log \pi_\Gamma(e^{\eta_\theta}) + \eta_\theta.$$

These Jacobian terms are required because the Gamma priors are defined for μ and θ on the original positive scale, whereas Stan samples their logarithms.

The prior-family sensitivity analysis uses 9 representative combinations of (μ_0, θ_0, n), with 1000 Monte Carlo replications for each scenario and prior family. Separately, 100,000 draws from each prior are used to summarize the corresponding induced prior predictive distribution before model fitting. This design directly addresses prior-family sensitivity by comparing changes in distributional form, not only changes in hyperparameters within a single lognormal family.

No prior is centered on $\log(\bar{y} + 0.1)$. The constant 0.1 and the corresponding empirical-Bayes prior specifications have been removed because they use the observed data to construct the prior and do not provide a fixed prior for the simulation comparison.

2.2. *Posterior distribution*

By Bayes' theorem, the joint posterior density is

$$p(\mu, \theta \mid \mathbf{y}) = \frac{L(\mu, \theta \mid \mathbf{y})\, \pi(\mu, \theta)}{\int_0^\infty \int_0^\infty L(u, v \mid \mathbf{y})\, \pi(u, v)\, du\, dv} \propto L(\mu, \theta \mid \mathbf{y})\, \pi(\mu, \theta), \qquad \mu > 0,\ \theta > 0. \tag{2.4}$$

The marginal posterior density of μ is

$$p(\mu \mid \mathbf{y}) = \int_0^\infty p(\mu, \theta \mid \mathbf{y})\, d\theta. \tag{2.5}$$

The normalizing constant and marginal posterior quantiles are not available in closed form. Posterior computation is therefore performed in Stan using Hamiltonian Monte Carlo with the No-U-Turn Sampler [27]. Sampling is conducted in terms of

$$\eta_\mu = \log \mu, \qquad \eta_\theta = \log \theta,$$

with

$$\mu = e^{\eta_\mu}, \qquad \theta = e^{\eta_\theta}.$$

For each simulated or observed dataset, four parallel chains are run. Each chain uses 1000 warmup iterations followed by 1000 retained iterations, giving 4000 post-warmup draws. The sampler settings are

$$\texttt{adapt\_delta} = 0.99, \qquad \texttt{max\_treedepth} = 14.$$

Convergence and sampling quality are assessed using rank-normalized $\widehat{R}$, bulk effective sample size, tail effective sample size, Monte Carlo standard errors, divergent transitions, maximum-treedepth events, and the energy Bayesian fraction of missing information [28, 29]. A fit is flagged when any monitored parameter has

$$\widehat{R} > 1.01, \qquad \mathrm{ESS}_{\mathrm{bulk}} < 400, \qquad \mathrm{ESS}_{\mathrm{tail}} < 400,$$

or when the fit has a divergent transition, reaches the maximum tree depth, or has

$$\text{E-BFMI} < 0.30.$$

Trace plots, autocorrelation plots, posterior predictive bar plots, and posterior predictive empirical distribution plots are used as graphical diagnostics.

2.3. Credible interval for μ

Let $F_{\mu|\mathbf{y}}$ denote the cumulative distribution function of the marginal posterior distribution in (2.5). In this study, we focus on the equal-tailed Bayesian credible interval for μ, since it is directly comparable to the two-sided frequentist confidence intervals used in the simulation study. A $(1-\alpha)$ equal-tailed credible interval for μ is

$$\mathrm{CrI}^{\mathrm{ET}}_{1-\alpha}(\mu) = \left[F^{-1}_{\mu|\mathbf{y}}\left(\frac{\alpha}{2}\right),\ F^{-1}_{\mu|\mathbf{y}}\left(1-\frac{\alpha}{2}\right)\right], \tag{2.6}$$

so that

$$\Pr\left(\mu \in \mathrm{CrI}^{\mathrm{ET}}_{1-\alpha}(\mu) \mid \mathbf{y}\right) = 1-\alpha.$$

In practice, the endpoints in (2.6) are estimated from the empirical quantiles of the MCMC posterior draws of μ. The resulting interval is evaluated in the simulation study using coverage probability and expected width.

2.4. Bayesian prediction interval

2.4.1. Single future observation

Let $\tilde{Y}$ be a single future observation from the same NB2 model. Given (μ, θ),

$$\tilde{Y} \mid \mu, \theta \sim \mathrm{NB2}(\mu, \theta).$$

The posterior predictive probability mass function of $\tilde{Y}$ is

$$p(\tilde{y} \mid \mathbf{y}) = \Pr(\tilde{Y} = \tilde{y} \mid \mathbf{y}) = \int_0^\infty \int_0^\infty \Pr(\tilde{Y} = \tilde{y} \mid \mu, \theta) p(\mu, \theta \mid \mathbf{y})\, d\mu\, d\theta, \qquad \tilde{y} = 0, 1, 2, \ldots, \tag{2.7}$$

where $\Pr(\tilde{Y} = \tilde{y} \mid \mu, \theta)$ is given by (2.1).

A $(1-\alpha)$ Bayesian prediction interval for $\tilde{Y}$ is any set $\mathcal{I}_{1-\alpha}(\tilde{Y}) \subseteq \{0, 1, 2, \ldots\}$ such that

$$\Pr\left(\tilde{Y} \in \mathcal{I}_{1-\alpha}(\tilde{Y}) \mid \mathbf{y}\right) \geq 1-\alpha. \tag{2.8}$$

Because $\tilde{Y}$ is discrete, the achieved posterior predictive probability may be slightly larger than the nominal level. A common equal-tailed prediction interval is defined using posterior predictive quantiles:

$$\mathcal{I}^{\mathrm{ET}}_{1-\alpha}(\tilde{Y}) = \left[Q_{\alpha/2}\left(\tilde{Y} \mid \mathbf{y}\right),\ Q_{1-\alpha/2}\left(\tilde{Y} \mid \mathbf{y}\right)\right], \tag{2.9}$$

where $Q_p(\tilde{Y} \mid \mathbf{y})$ is the smallest integer q such that $\Pr(\tilde{Y} \leq q \mid \mathbf{y}) \geq p$.

2.4.2. Future mean of m observations

Let $\tilde{Y}_1, \ldots, \tilde{Y}_m$ be conditionally independent future observations given (μ, θ):

$$\tilde{Y}_j \mid \mu, \theta \overset{\text{ind}}{\sim} \text{NB2}(\mu, \theta), \qquad j = 1, \ldots, m.$$

Define the future sample mean

$$\overline{\tilde{Y}} = \frac{1}{m} \sum_{j=1}^{m} \tilde{Y}_j.$$

Equivalently, letting $\tilde{S} = \sum_{j=1}^{m} \tilde{Y}_j$, we have

$$\tilde{S} \mid \mu, \theta \sim \text{NB2}(m\mu, m\theta), \qquad \overline{\tilde{Y}} = \frac{\tilde{S}}{m}.$$

The posterior predictive distribution of $\overline{\tilde{Y}}$ is

$$p(\overline{\tilde{Y}} \mid \mathbf{y}) = \int_0^\infty \int_0^\infty p(\overline{\tilde{Y}} \mid \mu, \theta) p(\mu, \theta \mid \mathbf{y}) \, d\mu \, d\theta. \tag{2.10}$$

A $(1 - \alpha)$ posterior predictive interval for $\overline{\tilde{Y}}$ is an interval $[L, U]$ such that

$$\Pr\left(L \le \overline{\tilde{Y}} \le U \mid \mathbf{y}\right) \ge 1 - \alpha.$$

The equal-tailed interval is

$$\text{PPI}_{1-\alpha}^{\text{ET}}(\overline{\tilde{Y}}) = \left[F^{-1}_{\overline{\tilde{Y}}|\mathbf{y}}\left(\frac{\alpha}{2}\right), \; F^{-1}_{\overline{\tilde{Y}}|\mathbf{y}}\left(1 - \frac{\alpha}{2}\right)\right], \tag{2.11}$$

where $F_{\overline{\tilde{Y}}|\mathbf{y}}$ is the posterior predictive CDF induced by (2.10).

2.5. Evaluation under likelihood misspecification

Robust Bayesian inference under misspecified high-dimensional linear regression models was investigated by Fan et al. [30, 31]. Those results do not establish robustness for the present NB2 count model because the likelihood, parameterization, and inferential targets differ. Zero inflation and structured zeros are also known to require explicit attention in Bayesian count modeling [19, 20]. We therefore evaluate likelihood misspecification directly rather than extrapolating robustness results from linear regression.

The fitted model remains the NB2 model in every scenario. Four data-generating mechanisms are considered.

First, the correctly specified mechanism is

$$Y_i \overset{\text{iid}}{\sim} \text{NB2}(\mu, \theta).$$

Second, the zero-inflated NB2 mechanism is

$$Y_i = (1 - Z_i)X_i, \qquad Z_i \sim \text{Bernoulli}(0.20), \qquad X_i \sim \text{NB2}\left(\frac{\mu}{0.80}, \theta\right),$$

with Z_i independent of X_i. This construction preserves

$$\mathrm{E}(Y_i) = \mu.$$

Third, the structured-zero mechanism assigns half of the observations to a low-zero group and half to a high-zero group:

$$p_i = \begin{cases} 0.05, & i \text{ in the low-zero group,} \\ 0.35, & i \text{ in the high-zero group.} \end{cases}$$

For the balanced design,

$$\frac{1}{n}\sum_{i=1}^{n} p_i = 0.20.$$

Counts are generated from

$$Y_i = (1 - Z_i)X_i, \qquad Z_i \sim \mathrm{Bernoulli}(p_i), \qquad X_i \sim \mathrm{NB2}\left(\frac{\mu}{0.80}, \theta\right).$$

All sample sizes used in the study are even, so the two groups have equal sizes.

Fourth, the Poisson-lognormal mechanism is

$$Y_i \mid \lambda_i \sim \mathrm{Poisson}(\lambda_i), \qquad \log \lambda_i \sim N\left(\log \mu - \frac{\sigma_\lambda^2}{2}, \sigma_\lambda^2\right),$$

where

$$\sigma_\lambda^2 = \log\left(1 + \frac{1}{\theta}\right).$$

This gives

$$\mathrm{E}(Y_i) = \mu, \qquad \mathrm{Var}(Y_i) = \mu + \frac{\mu^2}{\theta},$$

which matches the first two moments of $\mathrm{NB2}(\mu, \theta)$ while changing the full distribution.

For every misspecification scenario, future outcomes are generated from the same mechanism as the observed sample. Coverage therefore evaluates whether an interval constructed under the fitted NB2 model contains a future quantity generated from the actual data-generating mechanism. These experiments measure sensitivity to specified departures from the NB2 likelihood. They do not establish general robustness against arbitrary model misspecification.

2.6. Comparison with frequentist intervals

The Bayesian intervals are compared with Wald, MLE-based normal, and parametric bootstrap procedures. The term "MLE-based normal" is used because the relevant intervals employ a Gaussian approximation with an NB2 variance evaluated at the maximum-likelihood estimates. They are not obtained by inverting a likelihood-ratio test or profiling the likelihood.

Let $\bar{Y}$ and S^2 denote the sample mean and variance. The Wald confidence interval [32] for μ is

$$\left[\max\left\{0, \bar{Y} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right\}, \ \bar{Y} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right].$$

Under the fitted NB2 model,

$$\widehat{\mu} = \bar{Y}, \qquad \widehat{\sigma}^2_{\mathrm{NB2}} = \widehat{\mu} + \frac{\widehat{\mu}^2}{\widehat{\theta}},$$

where $\widehat{\theta}$ is the maximum-likelihood estimate of the dispersion parameter. The MLE-based normal confidence interval is

$$\left[\max\left\{0, \widehat{\mu} - z_{1-\alpha/2}\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{NB2}}}{n}}\right\},\ \widehat{\mu} + z_{1-\alpha/2}\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{NB2}}}{n}}\right].$$

The parametric bootstrap confidence interval [33] is a percentile interval. For each bootstrap replication, a sample of size n is generated from

$$\mathrm{NB2}(\widehat{\mu}, \widehat{\theta}),$$

the NB2 model is refitted, and the bootstrap estimate of μ is retained. The interval endpoints are the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of the retained estimates.

For the future mean

$$\widetilde{\bar{Y}}_m = \frac{1}{m}\sum_{j=1}^{m} \widetilde{Y}_j,$$

the Wald prediction interval is

$$\bar{Y} \pm z_{1-\alpha/2} S \sqrt{\frac{1}{n} + \frac{1}{m}}.$$

The MLE-based normal prediction interval is

$$\widehat{\mu} \pm z_{1-\alpha/2}\sqrt{\widehat{\sigma}^2_{\mathrm{NB2}}\left(\frac{1}{n} + \frac{1}{m}\right)}.$$

The lower endpoint is truncated at zero. For $m = 1$, endpoints are rounded outward to integers. For $m > 1$, endpoints are rounded outward to the support

$$\left\{0, \frac{1}{m}, \frac{2}{m}, \ldots\right\}.$$

For each parametric bootstrap prediction replication, an NB2 sample of size n is generated and refitted. A future sample of size m is then generated from the bootstrap-fitted model, and its mean is retained. The prediction endpoints are the type-1 empirical quantiles of the resulting future means, followed by outward rounding to the appropriate support.

Algorithm 1 Simulation algorithm for NB2 interval comparisons

1) **Specify the design.** Choose (μ_0, θ_0, n), the data-generating mechanism, future sample sizes m, nominal level $1 - \alpha$, Monte Carlo replications R, retained posterior draws S, and the prior specification.

2) **For each scenario and replication** $r = 1, \ldots, R$**:**

   (a) Generate the observed sample from the specified mechanism: NB2, zero-inflated NB2, structured-zero, or Poisson-lognormal.

   (b) Fit the NB2 model using NUTS. Retain $\{(\mu^{(s)}, \theta^{(s)})\}_{s=1}^{S}$ and record MCMC diagnostics and computational failures.

   (c) Construct the Bayesian credible interval for μ and posterior predictive intervals for a single future count and the future mean $\overline{\overline{Y}}_m$.

   (d) Construct the corresponding Wald, MLE-based normal, and parametric bootstrap confidence and prediction intervals. Round prediction endpoints outward to the appropriate support.

   (e) Generate the true future quantity from the same mechanism as the observed sample. Record interval endpoints, coverage, width, failures, and boundary estimates.

3) **Summarize performance.** Let

$$\mathcal{S} = \{r : F^{(r)} = 0\}, \qquad R_{\text{suc}} = |\mathcal{S}|$$

where $F^{(r)}$ indicates interval failure. For successful replications, let $I^{(r)}$ and $\ell^{(r)}$ denote the coverage indicator and interval width. Compute

$$\widehat{f} = 1 - \frac{R_{\text{suc}}}{R},$$
$$\widehat{\text{CP}}_{\text{op}} = \frac{1}{R} \sum_{r \in \mathcal{S}} I^{(r)},$$
$$\widehat{\text{CP}}_{\text{suc}} = \frac{1}{R_{\text{suc}}} \sum_{r \in \mathcal{S}} I^{(r)},$$
$$\widehat{\text{EW}}_{\text{suc}} = \frac{1}{R_{\text{suc}}} \sum_{r \in \mathcal{S}} \ell^{(r)}.$$

Thus, $\widehat{\text{CP}}_{\text{op}}$ counts failures as noncoverage, whereas $\widehat{\text{CP}}_{\text{suc}}$ conditions on successful interval construction.

4) **Report results.** Report both coverage measures, expected width, Monte Carlo standard errors, method-specific failure rates, boundary-estimate rates, and MCMC diagnostics.

Note: CP denotes coverage probability, EW denotes expected width, and R_{suc} denotes the number of successful replications.

3. Simulation studies

In this section, we evaluate the finite-sample operating characteristics of the Bayesian credible and posterior predictive intervals for the NB2 model and compare them with three frequentist competitors: Wald intervals, MLE-based normal intervals, and parametric bootstrap intervals. Throughout the tables, CP denotes operational coverage probability and EW denotes average interval width. Operational coverage counts fail interval constructions as noncoverage, while EW is computed among successful interval constructions. This convention gives a failure-aware comparison of the methods.

The simulation study is organized into four parts. First, correctly specified NB2 data-generating mechanisms are used to compare interval estimation for the mean parameter μ. Second, posterior predictive and frequentist prediction intervals are compared for a single future count and for future sample means. Third, robustness is evaluated under likelihood misspecification. Fourth, the prior-family sensitivity analysis and MCMC diagnostics are reported. To keep the main text readable, Tables 1–3 display the representative $\mu_0 = 1$ case. The remaining mean settings are reported in Appendix Tables S1–S3. A graphical summary of the operational coverage probabilities reported in Table 1 is provided in Figure S1. Additional graphical summaries of the likelihood-misspecification results are provided in Figures S2–S4.

3.1. Correctly specified NB2 simulations

For the correctly specified simulations, data are generated from the NB2 model using combinations of $n \in \{20, 50, 100, 200\}$, $\mu_0 \in \{1, 3, 5, 10\}$, and $\theta_0 \in \{0.1, 1, 4, 10\}$. The values of θ_0 range from severe overdispersion ($\theta_0 = 0.1$) to comparatively mild overdispersion ($\theta_0 = 10$). Table 1 presents the representative low-mean setting $\mu_0 = 1$, while the remaining mean settings are reported in Appendix Table S1. Figure S1 provides a graphical summary of the operational coverage probabilities reported in Table 1.

For $\mu_0 = 1$, Table 1 and Figure S1 show that the Bayesian credible interval generally gives coverage close to or above the nominal level across the correctly specified NB2 settings, although it can be wider than the frequentist intervals when the sample size is small and overdispersion is severe. This behavior is expected because the posterior must account for uncertainty in the mean and dispersion parameters. The Wald CI improves with increasing n, but it undercovers in the most difficult high-overdispersion cases. The MLE-based normal and bootstrap CIs are competitive in the more regular cells but can undercover when θ_0 is small, reflecting sensitivity to dispersion estimation in highly overdispersed samples.

Table 1. Coverage probability (CP) and expected width (EW) of 95% intervals for the NB2 mean μ under correctly specified NB2 data generation for the representative case $\mu_0 = 1$. Entries are CP (EW).

n	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
20	Bayesian CrI	0.930 (7.43)	0.949 (1.47)	0.965 (1.10)	0.971 (1.04)
	Wald CI	0.746 (2.11)	0.896 (1.18)	0.922 (0.96)	0.926 (0.90)
	MLE-based normal CI	0.709 (2.13)	0.872 (1.00)	0.929 (0.91)	0.943 (0.90)
	Bootstrap CI	0.728 (2.38)	0.883 (0.99)	0.936 (0.90)	0.950 (0.89)
50	Bayesian CrI	0.949 (2.89)	0.932 (0.82)	0.946 (0.64)	0.961 (0.61)
	Wald CI	0.835 (1.62)	0.929 (0.77)	0.939 (0.61)	0.939 (0.58)
	MLE-based normal CI	0.647 (1.16)	0.896 (0.64)	0.941 (0.58)	0.945 (0.56)
	Bootstrap CI	0.678 (1.15)	0.897 (0.64)	0.944 (0.57)	0.948 (0.56)
100	Bayesian CrI	0.957 (1.63)	0.945 (0.56)	0.951 (0.44)	0.954 (0.42)
	Wald CI	0.882 (1.22)	0.934 (0.55)	0.945 (0.44)	0.943 (0.41)
	MLE-based normal CI	0.672 (0.83)	0.877 (0.46)	0.925 (0.41)	0.936 (0.40)
	Bootstrap CI	0.688 (0.82)	0.880 (0.45)	0.929 (0.40)	0.939 (0.39)
200	Bayesian CrI	0.944 (1.02)	0.952 (0.40)	0.946 (0.31)	0.945 (0.29)
	Wald CI	0.911 (0.89)	0.946 (0.39)	0.947 (0.31)	0.949 (0.29)
	MLE-based normal CI	0.696 (0.59)	0.894 (0.32)	0.936 (0.29)	0.944 (0.28)
	Bootstrap CI	0.701 (0.58)	0.890 (0.32)	0.931 (0.28)	0.941 (0.28)

Note: CP is operational coverage, with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. The method labelled as likelihood CI in the raw frequentist output is reported here as the MLE-based normal CI. Bayesian rows use the updated Bayesian simulation results.

Prediction is more demanding than estimation of the mean because future sampling variation must be included in addition to parameter uncertainty. Table 2 reports the results for $\mu_0 = 1$. Row $m = 1$ corresponds to prediction of a single future count, whereas $m = 5, 10, 20, 30$ correspond to prediction of the future sample mean $\overline{Y}_{m,\text{new}}$.

Table 2. Coverage probability (CP) and expected width (EW) of 95% prediction intervals under correctly specified NB2 data generation for $\mu_0 = 1$. Entries are CP (EW). The case $m = 1$ corresponds to prediction of a single future count, whereas $m > 1$ corresponds to prediction of the future sample mean $\overline{Y}_{m,\text{new}}$.

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
			$\mu_0 = 1$			
20	1	Bayesian PI	0.955 (12.81)	0.977 (5.30)	0.987 (4.19)	0.992 (4.02)
		Wald PI	0.942 (6.78)	0.967 (4.21)	0.980 (3.70)	0.984 (3.57)
		MLE-normal PI	0.930 (7.20)	0.960 (3.82)	0.979 (3.64)	0.987 (3.61)
		Bootstrap PI	0.928 (7.15)	0.956 (3.65)	0.975 (3.52)	0.984 (3.49)
	5	Bayesian PPI	0.951 (10.26)	0.965 (3.05)	0.974 (2.38)	0.981 (2.26)
		Wald PI	0.879 (3.68)	0.949 (2.42)	0.967 (2.16)	0.972 (2.08)
		MLE-normal PI	0.848 (3.89)	0.931 (2.18)	0.963 (2.10)	0.975 (2.08)
		Bootstrap PI	0.837 (5.06)	0.905 (2.10)	0.953 (1.99)	0.963 (1.97)

Continued on next page

Table 2 continued.

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
	10	Bayesian PPI	0.952 (9.32)	0.963 (2.44)	0.972 (1.87)	0.976 (1.77)
		Wald PI	0.842 (3.05)	0.937 (2.03)	0.956 (1.74)	0.958 (1.65)
		MLE-normal PI	0.793 (3.18)	0.907 (1.76)	0.953 (1.66)	0.963 (1.64)
		Bootstrap PI	0.785 (4.18)	0.888 (1.65)	0.943 (1.55)	0.955 (1.54)
	20	Bayesian PPI	0.942 (8.58)	0.954 (2.02)	0.962 (1.54)	0.973 (1.46)
		Wald PI	0.810 (2.65)	0.920 (1.70)	0.946 (1.40)	0.949 (1.33)
		MLE-normal PI	0.757 (2.73)	0.880 (1.45)	0.942 (1.34)	0.957 (1.32)
		Bootstrap PI	0.758 (3.53)	0.868 (1.37)	0.934 (1.27)	0.952 (1.26)
	30	Bayesian PPI	0.936 (8.26)	0.945 (1.86)	0.966 (1.41)	0.972 (1.34)
		Wald PI	0.794 (2.50)	0.913 (1.55)	0.936 (1.27)	0.941 (1.20)
		MLE-normal PI	0.747 (2.56)	0.877 (1.32)	0.935 (1.21)	0.952 (1.19)
		Bootstrap PI	0.757 (3.24)	0.871 (1.26)	0.931 (1.17)	0.949 (1.15)
50	1	Bayesian PI	0.975 (11.51)	0.977 (4.92)	0.986 (3.89)	0.989 (3.74)
		Wald PI	0.958 (7.36)	0.970 (4.24)	0.985 (3.68)	0.986 (3.54)
		MLE-normal PI	0.943 (5.68)	0.959 (3.73)	0.980 (3.49)	0.985 (3.47)
		Bootstrap PI	0.936 (4.74)	0.955 (3.60)	0.979 (3.42)	0.983 (3.39)
	5	Bayesian PPI	0.978 (7.18)	0.970 (2.61)	0.968 (2.07)	0.982 (1.99)
		Wald PI	0.920 (3.82)	0.963 (2.38)	0.975 (2.11)	0.977 (2.04)
		MLE-normal PI	0.821 (3.00)	0.945 (2.16)	0.972 (2.05)	0.978 (2.03)
		Bootstrap PI	0.750 (3.00)	0.909 (1.96)	0.955 (1.85)	0.965 (1.83)
	10	Bayesian PPI	0.970 (5.79)	0.954 (1.96)	0.973 (1.55)	0.967 (1.49)
		Wald PI	0.906 (3.06)	0.958 (1.94)	0.966 (1.60)	0.971 (1.51)
		MLE-normal PI	0.746 (2.32)	0.926 (1.65)	0.959 (1.50)	0.970 (1.48)
		Bootstrap PI	0.716 (2.39)	0.897 (1.47)	0.943 (1.38)	0.958 (1.37)
	20	Bayesian PPI	0.956 (4.74)	0.953 (1.51)	0.962 (1.19)	0.968 (1.14)
		Wald PI	0.893 (2.56)	0.951 (1.49)	0.958 (1.20)	0.962 (1.13)
		MLE-normal PI	0.716 (1.87)	0.909 (1.25)	0.950 (1.12)	0.962 (1.10)
		Bootstrap PI	0.702 (1.94)	0.885 (1.14)	0.940 (1.06)	0.953 (1.05)
	30	Bayesian PPI	0.950 (4.27)	0.952 (1.32)	0.965 (1.04)	0.964 (1.00)
		Wald PI	0.882 (2.35)	0.946 (1.29)	0.958 (1.04)	0.957 (0.98)
		MLE-normal PI	0.702 (1.69)	0.903 (1.08)	0.949 (0.97)	0.957 (0.95)
		Bootstrap PI	0.693 (1.72)	0.884 (1.01)	0.939 (0.93)	0.949 (0.92)
100	1	Bayesian PI	0.975 (11.00)	0.980 (4.86)	0.988 (3.79)	0.988 (3.62)
		Wald PI	0.961 (7.63)	0.972 (4.25)	0.984 (3.74)	0.986 (3.56)
		MLE-normal PI	0.945 (5.68)	0.958 (3.78)	0.979 (3.51)	0.985 (3.46)
		Bootstrap PI	0.938 (4.74)	0.955 (3.60)	0.975 (3.38)	0.982 (3.33)
	5	Bayesian PPI	0.976 (6.18)	0.966 (2.52)	0.965 (1.97)	0.974 (1.89)
		Wald PI	0.933 (3.89)	0.963 (2.36)	0.978 (2.10)	0.982 (2.03)
		MLE-normal PI	0.833 (2.96)	0.943 (2.14)	0.972 (2.03)	0.982 (2.01)
		Bootstrap PI	0.751 (2.86)	0.906 (1.92)	0.955 (1.81)	0.968 (1.79)
	10	Bayesian PPI	0.978 (4.73)	0.959 (1.84)	0.969 (1.45)	0.970 (1.38)
		Wald PI	0.927 (3.07)	0.962 (1.90)	0.971 (1.55)	0.973 (1.46)
		MLE-normal PI	0.751 (2.26)	0.924 (1.61)	0.963 (1.45)	0.972 (1.42)
		Bootstrap PI	0.717 (2.22)	0.890 (1.41)	0.944 (1.32)	0.957 (1.30)
	20	Bayesian PPI	0.962 (3.67)	0.958 (1.37)	0.960 (1.07)	0.962 (1.03)
		Wald PI	0.924 (2.52)	0.952 (1.39)	0.963 (1.12)	0.965 (1.05)
		MLE-normal PI	0.725 (1.78)	0.907 (1.18)	0.953 (1.05)	0.963 (1.03)
		Bootstrap PI	0.695 (1.72)	0.880 (1.06)	0.940 (0.98)	0.951 (0.97)
	30	Bayesian PPI	0.959 (3.18)	0.950 (1.17)	0.971 (0.91)	0.961 (0.88)
		Wald PI	0.917 (2.28)	0.959 (1.18)	0.958 (0.94)	0.960 (0.88)
		MLE-normal PI	0.714 (1.58)	0.912 (0.99)	0.947 (0.88)	0.959 (0.86)
		Bootstrap PI	0.695 (1.49)	0.889 (0.90)	0.933 (0.83)	0.949 (0.82)
200	1	Bayesian PI	0.982 (10.61)	0.985 (4.86)	0.983 (3.78)	0.984 (3.56)
		Wald PI	0.964 (7.78)	0.972 (4.23)	0.986 (3.81)	0.987 (3.62)
		MLE-normal PI	0.948 (5.70)	0.960 (3.77)	0.980 (3.54)	0.984 (3.48)
		Bootstrap PI	0.939 (4.71)	0.954 (3.57)	0.974 (3.32)	0.981 (3.29)
	5	Bayesian PPI	0.974 (5.68)	0.973 (2.49)	0.966 (1.92)	0.970 (1.85)
		Wald PI	0.941 (3.94)	0.971 (2.35)	0.980 (2.09)	0.981 (2.03)
		MLE-normal PI	0.847 (2.96)	0.951 (2.13)	0.973 (2.02)	0.979 (2.00)
		Bootstrap PI	0.740 (2.82)	0.909 (1.89)	0.956 (1.78)	0.967 (1.77)
	10	Bayesian PPI	0.982 (4.24)	0.964 (1.79)	0.962 (1.40)	0.967 (1.34)
		Wald PI	0.938 (3.08)	0.970 (1.88)	0.973 (1.52)	0.976 (1.43)
		MLE-normal PI	0.756 (2.24)	0.937 (1.58)	0.963 (1.42)	0.972 (1.39)
		Bootstrap PI	0.718 (2.15)	0.899 (1.37)	0.944 (1.28)	0.958 (1.27)

Continued on next page

Table 2 continued.

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
	20	Bayesian PPI	0.956 (3.19)	0.953 (1.30)	0.959 (1.02)	0.964 (0.97)
		Wald PI	0.940 (2.49)	0.963 (1.34)	0.970 (1.08)	0.968 (1.01)
		MLE-normal PI	0.727 (1.74)	0.916 (1.13)	0.958 (1.01)	0.964 (0.99)
		Bootstrap PI	0.688 (1.62)	0.887 (1.00)	0.942 (0.93)	0.953 (0.93)
	30	Bayesian PPI	0.959 (2.70)	0.951 (1.09)	0.959 (0.85)	0.963 (0.81)
		Wald PI	0.935 (2.24)	0.962 (1.11)	0.964 (0.89)	0.963 (0.84)
		MLE-normal PI	0.715 (1.53)	0.910 (0.93)	0.950 (0.83)	0.960 (0.81)
		Bootstrap PI	0.684 (1.38)	0.885 (0.84)	0.935 (0.78)	0.948 (0.77)

Note: CP is operational coverage, with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. MLE-normal PI denotes the MLE-based normal prediction interval. Bayesian rows use posterior predictive intervals from the updated Bayesian simulation results.

The prediction results show the expected decrease in width as m increases, because averaging future counts reduces future-sample variability. For $\mu_0 = 1$, Bayesian posterior predictive intervals tend to maintain coverage well in the severe-overdispersion settings, often at the cost of larger width. The normal-approximation and bootstrap prediction intervals are shorter in many cells but can lose coverage in the highly overdispersed cases, especially for future means where the interval width becomes narrower.

3.2. Likelihood misspecification

To examine robustness beyond the fitted NB2 likelihood, the same NB2 model is fitted to data generated from three misspecified mechanisms: a Poisson–lognormal mixture, a zero-inflated NB2 model, and a structured-zero mechanism. These experiments are not intended to show that the NB2 model is correct under misspecification. Instead, they quantify how the Bayesian and frequentist NB2 interval procedures behave when the likelihood is only an approximation to the data-generating process. Table 3 summarizes the representative $\mu_0 = 1$ case. The remaining misspecification mean settings, $\mu_0 = 5$ and $\mu_0 = 10$, are reported in Appendix Table S3. Figures S2–S4 provide graphical summaries of the operational coverage probabilities for the three representative (n, θ_0) settings.

Table 3. Bayesian and frequentist interval performance under model misspecification for the representative case $\mu_0 = 1$. Entries are CP (EW).

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
		Poisson–lognormal		
		$\mu_0 = 1$		
CI for μ	Bayesian CrI	0.878 (3.15)	0.946 (0.54)	0.954 (0.29)
	Wald CI	0.761 (1.79)	0.936 (0.55)	0.947 (0.29)
	MLE-based normal CI	0.725 (1.49)	0.873 (0.44)	0.944 (0.28)
	Bootstrap CI	0.740 (1.51)	0.879 (0.44)	0.941 (0.28)
Y_{new}	Bayesian PI	0.970 (8.81)	0.987 (4.64)	0.984 (3.57)
	Wald PI	0.955 (5.84)	0.973 (4.24)	0.986 (3.61)
	MLE-based normal PI	0.944 (5.05)	0.963 (3.71)	0.984 (3.47)
	Bootstrap PI	0.943 (5.21)	0.960 (3.54)	0.980 (3.29)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.958 (5.58)	0.958 (2.41)	0.973 (1.86)
	Wald PI	0.905 (3.22)	0.964 (2.35)	0.982 (2.03)
	MLE-based normal PI	0.869 (2.81)	0.943 (2.11)	0.979 (2.00)
	Bootstrap PI	0.845 (3.17)	0.911 (1.89)	0.965 (1.77)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.918 (4.71)	0.945 (1.76)	0.973 (1.34)
	Wald PI	0.870 (2.69)	0.960 (1.88)	0.977 (1.43)
	MLE-based normal PI	0.817 (2.32)	0.930 (1.57)	0.973 (1.39)
	Bootstrap PI	0.794 (2.56)	0.898 (1.38)	0.960 (1.27)

Continued on next page

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.894 (4.09)	0.948 (1.31)	0.959 (0.97)
	Wald PI	0.834 (2.35)	0.953 (1.39)	0.968 (1.01)
	MLE-based normal PI	0.766 (1.98)	0.909 (1.14)	0.965 (0.98)
	Bootstrap PI	0.757 (2.14)	0.881 (1.03)	0.953 (0.92)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.880 (3.81)	0.953 (1.12)	0.959 (0.81)
	Wald PI	0.813 (2.20)	0.950 (1.17)	0.962 (0.84)
	MLE-based normal PI	0.740 (1.85)	0.904 (0.96)	0.957 (0.81)
	Bootstrap PI	0.732 (1.96)	0.880 (0.88)	0.948 (0.77)
	ZI–NB2			
	$\mu_0 = 1$			
CI for μ	Bayesian CrI	0.946 (8.78)	0.950 (0.66)	0.941 (0.33)
	Wald CI	0.713 (2.22)	0.931 (0.61)	0.943 (0.32)
	MLE-based normal CI	0.689 (2.39)	0.853 (0.50)	0.921 (0.30)
	Bootstrap CI	0.719 (2.77)	0.860 (0.50)	0.919 (0.29)
Y_{new}	Bayesian PI	0.965 (12.98)	0.981 (5.65)	0.981 (4.03)
	Wald PI	0.945 (7.21)	0.968 (4.58)	0.986 (3.91)
	MLE-based normal PI	0.927 (8.20)	0.955 (4.01)	0.980 (3.62)
	Bootstrap PI	0.922 (7.84)	0.950 (3.81)	0.975 (3.39)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.967 (11.35)	0.973 (2.89)	0.962 (2.06)
	Wald PI	0.873 (3.90)	0.963 (2.50)	0.981 (2.14)
	MLE-based normal PI	0.844 (4.39)	0.938 (2.24)	0.973 (2.05)
	Bootstrap PI	0.837 (5.83)	0.896 (2.05)	0.950 (1.80)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.949 (10.51)	0.961 (2.13)	0.974 (1.50)
	Wald PI	0.836 (3.21)	0.964 (2.05)	0.971 (1.58)
	MLE-based normal PI	0.792 (3.58)	0.917 (1.72)	0.958 (1.47)
	Bootstrap PI	0.783 (4.91)	0.876 (1.51)	0.930 (1.30)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.936 (9.87)	0.953 (1.59)	0.951 (1.09)
	Wald PI	0.785 (2.79)	0.952 (1.55)	0.966 (1.12)
	MLE-based normal PI	0.745 (3.06)	0.893 (1.29)	0.954 (1.04)
	Bootstrap PI	0.749 (4.20)	0.859 (1.13)	0.933 (0.95)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.935 (9.55)	0.959 (1.35)	0.970 (0.91)
	Wald PI	0.769 (2.63)	0.952 (1.31)	0.960 (0.93)
	MLE-based normal PI	0.732 (2.87)	0.891 (1.09)	0.942 (0.86)
	Bootstrap PI	0.743 (3.83)	0.867 (0.97)	0.926 (0.79)
	Structured-zero			
	$\mu_0 = 1$			
CI for μ	Bayesian CrI	0.947 (9.17)	0.952 (0.66)	0.948 (0.33)
	Wald CI	0.723 (2.21)	0.936 (0.61)	0.946 (0.32)
	MLE-based normal CI	0.691 (2.37)	0.854 (0.50)	0.926 (0.30)
	Bootstrap CI	0.720 (2.73)	0.860 (0.50)	0.922 (0.29)
Y_{new}	Bayesian PI	0.975 (13.91)	0.985 (5.63)	0.990 (4.04)
	Wald PI	0.936 (7.16)	0.958 (4.57)	0.984 (3.92)
	MLE-based normal PI	0.920 (8.11)	0.941 (3.99)	0.977 (3.62)
	Bootstrap PI	0.916 (7.80)	0.937 (3.79)	0.967 (3.40)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.965 (11.99)	0.980 (2.89)	0.973 (2.05)
	Wald PI	0.872 (3.87)	0.962 (2.50)	0.978 (2.14)
	MLE-based normal PI	0.843 (4.35)	0.932 (2.23)	0.971 (2.05)
	Bootstrap PI	0.833 (5.79)	0.899 (2.04)	0.951 (1.81)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.955 (11.14)	0.965 (2.12)	0.968 (1.50)
	Wald PI	0.829 (3.19)	0.967 (2.05)	0.973 (1.59)
	MLE-based normal PI	0.782 (3.55)	0.913 (1.71)	0.959 (1.47)
	Bootstrap PI	0.773 (4.87)	0.875 (1.50)	0.935 (1.30)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.952 (10.39)	0.967 (1.58)	0.970 (1.09)
	Wald PI	0.795 (2.77)	0.955 (1.55)	0.972 (1.13)
	MLE-based normal PI	0.746 (3.03)	0.893 (1.28)	0.951 (1.04)
	Bootstrap PI	0.757 (4.12)	0.864 (1.13)	0.934 (0.95)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.950 (10.00)	0.970 (1.35)	0.958 (0.91)
	Wald PI	0.775 (2.61)	0.948 (1.31)	0.962 (0.93)
	MLE-based normal PI	0.738 (2.84)	0.886 (1.08)	0.946 (0.86)
	Bootstrap PI	0.751 (3.78)	0.865 (0.96)	0.925 (0.79)

Note: CP is operational coverage, with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. The methods labelled as likelihood CI/PI in the raw frequentist output are reported here as MLE-based normal CI/PI. Bayesian rows are placed first and use the updated Bayesian simulation results.

The misspecification results should be interpreted as operating characteristics under alternative data-generating mechanisms rather than as evidence of model adequacy. Table 3 and Figures S2–S4 show that performance depends on the data-generating mechanism, inferential target, and (n, θ_0) setting. For $(n, \theta_0) = (20, 0.1)$, the Bayesian intervals generally provide higher coverage than the frequentist procedures, particularly for future sample means, although they can be wider. Undercoverage of the Wald, MLE-based normal, and bootstrap procedures becomes more pronounced as the future sample size increases, especially under the zero-inflated NB2 and structured-zero mechanisms. The differences among the methods generally decrease for $(n, \theta_0) = (100, 1)$ and $(n, \theta_0) = (200, 10)$, although the MLE-based normal and bootstrap procedures continue to show undercoverage for some targets. These findings quantify sensitivity to the specified departures from the NB2 likelihood and do not establish robustness under arbitrary misspecification.

3.3. Prior sensitivity analysis

To address robustness to the prior distribution, we replace the earlier empirical-Bayes and same-family lognormal sensitivity analysis with a prior-family sensitivity analysis. The sensitivity study uses three distinct weakly informative prior families rather than four variants of a single lognormal family:

$$\begin{aligned} P_1 &: \quad \log \mu \sim N(0, 2^2), \qquad \log \theta \sim N(0, 2.5^2), \\ P_2 &: \quad \mu \sim \Gamma(0.2271, 0.03226), \qquad \theta \sim \Gamma(0.1669, 0.01011), \\ P_3 &: \quad \log \mu \sim t_3(0, 1.232), \qquad \log \theta \sim t_3(0, 1.540). \end{aligned}$$

For the Gamma prior, the shape–rate parametrization is used. The Gamma and log-Student-t hyperparameters are calibrated so that their medians are approximately one, and their upper 0.975 quantiles are comparable to those of the baseline lognormal prior. No prior is centered at $\log(\bar{y} + 0.1)$; the data-centered empirical-Bayes specifications are removed so that the simulation compares fixed prior distributions.

Before fitting the model, we summarize the prior predictive implications of each prior using 100,000 prior draws. Table 4 shows that the three prior families have comparable central locations and similar upper prior quantiles for μ and θ, but they induce different tail behavior and different prior predictive zero probabilities. This is intentional because the purpose is to assess sensitivity to prior family, not only to small changes in hyperparameters.

Table 4. Prior predictive summaries based on 100,000 prior draws. Entries for μ and θ are the 0.025 quantile, median, and 0.975 quantile.

Prior	μ quantiles	θ quantiles	$\Pr(Y = 0)$	$Y_{0.95}$	$Y_{0.99}$
P_1 Lognormal	$(0.020, 0.99, 50.75)$	$(0.008, 1.02, 129.04)$	0.576	22	106
P_2 Gamma	$(2.01 \times 10^{-6}, 1.01, 50.09)$	$(1.59 \times 10^{-8}, 1.02, 134.45)$	0.661	27	79
P_3 log-Student-t	$(0.019, 1.00, 49.62)$	$(0.007, 1.00, 139.47)$	0.556	16	181

The simulation part of the prior-family analysis uses nine representative combinations of (μ_0, θ_0, n):

$$\begin{aligned}
\mu_0 = 1: &\quad (1, 0.1, 20),\ (1, 1, 100),\ (1, 10, 200),\\
\mu_0 = 5: &\quad (5, 0.1, 20),\ (5, 1, 100),\ (5, 10, 200),\\
\mu_0 = 10: &\quad (10, 0.1, 20),\ (10, 1, 100),\ (10, 10, 200).
\end{aligned}$$

Thus, the selected scenarios cover small, moderate, and large values of the mean, dispersion, and sample size, while keeping the sensitivity analysis computationally focused. For each scenario and prior family, 1000 Monte Carlo replications are used. Table 5 reports operational coverage probability (CP) and expected width (EW) for the equal-tailed credible interval for μ, the posterior predictive interval for one future count ($m = 1$), and the posterior predictive interval for the future mean with $m = 10$. Failed interval constructions, if any, are counted as noncoverage in the operational CP. In this prior-sensitivity run, all interval procedures are computed successfully, so operational and conditional coverage coincide.

Overall, the major coverage conclusions are reasonably stable across the three prior families for the moderate and regular scenarios. For example, when $n = 100$ and $\theta_0 = 1$, the credible-interval CPs are close to the nominal level for all three priors, and the widths are nearly identical for each value of μ_0. Similar behavior occurs in the $n = 200, \theta_0 = 10$ scenarios. The most visible sensitivity occurs in the difficult small-sample, high-overdispersion settings with $n = 20$ and $\theta_0 = 0.1$. In these cases, the prior family can materially affect interval width and, for the log-Student-t prior at $\mu_0 = 5$ and $\mu_0 = 10$, the credible interval coverage falls below the nominal level. The Gamma prior tends to produce wider intervals in these challenging settings, while the log-Student-t prior can produce shorter intervals and lower coverage for the mean.

Table 5. Prior-family sensitivity analysis for selected NB2 scenarios. Entries are CP (EW).

Scenario (μ_0, θ_0, n)	Prior	CrI for μ	PI ($m = 1$)	PPI ($m = 10$)
$(1, 0.1, 20)$	P_1 Lognormal	0.937 (7.51)	0.953 (12.79)	0.956 (9.36)
	P_2 Gamma	0.936 (15.20)	0.961 (16.58)	0.957 (16.39)
	P_3 log-Student-t	0.937 (4.62)	0.952 (10.93)	0.946 (6.75)
$(1, 1, 100)$	P_1 Lognormal	0.954 (0.56)	0.986 (4.88)	0.964 (1.84)
	P_2 Gamma	0.950 (0.56)	0.986 (4.83)	0.957 (1.83)
	P_3 log-Student-t	0.950 (0.56)	0.985 (4.87)	0.961 (1.84)
$(1, 10, 200)$	P_1 Lognormal	0.956 (0.29)	0.982 (3.55)	0.965 (1.34)
	P_2 Gamma	0.949 (0.29)	0.979 (3.50)	0.967 (1.32)
	P_3 log-Student-t	0.959 (0.30)	0.986 (3.63)	0.970 (1.35)
$(5, 0.1, 20)$	P_1 Lognormal	0.937 (19.96)	0.960 (50.62)	0.932 (30.20)
	P_2 Gamma	0.962 (26.08)	0.971 (61.93)	0.959 (40.04)
	P_3 log-Student-t	0.886 (15.71)	0.956 (43.14)	0.907 (24.30)
$(5, 1, 100)$	P_1 Lognormal	0.951 (2.19)	0.986 (20.11)	0.956 (7.20)
	P_2 Gamma	0.950 (2.21)	0.985 (20.22)	0.957 (7.23)
	P_3 log-Student-t	0.949 (2.17)	0.986 (20.00)	0.957 (7.16)
$(5, 10, 200)$	P_1 Lognormal	0.942 (0.76)	0.964 (10.31)	0.953 (3.47)
	P_2 Gamma	0.940 (0.75)	0.963 (10.26)	0.949 (3.44)
	P_3 log-Student-t	0.940 (0.76)	0.964 (10.35)	0.954 (3.48)
$(10, 0.1, 20)$	P_1 Lognormal	0.940 (36.97)	0.970 (101.92)	0.935 (58.45)
	P_2 Gamma	0.967 (37.84)	0.976 (114.64)	0.958 (65.44)
	P_3 log-Student-t	0.892 (35.52)	0.964 (92.97)	0.912 (53.29)

Continued on next page

Scenario (μ_0, θ_0, n)	Prior	CrI for μ	PI ($m = 1$)	PPI ($m = 10$)
(10, 1, 100)	P_1 Lognormal	0.954 (4.18)	0.977 (38.85)	0.958 (13.76)
	P_2 Gamma	0.956 (4.21)	0.976 (38.96)	0.957 (13.80)
	P_3 log-Student-t	0.955 (4.16)	0.978 (38.71)	0.955 (13.68)
(10, 10, 200)	P_1 Lognormal	0.964 (1.24)	0.960 (17.31)	0.958 (5.70)
	P_2 Gamma	0.965 (1.24)	0.962 (17.29)	0.956 (5.67)
	P_3 log-Student-t	0.961 (1.25)	0.961 (17.40)	0.960 (5.71)

Note: P_1 is the baseline lognormal prior, P_2 is the calibrated Gamma prior, and P_3 is the calibrated log-Student-t prior. CP denotes operational coverage probability and EW denotes expected width.

The sensitivity analysis therefore supports two conclusions. First, the substantive comparison between Bayesian credible or predictive intervals and the frequentist competitors is not driven solely by one lognormal prior choice in the moderate and large-sample settings. Second, prior choice remains important in the most weakly informed data settings, especially when the sample size is small and overdispersion is severe. This qualification is important because the one-sample NB2 model contains an unknown mean and an unknown dispersion parameter, and the dispersion parameter is difficult to estimate reliably in small highly overdispersed samples. Replication-level MCMC and sampler diagnostics are retained for this analysis. Although no interval construction failures occurr, diagnostic flags are more common for the heavy-tailed log-Student-t prior in some scenarios; these flags are reported separately and should be considered when interpreting the most difficult cells.

3.4. MCMC diagnostics

MCMC convergence and numerical accuracy are monitored for all fitted Bayesian NB2 models. The diagnostics are summarized in two blocks. Scenarios 1–64 correspond to the correctly specified NB2 data-generating mechanism, whereas scenarios 65–91 correspond to the likelihood-misspecification experiments. The diagnostic summaries include the Gelman–Rubin statistic $\widehat{R}$, bulk and tail effective sample sizes (ESS), Monte Carlo standard errors (MCSE), divergent transitions, maximum tree-depth hits, and E-BFMI.

Table 6 summarizes the parameter-level MCMC diagnostics for μ and θ. No MCMC fits fail in either block. For the correctly specified NB2 scenarios, the largest observed value of $\widehat{R}$ is 1.015 for μ and 1.018 for θ. For the misspecification scenarios, the corresponding largest values are 1.013 and 1.015. Minimum bulk ESS values remain above 445 for all blocks and parameters, while minimum tail ESS values are smaller for some dispersion parameter fits. Some diagnostic flags are retained, especially for θ, reflecting the known difficulty of estimating the dispersion parameter in small-sample or weakly identified settings. Fits producing numerical diagnostic flags are retained in the operational coverage calculations when interval construction was successful; the flags are recorded and reported separately rather than treated as computational failures.

Table 6. Parameter-level MCMC diagnostics for correctly specified NB2 scenarios 1–64 and misspecification scenarios 65–91.

Scenario block	Parameter	Scenarios	Fits	Failed	Max $\widehat{R}$	$\widehat{R} > 1.01$	Min ESS_{bulk}	Min ESS_{tail}	Max MCSE mean	Diagnostic flags
Scenarios 1–64 (NB2)	μ	64	128,000	0	1.015	25	445	330	3.924	242
Scenarios 1–64 (NB2)	θ	64	128,000	0	1.018	163	524	204	2712.418	242
Scenarios 65–91 (misspecification)	μ	27	27,000	0	1.013	9	597	369	5.347	44
Scenarios 65–91 (misspecification)	θ	27	27,000	0	1.015	19	656	255	411.702	44

Table 7 gives the sampler-level summary for the same two blocks. There are no divergent transitions and no maximum tree-depth hits in either block. The minimum E-BFMI is 0.249 for scenarios 1–64 and 0.329 for scenarios 65–91. These summaries indicate generally satisfactory numerical behavior, although a small number of fits, primarily involving the dispersion parameter in difficult scenarios, does not satisfy all prespecified diagnostic thresholds. These diagnostic results should not be interpreted as evidence that the fitted NB2 likelihood is adequate under every data-generating mechanism.

Because scenarios 65–91 use three different misspecified data-generating mechanisms, Table 8 further separates the misspecification diagnostics by DGM. The fitted model remains the NB2 model in all cases; the table therefore describes computational behavior of the fitted Bayesian NB2 procedure under misspecification, not goodness of fit.

Table 7. Sampler-level MCMC diagnostics for correctly specified NB2 scenarios 1–64 and misspecification scenarios 65–91.

Scenario block	Scenarios	Fits	Failed	Divergent fits	Total divergences	MaxTD fits	Total MaxTD	Min E-BFMI	Fit-level diagnostic flags
Scenarios 1–64 (NB2)	64	128,000	0	0	0	0	0	0.249	242
Scenarios 65–91 (misspecification)	27	27,000	0	0	0	0	0	0.329	44

Table 8. Parameter-level MCMC diagnostics for misspecification scenarios 65–91 by data-generating mechanism.

DGM	Parameter	Scenarios	Fits	Failed	Max $\widehat{R}$	$\widehat{R} > 1.01$	Min $\mathrm{ESS}_{\mathrm{bulk}}$	Min $\mathrm{ESS}_{\mathrm{tail}}$	Diagnostic flags
Poisson–lognormal	μ	9	9,000	0	1.013	3	785	434	17
Poisson–lognormal	θ	9	9,000	0	1.015	9	684	330	17
ZI–NB2	μ	9	9,000	0	1.011	3	597	411	9
ZI–NB2	θ	9	9,000	0	1.011	4	656	348	9
Structured–zero	μ	9	9,000	0	1.013	3	724	369	18
Structured–zero	θ	9	9,000	0	1.014	6	680	255	18

Representative diagnostic plots are shown one at a time rather than in a multi-panel layout. The example shown is scenario 20, replication 1, with $n = 50$, $\mu_0 = 1$, and $\theta_0 = 10$. These plots are visual illustrations; the full numerical assessment is given in Tables 6–8.

The apparent jumps in the $\log(\theta)$ trace reflect movement across a skewed and weakly identified dispersion posterior in this representative count-data setting; however, the numerical diagnostics, including $\widehat{R}$, ESS, divergent transitions, maximum tree-depth hits, and E-BFMI, do not indicate a systematic convergence failure for this fit.

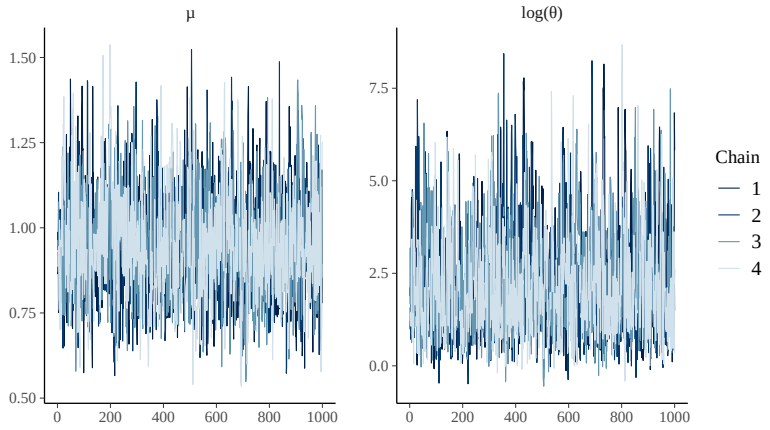


Figure 1. Trace plots for μ and log(θ) for a representative NB2 simulation fit.

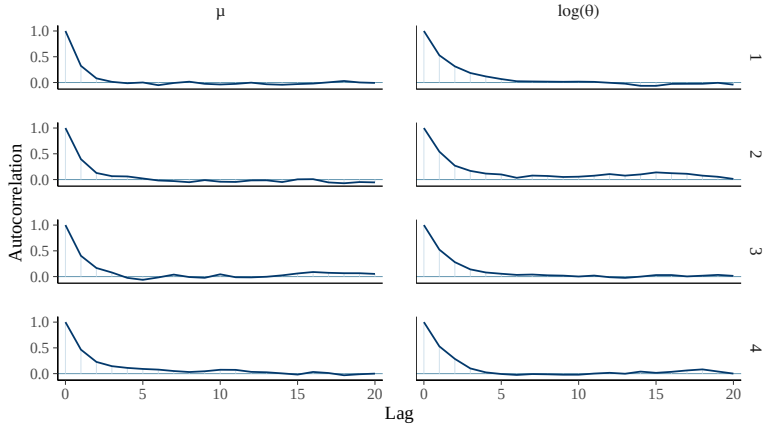


Figure 2. Autocorrelation functions for μ and log(θ) for the same representative NB2 simulation fit.

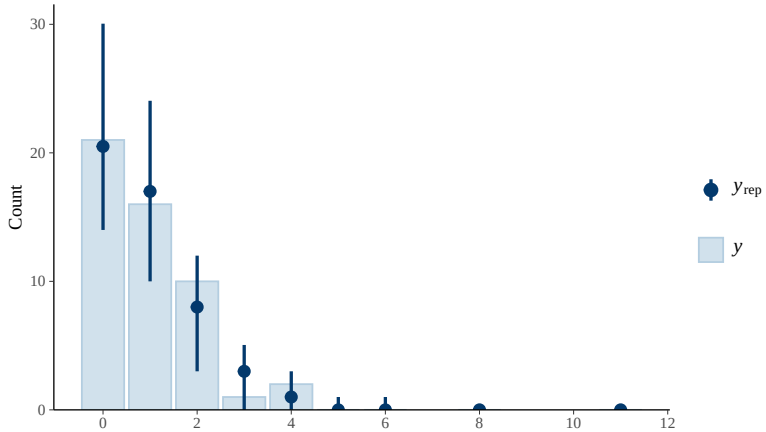


Figure 3. Posterior predictive bar plot comparing observed and replicated counts for the same representative NB2 simulation fit.

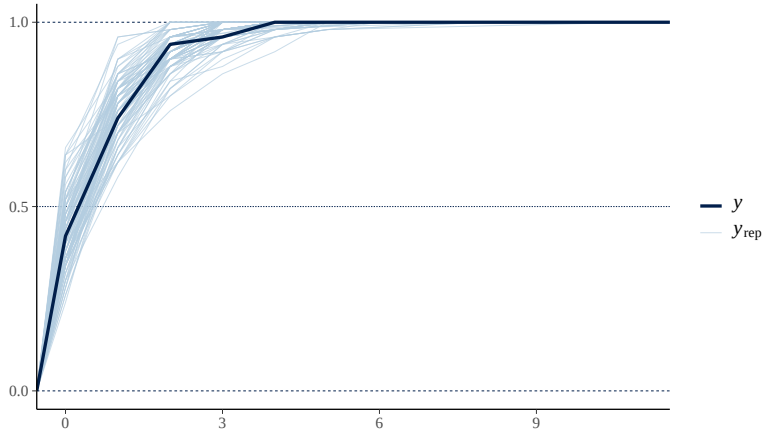


Figure 4. Posterior predictive empirical CDF comparing observed and replicated counts for the same representative NB2 simulation fit.

Note: Diagnostic flags indicate fits that did not satisfy at least one of the numerical diagnostic thresholds used in the simulation workflow. MaxTD denotes maximum tree-depth hits. E-BFMI is reported as the minimum value across the relevant fits. These diagnostics assess numerical behavior of the MCMC algorithm under the specified fitted model; they are not a substitute for posterior predictive model checking or likelihood adequacy assessment.

4. Real data illustrations

The two datasets are used as computational illustrations of the one-sample NB2 interval procedures. Our objective is to show how the Bayesian and frequentist confidence and prediction intervals are implemented and interpreted for overdispersed real count data. These intercept-only analyses are not intended to replace covariate-adjusted or longitudinal models that would be required for subject-matter conclusions.

4.1. Data sources and construction

Both datasets are distributed with the R package `MASS` [34, 35]. The `quine` data contain the number of school days absent (`Days`) for 146 children over a school-year observation window [36]. We analyze the 146 student-level counts without covariates because the target of this illustration is the marginal mean of the one-sample NB2 model for the combined sample. This does not estimate a covariate-adjusted absence rate.

The `epil` data arise from a randomized epilepsy study involving 59 subjects [37]. For each subject, the dataset contains an 8-week pre-randomization seizure count (`base`) and four post-randomization seizure counts, each measured over a 2-week period. In the long-format `MASS::epil` data, the same baseline value is repeated for the four follow-up rows belonging to a subject. We therefore retain exactly one unique baseline value per subject, giving 59 independent subject-level baseline counts. The accompanying code verifies that every subject has one and only one unique baseline value before fitting the model.

We explain why only the epilepsy baseline counts are analyzed.

Restricting the epilepsy illustration to one baseline count per subject is deliberate. First, the one-sample NB2 model assumes independent and identically distributed subject-level counts $Y_1, \ldots, Y_n$. The 59 pre-randomization baseline counts provide one count per subject, share the same 8-week exposure window, and are unaffected by the randomized treatment. Second, treating the four duplicated copies of `base` as 236 observations would constitute pseudoreplication and would artificially understate uncertainty. Third, the post-randomization `y` values are repeated measurements within subjects and therefore do not satisfy the independence assumption of the one-sample model.

A longitudinal analysis could lead to different numerical and substantive conclusions because it addresses a different estimand. For example, the four post-randomization counts could be analyzed using an NB2 generalized linear mixed model with a subject-specific random intercept, or a marginal NB2 model with an explicit within-subject correlation structure. Such a model would normally include treatment, period, treatment-by-period interaction, baseline seizure rate, and age. It could estimate treatment and time effects and would account for within-subject correlation, both of which can change standard errors and predictive uncertainty. If baseline and follow-up counts are modeled jointly, their different 8-week and 2-week exposure durations would also require an exposure adjustment. Consequently, this baseline analysis describes only the marginal distribution of pre-randomization 8-week seizure counts; it does not estimate a treatment effect or make a longitudinal claim. A longitudinal extension is scientifically important, but it lies outside the one-sample NB2 estimand examined here.

Table 9. Descriptive summaries for the two real datasets.

Dataset	n	Mean	Variance	Variance/Mean	Range
Quine absence	146	16.46	264.17	16.05	0–81
Epilepsy baseline	59	31.22	722.38	23.14	6–151

The sample mean and variance are 16.46 and 264.17, respectively, for the Quine counts, and 31.22 and 722.38 for the epilepsy baseline counts. The variance-to-mean ratios are 16.05 and 23.14 for the Quine and epilepsy baseline counts, respectively, which motivate consideration of an overdispersed count model. These ratios show strong departure from a simple Poisson variance assumption, but they should not be interpreted as a goodness-of-fit test for the full NB2 likelihood. The NB2 maximum likelihood estimates $(\widehat{\mu}, \widehat{\theta})$ are $(16.46, 1.07)$ for the Quine data and $(31.22, 2.00)$ for the epilepsy baseline data.

4.2. Estimation and computational details

For each dataset, we fit the intercept-only NB2 model

$$Y_i \mid \mu, \theta \sim \mathrm{NB2}(\mu, \theta), \qquad \mathrm{E}(Y_i \mid \mu, \theta) = \mu, \qquad \mathrm{Var}(Y_i \mid \mu, \theta) = \mu + \frac{\mu^2}{\theta}.$$

The Bayesian analysis uses the priors

$$\log(\mu) \sim N(0, 2^2) \qquad \text{and} \qquad \log(\theta) \sim N(0, 2.5^2).$$

Posterior sampling is performed with the No-U-Turn Sampler through `CmdStanR`, using four chains, 1,000 warm-up iterations per chain, and 5,000 retained iterations per chain, for 20,000 post-warm-up draws in total. The sampler settings are `adapt_delta=` 0.99 and `max_treedepth=` 14. The Bayesian point estimate is the posterior mean of μ, and the 95% credible interval is the equal-tailed interval formed from the 0.025 and 0.975 posterior quantiles. The code also saves the random seeds, analysis settings, CmdStan version, and R session information with the numerical output.

The frequentist comparisons are the sample-variance Wald interval, the MLE-based normal interval, and a parametric-bootstrap percentile interval. For clarity, the interval called "MLE-based normal" uses

$$\widehat{\mu} \pm z_{0.975} \sqrt{\frac{\widehat{\mu} + \widehat{\mu}^2/\widehat{\theta}}{n}}.$$

It is not a profile-likelihood or likelihood-ratio interval. The parametric bootstrap uses 20,000 samples generated from the fitted NB2 model; the NB2 model is refitted to each sample, and the empirical 0.025 and 0.975 quantiles of the bootstrap estimates are reported. Failed bootstrap refits are recorded rather than silently ignored.

Table 10. Interval estimates for the NB2 mean μ for the two real datasets.

Dataset	Method	Estimate	95% interval	Width
Quine absence	Bayesian CrI	16.50	[14.01, 19.39]	5.38
Quine absence	Wald CI	16.46	[13.82, 19.10]	5.27
Quine absence	MLE-based normal CI	16.46	[13.79, 19.13]	5.33
Quine absence	Parametric bootstrap CI	16.46	[13.92, 19.25]	5.34
Epilepsy baseline	Bayesian CrI	31.26	[25.71, 37.80]	12.08
Epilepsy baseline	Wald CI	31.22	[24.36, 38.08]	13.72
Epilepsy baseline	MLE-based normal CI	31.22	[25.40, 37.04]	11.63
Epilepsy baseline	Parametric bootstrap CI	31.22	[25.66, 37.17]	11.51

The last column of Table 10 is the width of the observed interval, calculated as $U - L$ from the unrounded endpoints and then rounded to two decimal places. Consequently, subtracting the displayed two-decimal endpoints can differ from the displayed width by 0.01 because of rounding. The column is labeled "Width," not "expected width," because no repeated-sampling average is involved in a single real-data analysis. For the Quine absence counts, the Bayesian credible interval for μ is $[14.01, 19.39]$, while the Wald, MLE-based normal, and bootstrap intervals are $[13.82, 19.10]$, $[13.79, 19.13]$, and $[13.92, 19.25]$, respectively. For the epilepsy baseline counts, the Bayesian credible interval is $[25.71, 37.80]$, while the corresponding frequentist intervals are $[24.36, 38.08]$, $[25.40, 37.04]$, and $[25.66, 37.17]$. Thus, the intervals are numerically similar and overlap substantially in these two examples, although their interpretations remain different.

Table 11. Parametric bootstrap refit diagnostics for the two real datasets.

Dataset	Requested replications	Successful refits	Failure proportion
Quine absence	20,000	20,000	0.0000
Epilepsy baseline	20,000	20,000	0.0000

The bootstrap diagnostics in Table 11 show that all 20,000 requested parametric bootstrap refits are successful for both datasets. Thus, the bootstrap confidence and prediction intervals in this section are based on the full requested number of bootstrap replications.

4.3. Prediction intervals

Two future targets are considered. The first is a single new count, Y_{new}. The second is the mean of 10 mutually independent future counts from the same one-sample population,

$$\overline{Y}_{10,\text{new}} = \frac{1}{10}\sum_{j=1}^{10} Y_{\text{new},j}.$$

For the Quine analysis, these targets refer to a new child observed over the same school-year window. For the epilepsy analysis, they refer to new pre-randomization 8-week baseline counts from comparable subjects. They do not refer to post-randomization repeated seizure counts.

Let

$$S_{m,\text{new}} = \sum_{j=1}^{m} Y_{\text{new},j}.$$

For a posterior draw $(\mu^{(s)}, \theta^{(s)})$, this future sum satisfies

$$S_{m,\text{new}} \mid \mu^{(s)}, \theta^{(s)} \sim \text{NB2}\left(m\mu^{(s)}, m\theta^{(s)}\right).$$

The Bayesian predictive limits are computed from the posterior mixture CDF,

$$\widehat{F}_{S_m|\mathbf{y}}(k) = \frac{1}{S}\sum_{s=1}^{S} F_{\text{NB}}\left(k;\, m\mu^{(s)}, m\theta^{(s)}\right),$$

where $F_{\text{NB}}(k; a, b)$ denotes the NB2 CDF at k with mean a and size parameter b. The predictive limits for $\overline{Y}_{m,\text{new}}$ are obtained by dividing the corresponding integer quantiles of $S_{m,\text{new}}$ by m. This Rao–Blackwellized calculation avoids the unnecessary Monte Carlo noise that would result from simulating only one future value for each posterior draw.

The Wald and MLE-based normal prediction intervals use, respectively, the sample variance and the fitted NB2 variance in

$$\widehat{\mu} \pm z_{0.975}\widehat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}.$$

The parametric-bootstrap prediction interval is obtained by generating a bootstrap sample from the fitted NB2 model, refitting the model, and simulating the corresponding future quantity from each bootstrap-fitted distribution. Prediction limits are rounded outward to the support of the target: Integers for Y_{new} and multiples of $1/m$ for $\overline{Y}_{m,\text{new}}$.

Table 12. Prediction intervals for the two real datasets.

Dataset	Quantity	Method	Estimated mean	95% interval	Width
Quine absence	Y_{new}	Bayesian PI	16.50	[0.00, 61.00]	61.00
Quine absence	Y_{new}	Wald PI	16.46	[0.00, 49.00]	49.00
Quine absence	Y_{new}	MLE-based normal PI	16.46	[0.00, 49.00]	49.00
Quine absence	Y_{new}	Parametric bootstrap PI	16.46	[0.00, 61.00]	61.00
Quine absence	$\overline{Y}_{10,\text{new}}$	Bayesian PI	16.50	[7.60, 28.80]	21.20
Quine absence	$\overline{Y}_{10,\text{new}}$	Wald PI	16.46	[6.00, 26.90]	20.90
Quine absence	$\overline{Y}_{10,\text{new}}$	MLE-based normal PI	16.46	[5.90, 27.00]	21.10
Quine absence	$\overline{Y}_{10,\text{new}}$	Parametric bootstrap PI	16.46	[7.80, 28.70]	20.90
Epilepsy baseline	Y_{new}	Bayesian PI	31.26	[3.00, 91.00]	88.00
Epilepsy baseline	Y_{new}	Wald PI	31.22	[0.00, 85.00]	85.00
Epilepsy baseline	Y_{new}	MLE-based normal PI	31.22	[0.00, 77.00]	77.00
Epilepsy baseline	Y_{new}	Parametric bootstrap PI	31.22	[3.00, 90.00]	87.00
Epilepsy baseline	$\overline{Y}_{10,\text{new}}$	Bayesian PI	31.26	[17.70, 49.10]	31.40
Epilepsy baseline	$\overline{Y}_{10,\text{new}}$	Wald PI	31.22	[13.20, 49.30]	36.10
Epilepsy baseline	$\overline{Y}_{10,\text{new}}$	MLE-based normal PI	31.22	[15.90, 46.60]	30.70
Epilepsy baseline	$\overline{Y}_{10,\text{new}}$	Parametric bootstrap PI	31.22	[17.90, 48.00]	30.10

In Table 12, the column labeled "Estimated mean" gives the posterior mean of μ for the Bayesian rows and the MLE $\widehat{\mu} = \overline{Y}$ for the frequentist rows. For the Quine data, the Bayesian and bootstrap prediction intervals for a single future absence count are both $[0, 61]$, while the Wald and MLE-based normal intervals are both $[0, 49]$. For the mean of 10 future absence counts, the Bayesian and bootstrap intervals are $[7.60, 28.80]$ and $[7.80, 28.70]$, respectively. For the epilepsy baseline data, the Bayesian prediction interval for a single future baseline count is $[3, 91]$, and the bootstrap interval is $[3, 90]$. For the mean of 10 future baseline counts, the corresponding Bayesian and bootstrap intervals are $[17.70, 49.10]$ and $[17.90, 48.00]$, respectively. Prediction intervals for one future count are much wider than those for the mean of 10 future counts, reflecting both overdispersion and the stabilizing effect of averaging independent future observations.

4.4. MCMC and numerical diagnostics

Convergence is assessed using $\widehat{R}$, bulk and tail effective sample sizes, Monte Carlo standard errors, divergent transitions, maximum tree-depth events, and E-BFMI. We use the following diagnostic criteria: $\widehat{R} \leq 1.01$, bulk and tail effective sample sizes of at least 400, no divergent transitions, no maximum tree-depth events, and minimum E-BFMI of at least 0.30. The final analyses satisfy these criteria. The divergence count, maximum-tree-depth count, and minimum E-BFMI are fit-level quantities and are repeated on the parameter rows only to keep the diagnostic table compact.

Table 13. MCMC diagnostics for the Bayesian NB2 analyses of the two real datasets.

Dataset	Parameter	$\widehat{R}$	$\mathrm{ESS}_{\mathrm{bulk}}$	$\mathrm{ESS}_{\mathrm{tail}}$	MCSE mean	MCSE SD	Divergences	MaxTD	E-BFMI
Quine absence	μ	1.000	10576	10050	0.0135	0.0120	0	0	0.926
Quine absence	θ	1.000	9774	9262	0.0013	0.0012	0	0	0.926
Epilepsy baseline	μ	1.000	9973	8877	0.0311	0.0308	0	0	0.949
Epilepsy baseline	θ	1.000	9509	8985	0.0037	0.0033	0	0	0.949

The MCMC diagnostics in Table 13 support the numerical reliability of the posterior summaries under the specified NB2 model. For both datasets and both monitored parameters, $\widehat{R} = 1.000$, the bulk and tail effective sample sizes are far above 400, there are no divergent transitions or maximum tree-depth events, and the minimum E-BFMI values are 0.926 for the Quine analysis and 0.949 for the epilepsy baseline analysis. These diagnostics indicate that the sampler explored the specified posterior well. They do not, by themselves, prove that the NB2 likelihood is the correct data-generating model.

4.5. Summary and limitations

The examples show that the Bayesian, Wald, MLE-based normal, and parametric bootstrap procedures yield broadly comparable inference for the marginal NB2 mean in these two real datasets. Prediction intervals for individual future counts are substantially wider than intervals for averages of 10 future counts. These conclusions are restricted to the intercept-only one-sample targets defined above. The Quine analysis estimates a marginal mean for the combined student sample and does not adjust for age, sex, ethnicity, or learner-status covariates. The epilepsy analysis concerns only the marginal distribution of pre-randomization 8-week baseline counts. It does not estimate treatment, period, or subject-specific effects. A longitudinal NB2 model is needed for treatment-response conclusions, and such an analysis can produce different estimates and uncertainty because it answers a different scientific question.

5. Concluding remarks

In this paper, we evaluated Bayesian credible and posterior predictive intervals for the two-parameter negative binomial (NB2) model in a one-sample, intercept-only setting. Using weakly informative priors and Hamiltonian Monte Carlo with the No-U-Turn Sampler, we constructed equal-tailed credible intervals for the mean parameter μ and posterior predictive intervals for a single future count and the mean of future observations.

The simulation results show that the Bayesian intervals generally provide coverage close to the nominal level across many correctly specified NB2 scenarios, although they may be wider in small-sample and highly overdispersed settings. Comparisons with Wald, MLE-based normal, and parametric bootstrap confidence and prediction intervals show that the Bayesian procedures are competitive and often more stable in difficult cells. The prior-family sensitivity analysis indicates that conclusions are reasonably stable in moderate and regular settings, but that prior choice can matter in small-sample, highly overdispersed scenarios.

The real-data applications to Quine school-absence counts and epilepsy baseline seizure counts illustrated the practical implementation of the one-sample NB2 interval procedures for overdispersed count data. In both examples, prediction intervals for individual future outcomes were substantially

wider than those for averaged future observations, emphasizing the distinction between uncertainty about the mean and variability in future counts.

A limitation of this study is that we focused on the one-sample NB2 model. The real-data examples were therefore used only as computational illustrations and did not incorporate available covariate information. Extensions to NB2 regression, zero-inflated NB2 models, longitudinal NB2 models, and covariate-adjusted posterior predictive intervals remain important directions for future research. Overall, the results provide a finite-sample comparison of Bayesian and frequentist interval procedures for NB2 mean estimation and prediction in overdispersed count-data analysis.

Use of AI tools declaration

While preparing this work, the authors used the generative pre-trained transformer models to verify grammar. After utilizing this tool/service, the authors reviewed and edited the content as necessary and take full responsibility for the publication's content.

Conflict of interest

The authors declare there are no conflicts of interest.

Data availability statement

The Quine and epilepsy datasets used in the real-data illustrations are distributed with the R package `MASS` [34]. The original data sources are Aitkin [36] for the Quine school-absence data and Thall and Vail [37] for the epilepsy seizure-count data. The simulation code and generated summary tables are available from the corresponding author upon reasonable request.

References

1. F. J. Anscombe, The statistical analysis of insect counts based on the negative binomial distribution, *Biometrics*, **5** (1949), 165–173. https://doi.org/10.2307/3001918
2. G. J. S. Ross, D. A. Preece, The negative binomial distribution, *Statistician*, **34** (1985), 323–336. https://doi.org/10.2307/2987659
3. J. F. Lawless, Negative binomial and mixed Poisson regression, *Can. J. Stat.*, **15** (1987), 209–225. https://doi.org/10.2307/3314912
4. B. J. Park, D. Lord, Adjustment for maximum likelihood estimate of negative binomial dispersion parameter, *Transp. Res. Rec.*, **2061** (2008), 9–19. https://doi.org/10.3141/2061-02
5. G. R. Wood, Confidence and prediction intervals for generalised linear accident models, *Accid. Anal. Prev.*, **37** (2005), 267–273. https://doi.org/10.1016/j.aap.2004.10.005
6. J. Sanchez, Y. He, Internet data analysis for the undergraduate statistics curriculum, *J. Stat. Educ.*, **13** (2005). https://doi.org/10.1080/10691898.2005.11910568
7. M. P. Sormani, L. Bonzano, L. Roccatagliata, G. R. Cutter, G. L. Mancardi, P. Bruzzi, Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: A meta-analytic approach, *Ann. Neurol.*, **65** (2009), 268–275. https://doi.org/10.1002/ana.21606

8. M. M. Hasan, K. Krishnamoorthy, Confidence intervals and prediction intervals for two-parameter negative binomial distributions, *J. Appl. Stat.*, **51** (2024), 2420–2435. https://doi.org/10.1080/02664763.2023.2297157

9. E. C. K. Pagui, A. Salvan, N. Sartori, Improved estimation in negative binomial regression, *Stat. Med.*, **41** (2022), 2403–2416. https://doi.org/10.1002/sim.9361

10. L. J. Willson, J. L. Folks, J. H. Young, Complete sufficiency and maximum likelihood estimation for the two-parameter negative binomial distribution, *Metrika*, **33** (1986), 349–362. https://doi.org/10.1007/BF01894768

11. J. Aragón, D. Eberly, S. Eberly, Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution, *Stat. Probab. Lett.*, **15** (1992), 375–379. https://doi.org/10.1016/0167-7152(92)90157-Z

12. B. A. Dang, K. Krishnamoorthy, Confidence intervals, prediction intervals and tolerance intervals for negative binomial distributions, *Stat. Pap.*, **63** (2022), 795–820. https://doi.org/10.1007/s00362-021-01255-y

13. W. W. Piegorsch, Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics*, **46** (1990), 863–867. https://doi.org/10.2307/2532104

14. U. Bandara, R. Gill, R. Mitra, On computing maximum likelihood estimates for the negative binomial distribution, *Stat. Probab. Lett.*, **148** (2019), 54–58. https://doi.org/10.1016/j.spl.2019.01.009

15. D. Shilane, S. Evans, A. Hubbard, Confidence intervals for negative binomial random variable of high dispersion, *Int. J. Biostat.*, **6** (2010), 28

16. K. Krishnamoorthy, M. M. Hasan, Inference on the difference and ratio of means of two-parameter negative binomial distributions, *J. Stat. Comput. Simul.*, **96** (2026), 1417–1438. https://doi.org/10.1080/00949655.2025.2592084

17. J. Y. Dauxois, P. Druilhet, D. Pommeret, A Bayesian choice between Poisson, binomial and negative binomial models, *Test*, **15** (2006), 423–432. https://doi.org/10.1007/BF02607060

18. S. Fu, A hierarchical Bayesian approach to negative binomial regression, *Methods Appl. Anal.*, **22** (2015), 409–428. https://dx.doi.org/10.4310/MAA.2015.v22.n4.a4

19. Q. He, H. H. Huang, A framework of zero-inflated Bayesian negative binomial regression models for spatiotemporal data, *J. Stat. Plann. Inference*, **229** (2024), 106098. https://doi.org/10.1016/j.jspi.2023.106098

20. Y. Hamura, K. Irie, S. Sugasawa, Robust Bayesian modeling of counts with zero inflation and outliers: Theoretical robustness and efficient computation, *J. Am. Stat. Assoc.*, **120** (2025), 1545–1557. https://doi.org/10.1080/01621459.2024.2447111

21. Y. Hamura, T. Kubokawa, Bayesian predictive distribution for a negative binomial model, *Math. Methods Stat.*, **28** (2019), 1–17. https://doi.org/10.3103/S1066530719010010

22. H. H. Huang, Y. H. Chen, M. Scott, Bayesian deep count regression and anomaly detection: Evidence from GDELT event panels, preprint, arXiv:2603.25970.

23. A. Gelman, A. Jakulin, M. G. Pittau, Y. Su, A weakly informative default prior distribution for logistic and other regression models, *Ann. Appl. Stat.*, **2** (2008), 1360–1383. https://doi.org/10.1214/08-AOAS191
24. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, *Bayesian Data Analysis*, 3rd edition, Chapman and Hall/CRC, Boca Raton, 2013. https://doi.org/10.1201/b16018
25. Stan Development Team, *Prior Choice Recommendations*, 2026. Available from: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.
26. J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, A. Gelman, Visualization in Bayesian workflow, *J. R. Stat. Soc. A*, **182** (2019), 389–402. https://doi.org/10.1111/rssa.12378
27. M. D. Hoffman, A. Gelman, The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.*, **15** (2014), 1593–1623. Available from: https://jmlr.org/papers/v15/hoffman14a.html.
28. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences, *Stat. Sci.*, **7** (1992), 457–472. https://doi.org/10.1214/ss/1177011136
29. A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P. C. Bürkner, Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC, *Bayesian Anal.*, **16** (2021), 667–718. https://doi.org/10.1214/20-BA1221
30. K. Fan, S. Subedi, G. Yang, X. Lu, J. Ren, C. Wu, Is seeing believing? A practitioner's perspective on high-dimensional statistical inference in cancer genomics studies, *Entropy*, **26** (2024), 794. https://doi.org/10.3390/e26090794
31. K. Fan, S. Subedi, V. Dissanayake, C. Wu, Robust Bayesian high-dimensional variable selection and inference with the horseshoe family of priors, *Comput. Stat. Data Anal.*, **219** (2026), 108358. https://doi.org/10.1016/j.csda.2026.108358
32. A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Am. Math. Soc.*, **54** (1943), 426–482. https://doi.org/10.2307/1990256
33. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, 1994. https://doi.org/10.1201/9780429246593
34. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S*, Springer, New York, 2002. https://doi.org/10.1007/978-0-387-21706-2
35. B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth, *MASS: Support Functions and Datasets for Venables and Ripley's Modern Applied Statistics with S*, R package, 2013.
36. M. Aitkin, The analysis of unbalanced cross-classifications, *J. R. Stat. Soc. A*, **141** (1978), 195–223. https://doi.org/10.2307/2344453
37. P. F. Thall, S. C. Vail, Some covariance models for longitudinal count data with overdispersion, *Biometrics*, **46** (1990), 657–671. https://doi.org/10.2307/2532086

Appendix

This appendix contains the parts of Tables 1–3 that were omitted from the main text to improve readability. The correctly specified NB2 confidence-interval and prediction-interval tables report the remaining mean settings $\mu_0 = 3, 5, 10$. The misspecification table reports the remaining misspecification mean settings $\mu_0 = 5, 10$.

Table S1. Coverage probability (CP) and expected width (EW) of 95% intervals for the NB2 mean μ under correctly specified NB2 data generation for the remaining mean settings. Entries are CP (EW).

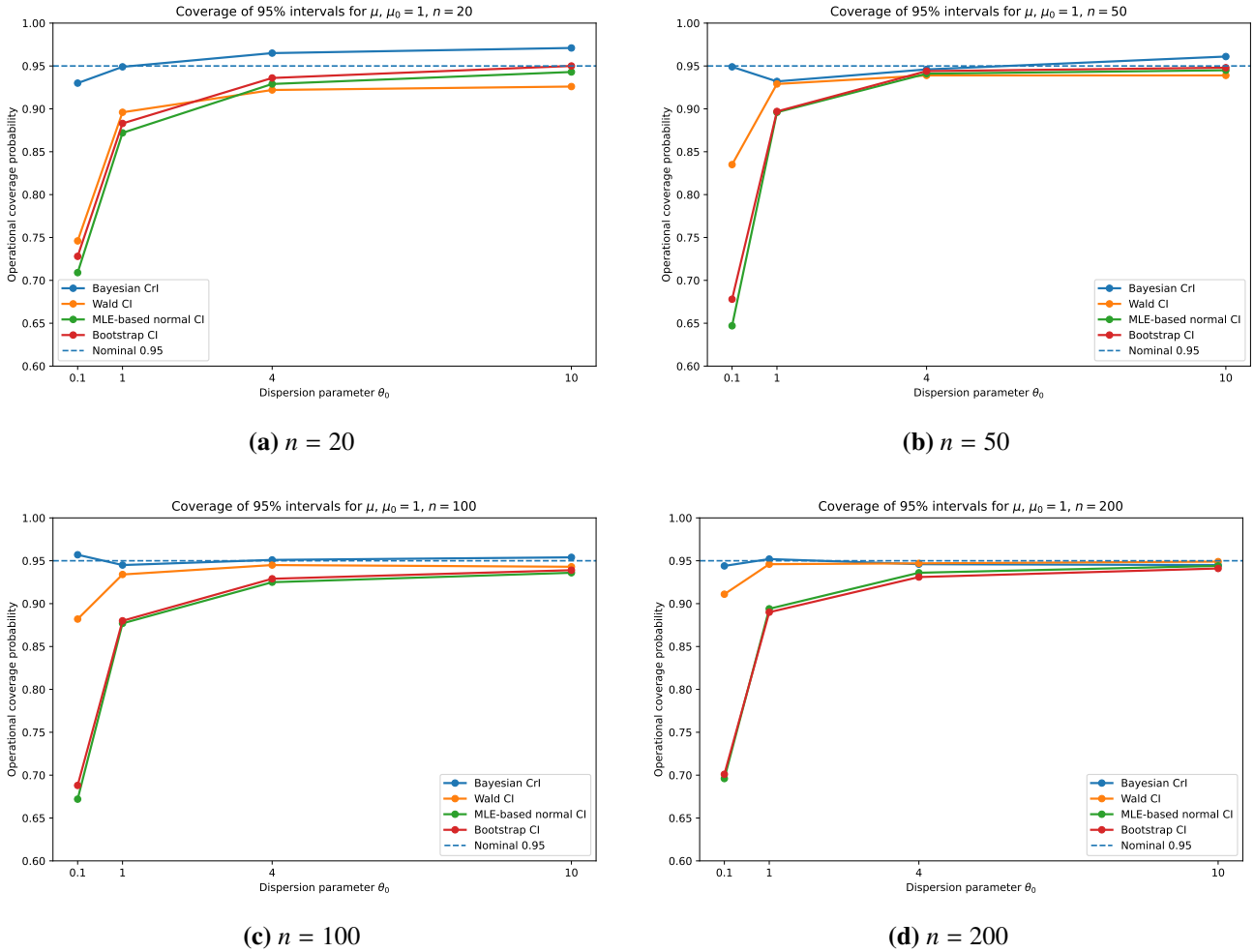
n	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
			$\mu_0 = 3$		
20	Bayesian CrI	0.938 (14.18)	0.943 (3.40)	0.954 (2.13)	0.959 (1.87)
	Wald CI	0.750 (6.19)	0.907 (2.91)	0.926 (1.97)	0.931 (1.70)
	MLE-based normal CI	0.754 (6.93)	0.771 (2.05)	0.879 (1.59)	0.925 (1.54)
	Bootstrap CI	0.779 (7.70)	0.766 (2.03)	0.874 (1.57)	0.924 (1.53)
50	Bayesian CrI	0.955 (7.24)	0.959 (2.01)	0.946 (1.28)	0.954 (1.12)
	Wald CI	0.843 (4.73)	0.930 (1.88)	0.937 (1.26)	0.940 (1.09)
	MLE-based normal CI	0.679 (4.40)	0.693 (1.02)	0.866 (0.99)	0.917 (0.97)
	Bootstrap CI	0.689 (4.34)	0.692 (1.01)	0.865 (0.98)	0.914 (0.96)
100	Bayesian CrI	0.947 (4.35)	0.949 (1.39)	0.948 (0.90)	0.943 (0.78)
	Wald CI	0.886 (3.56)	0.937 (1.35)	0.943 (0.89)	0.951 (0.77)
	MLE-based normal CI	0.753 (3.26)	0.687 (0.71)	0.866 (0.70)	0.922 (0.69)
	Bootstrap CI	0.754 (3.20)	0.685 (0.71)	0.865 (0.69)	0.919 (0.68)
200	Bayesian CrI	0.952 (2.89)	0.946 (0.97)	0.953 (0.64)	0.944 (0.55)
	Wald CI	0.917 (2.59)	0.945 (0.96)	0.948 (0.63)	0.949 (0.55)
	MLE-based normal CI	0.824 (2.41)	0.688 (0.50)	0.870 (0.49)	0.920 (0.49)
	Bootstrap CI	0.823 (2.38)	0.685 (0.50)	0.867 (0.49)	0.918 (0.48)
			$\mu_0 = 5$		
20	Bayesian CrI	0.941 (21.39)	0.948 (5.30)	0.943 (3.09)	0.954 (2.56)
	Wald CI	0.743 (10.30)	0.902 (4.60)	0.924 (2.88)	0.930 (2.36)
	MLE-based normal CI	0.777 (11.78)	0.765 (3.61)	0.822 (2.08)	0.892 (1.99)
	Bootstrap CI	0.811 (13.01)	0.777 (3.57)	0.831 (2.06)	0.900 (1.97)
50	Bayesian CrI	0.949 (11.19)	0.956 (3.14)	0.953 (1.89)	0.944 (1.54)
	Wald CI	0.837 (7.84)	0.929 (2.98)	0.941 (1.84)	0.949 (1.51)
	MLE-based normal CI	0.833 (8.39)	0.593 (1.31)	0.813 (1.25)	0.903 (1.26)
	Bootstrap CI	0.843 (8.24)	0.593 (1.30)	0.813 (1.24)	0.903 (1.24)
100	Bayesian CrI	0.951 (7.09)	0.955 (2.19)	0.954 (1.32)	0.951 (1.08)
	Wald CI	0.883 (5.86)	0.944 (2.13)	0.941 (1.31)	0.947 (1.07)
	MLE-based normal CI	0.896 (6.15)	0.589 (0.89)	0.808 (0.88)	0.889 (0.89)
	Bootstrap CI	0.908 (6.04)	0.591 (0.89)	0.806 (0.87)	0.891 (0.88)
200	Bayesian CrI	0.952 (4.69)	0.953 (1.53)	0.957 (0.93)	0.949 (0.76)
	Wald CI	0.916 (4.29)	0.942 (1.51)	0.946 (0.93)	0.950 (0.76)
	MLE-based normal CI	0.927 (4.43)	0.568 (0.62)	0.809 (0.62)	0.898 (0.63)
	Bootstrap CI	0.933 (4.37)	0.565 (0.62)	0.805 (0.62)	0.894 (0.62)
			$\mu_0 = 10$		
20	Bayesian CrI	0.926 (35.16)	0.956 (10.00)	0.947 (5.44)	0.953 (4.13)
	Wald CI	0.744 (20.57)	0.905 (8.82)	0.932 (5.07)	0.930 (3.85)
	MLE-based normal CI	0.787 (23.74)	0.892 (8.76)	0.829 (4.09)	0.851 (3.02)
	Bootstrap CI	0.826 (26.22)	0.899 (8.64)	0.833 (4.05)	0.856 (2.99)
50	Bayesian CrI	0.943 (21.31)	0.943 (5.97)	0.943 (3.33)	0.943 (2.52)
	Wald CI	0.845 (15.61)	0.929 (5.70)	0.938 (3.25)	0.938 (2.46)
	MLE-based normal CI	0.867 (17.45)	0.841 (5.17)	0.699 (1.78)	0.832 (1.75)
	Bootstrap CI	0.894 (17.10)	0.839 (5.11)	0.701 (1.76)	0.834 (1.74)

Continued on next page

n	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
100	Bayesian CrI	0.949 (13.60)	0.943 (4.18)	0.954 (2.34)	0.958 (1.77)
	Wald CI	0.885 (11.74)	0.938 (4.07)	0.945 (2.31)	0.946 (1.75)
	MLE-based normal CI	0.900 (12.47)	0.886 (3.84)	0.703 (1.24)	0.834 (1.24)
	Bootstrap CI	0.915 (12.24)	0.884 (3.81)	0.699 (1.23)	0.832 (1.23)
200	Bayesian CrI	0.952 (9.29)	0.939 (2.92)	0.947 (1.64)	0.939 (1.24)
	Wald CI	0.914 (8.52)	0.946 (2.90)	0.952 (1.63)	0.950 (1.24)
	MLE-based normal CI	0.923 (8.79)	0.921 (2.81)	0.706 (0.88)	0.838 (0.88)
	Bootstrap CI	0.929 (8.67)	0.919 (2.79)	0.702 (0.87)	0.834 (0.87)

Note: CP is operational coverage, with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. The method labelled as likelihood CI in the raw frequentist output is reported here as the MLE-based normal CI. Bayesian rows use the updated Bayesian simulation results.

Figure S1 provides a graphical summary of the operational coverage probabilities for the 95% interval procedures for μ under correctly specified NB2 data generation for the representative case $\mu_0 = 1$.



(a) $n = 20$ **(b)** $n = 50$

(c) $n = 100$ **(d)** $n = 200$

Figure S1. Graphical summary of the operational coverage probabilities of the 95% interval procedures for the NB2 mean μ under correctly specified NB2 data generation for the representative case $\mu_0 = 1$. The four panels correspond to $n = 20$, 50, 100, and 200. The horizontal dashed line marks the nominal coverage level of 0.95.

Table S2. Coverage probability (CP) and expected width (EW) of 95% prediction intervals under correctly specified NB2 data generation for the remaining mean settings. Entries are CP (EW). The case $m = 1$ corresponds to prediction of a single future count, whereas $m > 1$ corresponds to prediction of the future sample mean $\overline{Y}_{m,\text{new}}$.

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
			$\mu_0 = 3$			
20	1	Bayesian PI	0.963 (32.93)	0.975 (13.48)	0.980 (9.04)	0.988 (8.12)
		Wald PI	0.935 (18.84)	0.952 (10.19)	0.970 (8.02)	0.975 (7.38)
		MLE-based normal PI	0.938 (22.05)	0.916 (8.25)	0.955 (7.17)	0.971 (7.06)
		Bootstrap PI	0.939 (25.41)	0.886 (8.00)	0.944 (6.93)	0.965 (6.91)
	5	Bayesian PPI	0.953 (23.94)	0.952 (7.33)	0.956 (4.71)	0.968 (4.15)
		Wald PI	0.866 (10.58)	0.931 (6.27)	0.950 (4.60)	0.954 (4.00)
		MLE-based normal PI	0.855 (12.13)	0.808 (4.58)	0.910 (3.75)	0.947 (3.65)
		Bootstrap PI	0.868 (17.03)	0.761 (4.23)	0.886 (3.45)	0.932 (3.40)
	10	Bayesian PPI	0.944 (20.66)	0.952 (5.77)	0.959 (3.67)	0.961 (3.23)
		Wald PI	0.835 (8.84)	0.927 (5.09)	0.944 (3.51)	0.946 (3.04)
		MLE-based normal PI	0.823 (10.01)	0.787 (3.62)	0.899 (2.85)	0.938 (2.77)
		Bootstrap PI	0.838 (13.95)	0.760 (3.33)	0.881 (2.69)	0.925 (2.65)
	20	Bayesian PPI	0.933 (18.19)	0.947 (4.75)	0.954 (3.00)	0.973 (2.64)
		Wald PI	0.808 (7.75)	0.915 (4.17)	0.940 (2.84)	0.945 (2.45)
		MLE-based normal PI	0.802 (8.68)	0.775 (2.95)	0.893 (2.30)	0.933 (2.23)
		Bootstrap PI	0.819 (11.62)	0.758 (2.77)	0.882 (2.21)	0.927 (2.17)
	30	Bayesian PPI	0.940 (17.11)	0.945 (4.36)	0.953 (2.75)	0.955 (2.41)
		Wald PI	0.794 (7.31)	0.916 (3.79)	0.938 (2.58)	0.938 (2.23)
		MLE-based normal PI	0.785 (8.16)	0.773 (2.68)	0.886 (2.08)	0.925 (2.02)
		Bootstrap PI	0.808 (10.57)	0.758 (2.55)	0.879 (2.03)	0.921 (1.98)
50	1	Bayesian PI	0.971 (32.45)	0.978 (12.78)	0.976 (8.63)	0.981 (7.74)
		Wald PI	0.952 (20.57)	0.953 (10.22)	0.975 (7.99)	0.981 (7.38)
		MLE-based normal PI	0.936 (19.40)	0.897 (7.12)	0.955 (7.04)	0.975 (6.99)
		Bootstrap PI	0.932 (21.92)	0.870 (6.79)	0.946 (6.83)	0.970 (6.84)
	5	Bayesian PPI	0.982 (19.61)	0.955 (6.49)	0.953 (4.22)	0.956 (3.70)
		Wald PI	0.913 (11.03)	0.951 (6.15)	0.963 (4.37)	0.964 (3.81)
		MLE-based normal PI	0.769 (10.27)	0.769 (3.54)	0.915 (3.47)	0.951 (3.42)
		Bootstrap PI	0.755 (13.44)	0.718 (3.21)	0.890 (3.19)	0.931 (3.18)
	10	Bayesian PPI	0.959 (15.51)	0.955 (4.85)	0.961 (3.13)	0.963 (2.75)
		Wald PI	0.898 (8.91)	0.946 (4.70)	0.956 (3.18)	0.961 (2.77)
		MLE-based normal PI	0.748 (8.16)	0.735 (2.59)	0.896 (2.52)	0.943 (2.48)
		Bootstrap PI	0.736 (10.48)	0.699 (2.39)	0.876 (2.37)	0.929 (2.36)
	20	Bayesian PPI	0.952 (12.48)	0.949 (3.73)	0.949 (2.39)	0.953 (2.10)
		Wald PI	0.888 (7.50)	0.939 (3.57)	0.952 (2.40)	0.957 (2.09)
		MLE-based normal PI	0.747 (6.77)	0.713 (1.95)	0.891 (1.90)	0.932 (1.87)
		Bootstrap PI	0.729 (8.23)	0.690 (1.84)	0.879 (1.81)	0.923 (1.80)
	30	Bayesian PPI	0.943 (11.15)	0.948 (3.26)	0.948 (2.09)	0.956 (1.83)
		Wald PI	0.879 (6.90)	0.942 (3.11)	0.948 (2.09)	0.953 (1.81)
		MLE-based normal PI	0.739 (6.20)	0.710 (1.69)	0.886 (1.65)	0.934 (1.62)
		Bootstrap PI	0.723 (7.22)	0.691 (1.61)	0.874 (1.59)	0.925 (1.58)
100	1	Bayesian PI	0.975 (30.71)	0.973 (12.58)	0.979 (8.54)	0.985 (7.56)
		Wald PI	0.957 (21.38)	0.960 (10.26)	0.972 (8.00)	0.976 (7.37)
		MLE-based normal PI	0.943 (19.88)	0.903 (7.09)	0.954 (7.03)	0.970 (6.99)
		Bootstrap PI	0.944 (23.19)	0.884 (6.80)	0.948 (6.83)	0.965 (6.83)
	5	Bayesian PPI	0.974 (17.09)	0.961 (6.24)	0.959 (4.09)	0.965 (3.54)
		Wald PI	0.929 (11.26)	0.960 (6.14)	0.968 (4.30)	0.970 (3.74)
		MLE-based normal PI	0.804 (10.38)	0.764 (3.46)	0.917 (3.41)	0.954 (3.36)
		Bootstrap PI	0.803 (13.38)	0.713 (3.12)	0.885 (3.11)	0.937 (3.11)
	10	Bayesian PPI	0.954 (13.02)	0.958 (4.56)	0.955 (2.97)	0.954 (2.57)
		Wald PI	0.925 (8.96)	0.958 (4.56)	0.961 (3.07)	0.963 (2.66)
		MLE-based normal PI	0.796 (8.16)	0.735 (2.46)	0.902 (2.42)	0.943 (2.38)
		Bootstrap PI	0.788 (10.19)	0.702 (2.27)	0.880 (2.26)	0.929 (2.25)
	20	Bayesian PPI	0.954 (10.01)	0.957 (3.39)	0.956 (2.20)	0.949 (1.90)
		Wald PI	0.921 (7.39)	0.954 (3.35)	0.959 (2.24)	0.956 (1.94)
		MLE-based normal PI	0.793 (6.65)	0.716 (1.79)	0.896 (1.76)	0.933 (1.74)
		Bootstrap PI	0.777 (7.76)	0.692 (1.69)	0.879 (1.67)	0.920 (1.67)

Continued on next page

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
	30	Bayesian PPI	0.956 (8.65)	0.957 (2.88)	0.948 (1.87)	0.952 (1.62)
		Wald PI	0.914 (6.71)	0.949 (2.84)	0.952 (1.90)	0.955 (1.64)
		MLE-based normal PI	0.786 (6.02)	0.700 (1.52)	0.886 (1.49)	0.932 (1.47)
		Bootstrap PI	0.770 (6.64)	0.679 (1.44)	0.872 (1.43)	0.921 (1.42)
200	1	Bayesian PI	0.986 (30.56)	0.978 (12.44)	0.978 (8.52)	0.975 (7.51)
		Wald PI	0.960 (21.88)	0.961 (10.27)	0.973 (8.00)	0.979 (7.35)
		MLE-based normal PI	0.951 (20.58)	0.901 (7.07)	0.955 (7.05)	0.973 (7.01)
		Bootstrap PI	0.954 (24.63)	0.889 (6.80)	0.949 (6.84)	0.971 (6.84)
	5	Bayesian PPI	0.978 (16.32)	0.966 (6.09)	0.953 (4.03)	0.965 (3.48)
		Wald PI	0.941 (11.40)	0.955 (6.14)	0.967 (4.25)	0.969 (3.70)
		MLE-based normal PI	0.850 (10.67)	0.765 (3.41)	0.916 (3.36)	0.950 (3.31)
		Bootstrap PI	0.857 (13.67)	0.710 (3.07)	0.887 (3.07)	0.931 (3.06)
	10	Bayesian PPI	0.961 (12.18)	0.966 (4.40)	0.954 (2.90)	0.960 (2.50)
		Wald PI	0.942 (9.00)	0.960 (4.48)	0.962 (3.00)	0.964 (2.60)
		MLE-based normal PI	0.846 (8.35)	0.737 (2.40)	0.906 (2.37)	0.941 (2.33)
		Bootstrap PI	0.841 (10.27)	0.699 (2.21)	0.881 (2.20)	0.929 (2.20)
	20	Bayesian PPI	0.953 (9.12)	0.956 (3.20)	0.965 (2.10)	0.947 (1.81)
		Wald PI	0.932 (7.33)	0.955 (3.22)	0.958 (2.15)	0.957 (1.86)
		MLE-based normal PI	0.841 (6.75)	0.723 (1.72)	0.892 (1.69)	0.932 (1.67)
		Bootstrap PI	0.829 (7.67)	0.695 (1.60)	0.872 (1.60)	0.919 (1.59)
	30	Bayesian PPI	0.946 (7.72)	0.949 (2.68)	0.965 (1.76)	0.951 (1.51)
		Wald PI	0.935 (6.60)	0.949 (2.68)	0.955 (1.79)	0.961 (1.55)
		MLE-based normal PI	0.847 (6.06)	0.708 (1.42)	0.886 (1.40)	0.933 (1.38)
		Bootstrap PI	0.832 (6.47)	0.689 (1.34)	0.868 (1.34)	0.922 (1.33)
			$\mu_0 = 5$			
20	1	Bayesian PI	0.965 (53.86)	0.967 (21.51)	0.970 (13.54)	0.978 (11.39)
		Wald PI	0.933 (30.94)	0.943 (16.02)	0.962 (12.08)	0.971 (10.83)
		MLE-based normal PI	0.939 (36.90)	0.896 (13.74)	0.927 (10.16)	0.959 (9.95)
		Bootstrap PI	0.944 (44.19)	0.836 (14.50)	0.878 (9.04)	0.936 (8.86)
	5	Bayesian PPI	0.957 (38.08)	0.957 (11.52)	0.950 (6.85)	0.962 (5.69)
		Wald PI	0.862 (17.52)	0.931 (9.96)	0.944 (6.63)	0.949 (5.47)
		MLE-based normal PI	0.869 (20.41)	0.802 (7.89)	0.856 (4.85)	0.919 (4.64)
		Bootstrap PI	0.889 (29.02)	0.768 (7.60)	0.832 (4.51)	0.902 (4.39)
	10	Bayesian PPI	0.949 (32.43)	0.936 (9.04)	0.942 (5.33)	0.965 (4.42)
		Wald PI	0.833 (14.67)	0.923 (8.00)	0.940 (5.08)	0.946 (4.18)
		MLE-based normal PI	0.843 (16.89)	0.792 (6.32)	0.846 (3.71)	0.913 (3.54)
		Bootstrap PI	0.863 (23.65)	0.767 (5.98)	0.829 (3.52)	0.901 (3.41)
	20	Bayesian PPI	0.936 (28.36)	0.947 (7.44)	0.952 (4.36)	0.953 (3.61)
		Wald PI	0.801 (12.87)	0.915 (6.55)	0.935 (4.12)	0.940 (3.38)
		MLE-based normal PI	0.815 (14.68)	0.794 (5.16)	0.840 (3.00)	0.905 (2.86)
		Bootstrap PI	0.844 (19.75)	0.772 (4.92)	0.831 (2.89)	0.896 (2.79)
	30	Bayesian PPI	0.944 (26.41)	0.947 (6.80)	0.952 (3.99)	0.956 (3.30)
		Wald PI	0.785 (12.16)	0.913 (5.97)	0.932 (3.75)	0.934 (3.07)
		MLE-based normal PI	0.804 (13.82)	0.784 (4.70)	0.835 (2.72)	0.902 (2.60)
		Bootstrap PI	0.839 (17.90)	0.770 (4.52)	0.830 (2.65)	0.896 (2.55)
50	1	Bayesian PI	0.967 (52.44)	0.972 (20.33)	0.975 (13.12)	0.962 (10.67)
		Wald PI	0.949 (33.76)	0.951 (16.15)	0.968 (12.06)	0.977 (10.88)
		MLE-based normal PI	0.946 (35.77)	0.839 (10.03)	0.927 (9.90)	0.962 (9.95)
		Bootstrap PI	0.948 (44.50)	0.681 (8.92)	0.876 (8.74)	0.933 (8.78)
	5	Bayesian PPI	0.966 (31.21)	0.952 (10.20)	0.950 (6.22)	0.958 (5.08)
		Wald PI	0.913 (18.23)	0.944 (9.81)	0.958 (6.30)	0.959 (5.20)
		MLE-based normal PI	0.870 (19.06)	0.642 (4.53)	0.857 (4.36)	0.923 (4.37)
		Bootstrap PI	0.885 (26.21)	0.607 (4.24)	0.829 (4.11)	0.904 (4.11)
	10	Bayesian PPI	0.950 (24.56)	0.945 (7.61)	0.948 (4.61)	0.961 (3.76)
		Wald PI	0.898 (14.75)	0.945 (7.40)	0.953 (4.61)	0.956 (3.79)
		MLE-based normal PI	0.858 (15.29)	0.630 (3.32)	0.842 (3.17)	0.915 (3.18)
		Bootstrap PI	0.863 (20.29)	0.603 (3.14)	0.825 (3.04)	0.901 (3.04)
	20	Bayesian PPI	0.950 (19.67)	0.950 (5.84)	0.959 (3.53)	0.952 (2.88)
		Wald PI	0.881 (12.43)	0.940 (5.63)	0.948 (3.49)	0.957 (2.87)
		MLE-based normal PI	0.847 (12.78)	0.614 (2.50)	0.830 (2.40)	0.913 (2.40)
		Bootstrap PI	0.849 (15.84)	0.597 (2.41)	0.818 (2.33)	0.902 (2.33)

Continued on next page

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
	30	Bayesian PPI	0.943 (17.51)	0.953 (5.11)	0.949 (3.08)	0.951 (2.51)
		Wald PI	0.878 (11.45)	0.934 (4.90)	0.951 (3.04)	0.951 (2.50)
		MLE-based normal PI	0.844 (11.75)	0.604 (2.18)	0.833 (2.08)	0.908 (2.09)
		Bootstrap PI	0.851 (13.89)	0.589 (2.11)	0.822 (2.03)	0.899 (2.04)
100	1	Bayesian PI	0.969 (50.85)	0.979 (20.08)	0.972 (13.02)	0.969 (10.44)
		Wald PI	0.955 (34.96)	0.952 (16.18)	0.968 (12.07)	0.975 (10.88)
		MLE-based normal PI	0.957 (36.39)	0.849 (9.89)	0.927 (9.90)	0.959 (9.95)
		Bootstrap PI	0.965 (46.97)	0.675 (8.73)	0.875 (8.74)	0.934 (8.77)
	5	Bayesian PPI	0.971 (28.25)	0.959 (9.84)	0.962 (6.01)	0.956 (4.92)
		Wald PI	0.928 (18.53)	0.956 (9.77)	0.962 (6.20)	0.967 (5.11)
		MLE-based normal PI	0.916 (19.15)	0.633 (4.29)	0.858 (4.24)	0.925 (4.26)
		Bootstrap PI	0.941 (25.97)	0.598 (4.04)	0.830 (4.01)	0.904 (4.01)
	10	Bayesian PPI	0.955 (21.44)	0.943 (7.20)	0.945 (4.36)	0.949 (3.57)
		Wald PI	0.919 (14.77)	0.950 (7.15)	0.952 (4.44)	0.958 (3.65)
		MLE-based normal PI	0.909 (15.19)	0.620 (3.07)	0.839 (3.03)	0.915 (3.04)
		Bootstrap PI	0.920 (19.61)	0.600 (2.93)	0.819 (2.90)	0.901 (2.91)
	20	Bayesian PPI	0.950 (16.45)	0.954 (5.34)	0.947 (3.23)	0.953 (2.64)
		Wald PI	0.917 (12.20)	0.953 (5.26)	0.953 (3.26)	0.959 (2.67)
		MLE-based normal PI	0.910 (12.49)	0.603 (2.24)	0.828 (2.21)	0.913 (2.22)
		Bootstrap PI	0.913 (14.90)	0.589 (2.17)	0.813 (2.15)	0.902 (2.15)
	30	Bayesian PPI	0.950 (14.20)	0.956 (4.55)	0.956 (2.74)	0.955 (2.24)
		Wald PI	0.910 (11.08)	0.948 (4.46)	0.950 (2.76)	0.953 (2.26)
		MLE-based normal PI	0.904 (11.34)	0.610 (1.90)	0.828 (1.87)	0.898 (1.88)
		Bootstrap PI	0.908 (12.75)	0.596 (1.84)	0.816 (1.83)	0.891 (1.83)
200	1	Bayesian PI	0.969 (49.99)	0.979 (19.87)	0.980 (13.04)	0.971 (10.29)
		Wald PI	0.960 (35.93)	0.955 (16.22)	0.968 (12.07)	0.978 (10.88)
		MLE-based normal PI	0.962 (36.91)	0.851 (9.87)	0.929 (9.92)	0.962 (9.98)
		Bootstrap PI	0.971 (48.74)	0.677 (8.70)	0.881 (8.74)	0.935 (8.78)
	5	Bayesian PPI	0.980 (26.67)	0.957 (9.62)	0.960 (5.94)	0.956 (4.83)
		Wald PI	0.941 (18.85)	0.957 (9.78)	0.962 (6.14)	0.963 (5.05)
		MLE-based normal PI	0.941 (19.29)	0.633 (4.18)	0.858 (4.18)	0.928 (4.21)
		Bootstrap PI	0.968 (25.92)	0.601 (3.96)	0.830 (3.96)	0.910 (3.96)
	10	Bayesian PPI	0.964 (19.87)	0.954 (6.96)	0.960 (4.27)	0.956 (3.46)
		Wald PI	0.940 (14.89)	0.955 (7.03)	0.960 (4.35)	0.961 (3.57)
		MLE-based normal PI	0.943 (15.20)	0.617 (2.95)	0.844 (2.95)	0.918 (2.97)
		Bootstrap PI	0.952 (19.29)	0.595 (2.84)	0.826 (2.83)	0.902 (2.84)
	20	Bayesian PPI	0.950 (14.86)	0.946 (5.05)	0.952 (3.09)	0.951 (2.51)
		Wald PI	0.937 (12.15)	0.951 (5.06)	0.955 (3.13)	0.958 (2.56)
		MLE-based normal PI	0.941 (12.37)	0.601 (2.11)	0.831 (2.11)	0.908 (2.13)
		Bootstrap PI	0.943 (14.39)	0.589 (2.06)	0.817 (2.05)	0.894 (2.06)
	30	Bayesian PPI	0.947 (12.56)	0.952 (4.22)	0.952 (2.59)	0.945 (2.10)
		Wald PI	0.932 (10.95)	0.952 (4.22)	0.956 (2.60)	0.957 (2.13)
		MLE-based normal PI	0.938 (11.15)	0.595 (1.75)	0.836 (1.75)	0.910 (1.77)
		Bootstrap PI	0.939 (12.14)	0.585 (1.72)	0.826 (1.72)	0.900 (1.72)
			$\mu_0 = 10$			
20	1	Bayesian PI	0.944 (97.40)	0.978 (41.39)	0.956 (24.08)	0.966 (18.56)
		Wald PI	0.931 (61.29)	0.940 (30.70)	0.958 (22.00)	0.963 (18.34)
		MLE-based normal PI	0.940 (73.74)	0.930 (30.50)	0.891 (18.36)	0.918 (14.74)
		Bootstrap PI	0.947 (90.49)	0.936 (37.06)	0.848 (17.23)	0.881 (13.22)
	5	Bayesian PPI	0.947 (66.51)	0.951 (21.83)	0.952 (12.09)	0.959 (9.20)
		Wald PI	0.866 (34.89)	0.930 (19.20)	0.939 (11.55)	0.948 (8.81)
		MLE-based normal PI	0.882 (40.95)	0.914 (19.11)	0.850 (9.34)	0.883 (6.96)
		Bootstrap PI	0.911 (59.07)	0.906 (19.24)	0.823 (8.63)	0.863 (6.55)
	10	Bayesian PPI	0.924 (55.71)	0.954 (17.12)	0.948 (9.40)	0.953 (7.14)
		Wald PI	0.828 (29.25)	0.925 (15.30)	0.938 (8.89)	0.942 (6.77)
		MLE-based normal PI	0.850 (33.94)	0.911 (15.25)	0.844 (7.19)	0.872 (5.34)
		Bootstrap PI	0.880 (47.85)	0.903 (14.97)	0.823 (6.73)	0.860 (5.10)
	20	Bayesian PPI	0.929 (47.78)	0.958 (14.07)	0.951 (7.69)	0.965 (5.84)
		Wald PI	0.806 (25.68)	0.919 (12.51)	0.938 (7.23)	0.939 (5.50)
		MLE-based normal PI	0.832 (29.53)	0.907 (12.44)	0.841 (5.83)	0.867 (4.33)
		Bootstrap PI	0.872 (39.58)	0.905 (12.23)	0.823 (5.56)	0.857 (4.19)

Continued on next page

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
	30	Bayesian PPI	0.924 (44.42)	0.956 (12.88)	0.948 (7.02)	0.956 (5.33)
		Wald PI	0.795 (24.27)	0.915 (11.41)	0.929 (6.58)	0.937 (5.01)
		MLE-based normal PI	0.825 (27.82)	0.902 (11.35)	0.832 (5.31)	0.863 (3.94)
		Bootstrap PI	0.866 (36.14)	0.901 (11.17)	0.821 (5.12)	0.855 (3.84)
50	1	Bayesian PI	0.970 (103.00)	0.980 (39.17)	0.962 (23.13)	0.964 (17.67)
		Wald PI	0.948 (66.72)	0.945 (30.83)	0.961 (22.07)	0.969 (18.47)
		MLE-based normal PI	0.952 (73.26)	0.890 (28.50)	0.814 (13.64)	0.907 (13.51)
		Bootstrap PI	0.962 (94.80)	0.887 (33.39)	0.756 (12.46)	0.869 (12.40)
	5	Bayesian PPI	0.970 (60.85)	0.952 (19.43)	0.958 (10.99)	0.957 (8.32)
		Wald PI	0.912 (36.23)	0.945 (18.78)	0.951 (10.96)	0.952 (8.37)
		MLE-based normal PI	0.920 (39.25)	0.868 (17.12)	0.747 (6.09)	0.860 (6.01)
		Bootstrap PI	0.947 (54.70)	0.848 (16.45)	0.724 (5.83)	0.845 (5.80)
	10	Bayesian PPI	0.950 (47.64)	0.949 (14.50)	0.946 (8.15)	0.961 (6.16)
		Wald PI	0.901 (29.36)	0.940 (14.05)	0.948 (8.05)	0.948 (6.13)
		MLE-based normal PI	0.911 (31.58)	0.871 (12.76)	0.736 (4.46)	0.852 (4.39)
		Bootstrap PI	0.927 (42.36)	0.852 (12.19)	0.720 (4.31)	0.837 (4.28)
	20	Bayesian PPI	0.947 (38.04)	0.940 (11.13)	0.953 (6.23)	0.952 (4.71)
		Wald PI	0.885 (24.77)	0.938 (10.71)	0.949 (6.12)	0.946 (4.66)
		MLE-based normal PI	0.897 (26.47)	0.858 (9.72)	0.734 (3.38)	0.845 (3.33)
		Bootstrap PI	0.916 (33.09)	0.842 (9.31)	0.721 (3.30)	0.836 (3.27)
	30	Bayesian PPI	0.942 (33.78)	0.952 (9.72)	0.954 (5.44)	0.951 (4.10)
		Wald PI	0.883 (22.82)	0.935 (9.34)	0.943 (5.33)	0.944 (4.05)
		MLE-based normal PI	0.897 (24.37)	0.856 (8.47)	0.717 (2.94)	0.841 (2.89)
		Bootstrap PI	0.916 (28.96)	0.843 (8.15)	0.709 (2.88)	0.835 (2.86)
100	1	Bayesian PI	0.971 (99.14)	0.981 (38.89)	0.966 (23.00)	0.964 (17.50)
		Wald PI	0.955 (69.53)	0.949 (30.96)	0.963 (22.09)	0.968 (18.54)
		MLE-based normal PI	0.958 (73.16)	0.913 (29.55)	0.815 (13.49)	0.904 (13.51)
		Bootstrap PI	0.971 (96.91)	0.920 (35.08)	0.755 (12.33)	0.865 (12.35)
	5	Bayesian PPI	0.972 (54.88)	0.953 (18.82)	0.955 (10.69)	0.953 (8.08)
		Wald PI	0.927 (37.02)	0.952 (18.70)	0.956 (10.77)	0.960 (8.21)
		MLE-based normal PI	0.931 (38.68)	0.909 (17.74)	0.735 (5.89)	0.863 (5.88)
		Bootstrap PI	0.965 (52.97)	0.891 (16.99)	0.714 (5.66)	0.843 (5.67)
	10	Bayesian PPI	0.948 (41.52)	0.942 (13.76)	0.962 (7.76)	0.956 (5.86)
		Wald PI	0.920 (29.54)	0.950 (13.60)	0.954 (7.75)	0.953 (5.90)
		MLE-based normal PI	0.930 (30.74)	0.905 (12.85)	0.727 (4.22)	0.852 (4.21)
		Bootstrap PI	0.941 (39.90)	0.890 (12.39)	0.712 (4.10)	0.838 (4.10)
	20	Bayesian PPI	0.946 (31.79)	0.949 (10.21)	0.962 (5.74)	0.947 (4.33)
		Wald PI	0.914 (24.42)	0.948 (10.02)	0.951 (5.70)	0.952 (4.33)
		MLE-based normal PI	0.923 (25.30)	0.907 (9.47)	0.727 (3.09)	0.852 (3.08)
		Bootstrap PI	0.936 (30.26)	0.891 (9.16)	0.717 (3.03)	0.845 (3.03)
	30	Bayesian PPI	0.945 (27.44)	0.956 (8.69)	0.948 (4.88)	0.955 (3.68)
		Wald PI	0.912 (22.19)	0.944 (8.51)	0.951 (4.84)	0.945 (3.67)
		MLE-based normal PI	0.924 (23.00)	0.899 (8.04)	0.725 (2.62)	0.843 (2.61)
		Bootstrap PI	0.933 (25.95)	0.887 (7.78)	0.717 (2.58)	0.834 (2.58)
200	1	Bayesian PI	0.971 (99.29)	0.980 (38.45)	0.958 (22.78)	0.958 (17.34)
		Wald PI	0.957 (70.91)	0.951 (31.03)	0.962 (22.09)	0.971 (18.52)
		MLE-based normal PI	0.960 (72.80)	0.933 (30.30)	0.813 (13.49)	0.911 (13.56)
		Bootstrap PI	0.972 (97.59)	0.947 (36.49)	0.753 (12.31)	0.871 (12.33)
	5	Bayesian PPI	0.970 (52.97)	0.961 (18.40)	0.951 (10.47)	0.955 (7.94)
		Wald PI	0.935 (37.38)	0.958 (18.67)	0.958 (10.67)	0.960 (8.12)
		MLE-based normal PI	0.940 (38.24)	0.933 (18.18)	0.750 (5.81)	0.864 (5.81)
		Bootstrap PI	0.967 (51.55)	0.917 (17.47)	0.730 (5.60)	0.847 (5.60)
	10	Bayesian PPI	0.959 (39.45)	0.943 (13.29)	0.949 (7.52)	0.949 (5.69)
		Wald PI	0.936 (29.58)	0.955 (13.37)	0.957 (7.59)	0.957 (5.77)
		MLE-based normal PI	0.938 (30.19)	0.932 (12.99)	0.737 (4.12)	0.860 (4.12)
		Bootstrap PI	0.947 (38.28)	0.919 (12.61)	0.723 (4.01)	0.846 (4.01)
	20	Bayesian PPI	0.948 (29.48)	0.946 (9.66)	0.957 (5.45)	0.952 (4.13)
		Wald PI	0.934 (24.16)	0.950 (9.65)	0.953 (5.47)	0.953 (4.15)
		MLE-based normal PI	0.939 (24.60)	0.929 (9.38)	0.721 (2.96)	0.848 (2.96)
		Bootstrap PI	0.949 (28.52)	0.918 (9.16)	0.711 (2.90)	0.838 (2.90)
	30	Bayesian PPI	0.945 (24.95)	0.954 (8.07)	0.948 (4.55)	0.954 (3.44)
		Wald PI	0.934 (21.77)	0.951 (8.05)	0.951 (4.56)	0.950 (3.46)
		MLE-based normal PI	0.939 (22.19)	0.927 (7.82)	0.721 (2.46)	0.845 (2.46)

Continued on next page

n	m	Method	$\theta_0 = 0.1$	$\theta_0 = 1$	$\theta_0 = 4$	$\theta_0 = 10$
		Bootstrap PI	0.944 (24.07)	0.918 (7.65)	0.711 (2.42)	0.840 (2.43)

Note: CP is operational coverage , with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. The method labelled as likelihood PI in the raw frequentist output is reported here as the MLE-based normal PI. Bayesian rows use posterior predictive intervals from the updated Bayesian simulation results.

Table S3. Bayesian and frequentist interval performance under model misspecification for the remaining mean settings. Entries are CP (EW).

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
		Poisson–lognormal		
		$\mu_0 = 5$		
CI for μ	Bayesian CrI	0.807 (8.83)	0.920 (1.90)	0.947 (0.76)
	Wald CI	0.732 (8.17)	0.934 (2.10)	0.944 (0.76)
	MLE-based normal CI	0.654 (6.70)	0.589 (0.93)	0.889 (0.62)
	Bootstrap CI	0.652 (6.58)	0.592 (0.92)	0.884 (0.62)
Y_{new}	Bayesian PI	0.960 (32.79)	0.973 (17.75)	0.968 (10.31)
	Wald PI	0.937 (24.99)	0.959 (16.06)	0.975 (10.87)
	MLE-based normal PI	0.918 (20.90)	0.886 (10.10)	0.959 (9.97)
	Bootstrap PI	0.913 (26.74)	0.777 (8.77)	0.935 (8.77)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.918 (18.44)	0.937 (8.59)	0.961 (4.83)
	Wald PI	0.872 (14.58)	0.949 (9.57)	0.968 (5.05)
	MLE-based normal PI	0.801 (12.29)	0.699 (4.47)	0.928 (4.20)
	Bootstrap PI	0.780 (14.63)	0.653 (4.06)	0.912 (3.96)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.874 (14.73)	0.934 (6.27)	0.950 (3.47)
	Wald PI	0.846 (12.31)	0.944 (7.04)	0.954 (3.57)
	MLE-based normal PI	0.770 (10.40)	0.654 (3.19)	0.909 (2.96)
	Bootstrap PI	0.750 (11.49)	0.625 (2.96)	0.893 (2.84)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.849 (12.28)	0.918 (4.65)	0.954 (2.51)
	Wald PI	0.820 (10.71)	0.943 (5.19)	0.957 (2.56)
	MLE-based normal PI	0.757 (9.06)	0.644 (2.33)	0.913 (2.12)
	Bootstrap PI	0.729 (9.43)	0.624 (2.19)	0.900 (2.06)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.807 (11.27)	0.926 (3.95)	0.959 (2.10)
	Wald PI	0.785 (10.02)	0.937 (4.40)	0.956 (2.13)
	MLE-based normal PI	0.721 (8.45)	0.619 (1.97)	0.909 (1.76)
	Bootstrap PI	0.699 (8.63)	0.598 (1.88)	0.899 (1.72)
		$\mu_0 = 10$		
CI for μ	Bayesian CrI	0.751 (16.04)	0.900 (3.52)	0.949 (1.24)
	Wald CI	0.737 (16.52)	0.927 (4.01)	0.947 (1.24)
	MLE-based normal CI	0.718 (13.59)	0.784 (3.01)	0.835 (0.88)
	Bootstrap CI	0.723 (13.33)	0.777 (2.98)	0.830 (0.87)

Continued on next page

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
Y_{new}	Bayesian PI	0.949 (61.83)	0.967 (33.47)	0.962 (17.26)
	Wald PI	0.932 (50.23)	0.954 (30.65)	0.972 (18.52)
	MLE-based normal PI	0.924 (41.75)	0.884 (24.95)	0.916 (13.55)
	Bootstrap PI	0.929 (55.77)	0.853 (26.03)	0.881 (12.36)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.902 (34.13)	0.918 (15.94)	0.963 (7.91)
	Wald PI	0.877 (29.50)	0.950 (18.22)	0.960 (8.12)
	MLE-based normal PI	0.848 (25.20)	0.840 (13.98)	0.874 (5.81)
	Bootstrap PI	0.837 (29.91)	0.799 (12.48)	0.860 (5.60)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.865 (27.12)	0.924 (11.61)	0.949 (5.67)
	Wald PI	0.843 (24.90)	0.943 (13.36)	0.955 (5.77)
	MLE-based normal PI	0.818 (21.51)	0.836 (10.07)	0.854 (4.12)
	Bootstrap PI	0.801 (23.51)	0.784 (9.09)	0.838 (4.01)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.825 (22.53)	0.905 (8.60)	0.948 (4.11)
	Wald PI	0.820 (21.65)	0.946 (9.86)	0.956 (4.15)
	MLE-based normal PI	0.792 (18.64)	0.823 (7.42)	0.843 (2.96)
	Bootstrap PI	0.769 (19.20)	0.777 (6.74)	0.837 (2.91)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.789 (20.64)	0.896 (7.32)	0.950 (3.43)
	Wald PI	0.787 (20.25)	0.939 (8.37)	0.956 (3.46)
	MLE-based normal PI	0.761 (17.29)	0.818 (6.29)	0.849 (2.46)
	Bootstrap PI	0.739 (17.52)	0.773 (5.75)	0.841 (2.42)
		ZI–NB2		
		$\mu_0 = 5$		
CI for μ	Bayesian CrI	0.940 (24.49)	0.966 (2.89)	0.980 (1.28)
	Wald CI	0.712 (10.63)	0.934 (2.52)	0.950 (1.05)
	MLE-based normal CI	0.755 (12.57)	0.569 (1.18)	0.767 (0.65)
	Bootstrap CI	0.794 (14.52)	0.562 (1.17)	0.762 (0.64)
Y_{new}	Bayesian PI	0.956 (53.89)	0.988 (25.77)	0.999 (16.95)
	Wald PI	0.935 (32.31)	0.949 (18.13)	0.979 (12.94)
	MLE-based normal PI	0.943 (40.75)	0.828 (11.32)	0.918 (10.08)
	Bootstrap PI	0.943 (44.74)	0.578 (10.58)	0.706 (8.73)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.951 (41.52)	0.954 (12.78)	0.983 (8.10)
	Wald PI	0.858 (18.13)	0.957 (10.85)	0.961 (6.92)
	MLE-based normal PI	0.867 (22.24)	0.614 (5.35)	0.797 (4.33)
	Bootstrap PI	0.890 (32.00)	0.576 (5.09)	0.763 (3.96)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.935 (36.05)	0.977 (9.41)	0.982 (5.83)
	Wald PI	0.818 (15.12)	0.949 (8.44)	0.962 (4.91)
	MLE-based normal PI	0.832 (18.29)	0.588 (3.99)	0.797 (3.06)
	Bootstrap PI	0.853 (26.54)	0.563 (3.72)	0.773 (2.85)

Continued on next page

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.921 (31.86)	0.972 (7.01)	0.982 (4.24)
	Wald PI	0.788 (13.22)	0.941 (6.22)	0.959 (3.53)
	MLE-based normal PI	0.810 (15.81)	0.589 (2.93)	0.782 (2.19)
	Bootstrap PI	0.839 (22.14)	0.565 (2.75)	0.759 (2.06)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.923 (30.01)	0.962 (5.97)	0.982 (3.54)
	Wald PI	0.762 (12.48)	0.944 (5.28)	0.956 (2.94)
	MLE-based normal PI	0.788 (14.85)	0.577 (2.48)	0.784 (1.82)
	Bootstrap PI	0.830 (20.18)	0.559 (2.35)	0.766 (1.73)
		$\mu_0 = 10$		
CI for μ	Bayesian CrI	0.931 (42.39)	0.976 (5.79)	0.990 (2.74)
	Wald CI	0.715 (21.47)	0.941 (4.91)	0.948 (1.91)
	MLE-based normal CI	0.765 (25.70)	0.966 (5.65)	0.682 (1.15)
	Bootstrap CI	0.806 (29.73)	0.967 (5.59)	0.681 (1.14)
Y_{new}	Bayesian PI	0.959 (105.19)	0.986 (51.84)	1.000 (36.25)
	Wald PI	0.936 (64.75)	0.947 (35.16)	0.980 (24.03)
	MLE-based normal PI	0.944 (82.85)	0.959 (38.88)	0.704 (15.85)
	Bootstrap PI	0.947 (91.93)	0.979 (51.16)	0.687 (15.71)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.949 (77.56)	0.965 (25.60)	0.990 (17.29)
	Wald PI	0.855 (36.52)	0.954 (21.30)	0.955 (12.42)
	MLE-based normal PI	0.873 (45.35)	0.969 (23.01)	0.711 (7.56)
	Bootstrap PI	0.901 (66.03)	0.963 (25.23)	0.697 (7.27)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.927 (66.06)	0.960 (18.86)	0.986 (12.47)
	Wald PI	0.817 (30.50)	0.951 (16.38)	0.949 (8.85)
	MLE-based normal PI	0.847 (37.34)	0.968 (18.73)	0.703 (5.37)
	Bootstrap PI	0.872 (54.36)	0.965 (18.53)	0.691 (5.22)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.934 (57.35)	0.973 (14.07)	0.990 (9.06)
	Wald PI	0.783 (26.70)	0.942 (12.08)	0.949 (6.38)
	MLE-based normal PI	0.812 (32.29)	0.965 (13.89)	0.694 (3.86)
	Bootstrap PI	0.852 (45.29)	0.965 (13.76)	0.688 (3.78)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.923 (53.48)	0.971 (11.99)	0.989 (7.57)
	Wald PI	0.769 (25.21)	0.949 (10.26)	0.953 (5.32)
	MLE-based normal PI	0.808 (30.33)	0.968 (11.80)	0.689 (3.22)
	Bootstrap PI	0.848 (41.46)	0.967 (11.70)	0.680 (3.16)

Continued on next page

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
		Structured-zero		
		$\mu_0 = 5$		
CI for μ	Bayesian CrI	0.946 (24.82)	0.961 (2.89)	0.988 (1.28)
	Wald CI	0.719 (11.03)	0.942 (2.52)	0.955 (1.05)
	MLE-based normal CI	0.762 (13.07)	0.578 (1.20)	0.779 (0.65)
	Bootstrap CI	0.800 (15.10)	0.578 (1.18)	0.774 (0.64)
Y_{new}	Bayesian PI	0.945 (54.68)	0.978 (25.76)	0.997 (17.01)
	Wald PI	0.919 (33.53)	0.941 (18.20)	0.975 (12.93)
	MLE-based normal PI	0.926 (42.35)	0.814 (11.44)	0.902 (10.07)
	Bootstrap PI	0.927 (46.74)	0.665 (10.69)	0.836 (8.73)
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.938 (41.87)	0.973 (12.78)	0.983 (8.12)
	Wald PI	0.855 (18.82)	0.946 (10.88)	0.966 (6.92)
	MLE-based normal PI	0.865 (23.13)	0.621 (5.42)	0.818 (4.33)
	Bootstrap PI	0.892 (33.05)	0.592 (5.18)	0.790 (3.96)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.961 (36.49)	0.975 (9.40)	0.981 (5.84)
	Wald PI	0.829 (15.70)	0.948 (8.46)	0.965 (4.91)
	MLE-based normal PI	0.846 (19.02)	0.594 (4.06)	0.812 (3.06)
	Bootstrap PI	0.869 (27.52)	0.566 (3.78)	0.787 (2.84)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.945 (32.24)	0.974 (7.01)	0.987 (4.24)
	Wald PI	0.786 (13.73)	0.950 (6.23)	0.960 (3.53)
	MLE-based normal PI	0.813 (16.44)	0.590 (2.98)	0.797 (2.19)
	Bootstrap PI	0.842 (22.98)	0.571 (2.80)	0.775 (2.06)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.950 (30.34)	0.977 (5.98)	0.985 (3.55)
	Wald PI	0.771 (12.96)	0.955 (5.29)	0.962 (2.94)
	MLE-based normal PI	0.797 (15.44)	0.593 (2.53)	0.792 (1.82)
	Bootstrap PI	0.837 (21.02)	0.575 (2.38)	0.779 (1.73)
		$\mu_0 = 10$		
CI for μ	Bayesian CrI	0.921 (40.43)	0.979 (5.84)	0.994 (2.73)
	Wald CI	0.731 (21.80)	0.937 (4.91)	0.959 (1.91)
	MLE-based normal CI	0.774 (26.04)	0.964 (5.65)	0.689 (1.13)
	Bootstrap CI	0.811 (30.16)	0.967 (5.59)	0.689 (1.12)
Y_{new}	Bayesian PI	0.961 (100.28)	0.975 (52.40)	1.000 (36.24)
	Wald PI	0.926 (65.76)	0.941 (35.15)	0.973 (24.03)
	MLE-based normal PI	0.937 (83.93)	0.954 (38.89)	0.801 (15.69)
	Bootstrap PI	0.937 (93.44)	0.978 (51.11)	0.786 (15.47)

Continued on next page

Quantity	Method	$(n, \theta_0) = (20, 0.1)$	$(n, \theta_0) = (100, 1)$	$(n, \theta_0) = (200, 10)$
$\overline{Y}_{5,\text{new}}$	Bayesian PPI	0.947 (73.98)	0.973 (25.79)	0.992 (17.24)
	Wald PI	0.857 (37.09)	0.949 (21.30)	0.967 (12.43)
	MLE-based normal PI	0.875 (45.95)	0.963 (23.02)	0.726 (7.43)
	Bootstrap PI	0.906 (66.73)	0.972 (25.20)	0.713 (7.15)
$\overline{Y}_{10,\text{new}}$	Bayesian PPI	0.930 (62.93)	0.975 (19.03)	0.997 (12.43)
	Wald PI	0.818 (30.96)	0.948 (16.37)	0.961 (8.85)
	MLE-based normal PI	0.848 (37.83)	0.967 (18.74)	0.726 (5.28)
	Bootstrap PI	0.874 (54.88)	0.967 (18.52)	0.714 (5.14)
$\overline{Y}_{20,\text{new}}$	Bayesian PPI	0.935 (54.53)	0.974 (14.18)	0.991 (9.04)
	Wald PI	0.796 (27.10)	0.939 (12.07)	0.967 (6.38)
	MLE-based normal PI	0.829 (32.71)	0.961 (13.90)	0.715 (3.80)
	Bootstrap PI	0.865 (46.04)	0.963 (13.78)	0.705 (3.72)
$\overline{Y}_{30,\text{new}}$	Bayesian PPI	0.929 (50.91)	0.974 (12.08)	0.989 (7.55)
	Wald PI	0.771 (25.59)	0.943 (10.25)	0.958 (5.32)
	MLE-based normal PI	0.811 (30.72)	0.967 (11.80)	0.704 (3.16)
	Bootstrap PI	0.855 (42.00)	0.967 (11.72)	0.697 (3.11)

Note: CP is operational coverage, with failed intervals counted as noncoverage. EW is the average interval width among successful intervals. The methods labelled as likelihood CI/PI in the raw frequentist output are reported here as MLE-based normal CI/PI. Bayesian rows are placed first and use the updated Bayesian simulation results.

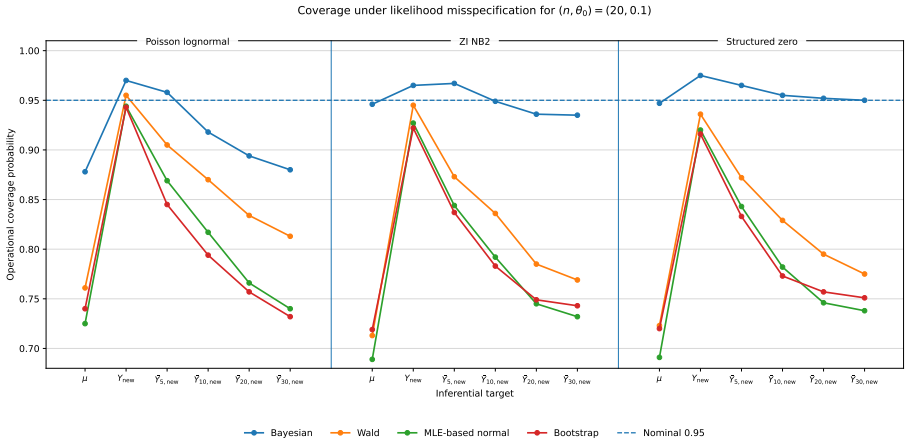


Figure S2. Operational coverage of the 95% interval procedures under likelihood misspecification for $\mu_0 = 1$ and $(n, \theta_0) = (20, 0.1)$. The three regions represent the Poisson–lognormal, ZI–NB2, and structured-zero mechanisms. The dashed line marks the nominal level of 0.95; exact values are reported in Table 3.

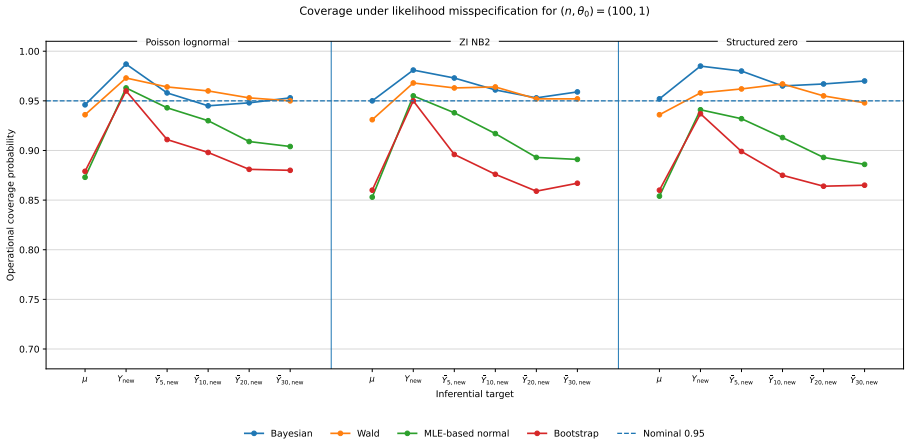


Figure S3. Operational coverage of the 95% interval procedures under likelihood misspecification for $\mu_0 = 1$ and $(n, \theta_0) = (100, 1)$. The three regions represent the Poisson–lognormal, ZI–NB2, and structured-zero mechanisms. The dashed line marks the nominal level of 0.95; exact values are reported in Table 3.

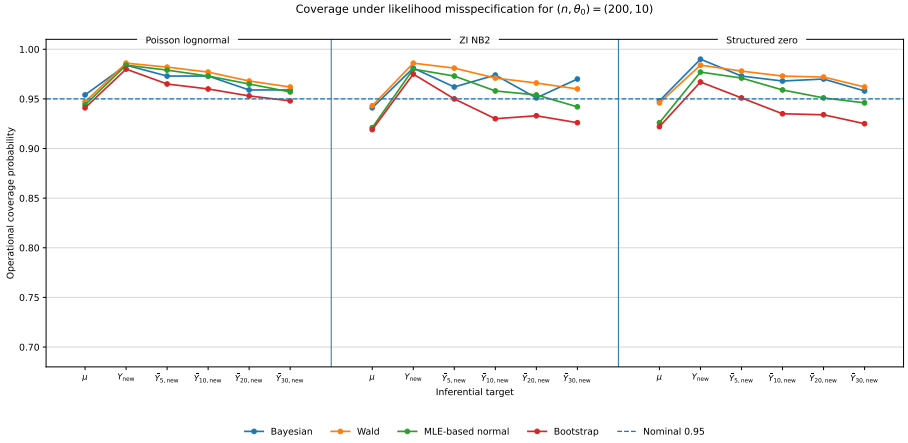


Figure S4. Operational coverage of the 95% interval procedures under likelihood misspecification for $\mu_0 = 1$ and $(n, \theta_0) = (200, 10)$. The three regions represent the Poisson–lognormal, ZI–NB2, and structured-zero mechanisms. The dashed line marks the nominal level of 0.95; exact values are reported in Table 3.