

Post-Hoc Analysis of Confidence Intervals

Impact of Erroneous Prevalence Assumptions on Study Precision

Kulig, Schäfer, Wilczek, Lange, Hensel, Jung

2026-10-15

Table of contents

1	Methodological Context of the Post-Hoc Analysis	2
1.1	The Problem of Binomial Variance	2
1.2	Consequences for the Confidence Interval	2
1.3	Reading Guide for the Tables	3
1.4	Brief Description of the Methods	3
2	Wald / Cochran (Normal Approximation)	4
3	Wilson Score Interval	5
4	Clopper-Pearson (Exact Method)	6
5	Bayesian Credible Interval	7
6	References	8

1 Methodological Context of the Post-Hoc Analysis

This report evaluates how erroneous planning assumptions, or sample sizes adopted without justification, affect the final precision and informative value of a prevalence study when the prevalence actually observed in the field deviates from the assumed value. It is part of a framework for sample size planning of binomially distributed traits (Kulig et al., 2026).

1.1 The Problem of Binomial Variance

In sample size planning for prevalence studies, the assumed prevalence is the most critical input. Together with the confidence level and the targeted margin of error (MoE), it determines the sample size (n) required to keep the estimate within the desired precision (Cochran, 1977; Dohoo et al., 2003; Thrusfield et al., 2018).

The variance of a binomial proportion, $p(1 - p)/n$, is not linear in p . It is smallest near the boundaries (0 % or 100 %) and reaches its maximum at exactly 50 % (Agresti, 2007). A sample size planned for a low prevalence therefore yields a wider interval when the observed prevalence turns out to be closer to 50 %.

In practice, for example in on-farm animal health monitoring, sample sizes are frequently adopted from the literature or calculated with standard formulas for a low assumed prevalence (e.g. 5 %), because the formula then returns an economically attractive, small sample size (Humphry et al., 2004; Lenth, 2001). If a considerably higher prevalence (e.g. 20 %) is observed during the subsequent field survey, the fixed sample size is no longer sufficient to compensate for the increased variance.

1.2 Consequences for the Confidence Interval

As a direct consequence of this planning error, the confidence interval widens after data collection. The resulting MoE exceeds the value that was required when the study was approved or planned, and the study loses precision.

This report simulates exactly this situation. It takes the **fixed sample size** (n) and analyses retrospectively (post hoc), using four statistical interval methods, how far the confidence interval widens at this sample size when the true prevalence in the herd varies.

Finite population correction. Because the sample is drawn without replacement from a finite herd of $N = 200$ animals, all methods account for the finite population correction (FPC), $(N - n)/(N - 1)$ (Cochran, 1977; Kish, 1965).

Cluster effects. Cluster effects are not taken into account (simple random sampling, $D_{eff} = 1$). If the study was planned with cluster sampling (Tool A), cluster sampling should be activated in the configuration, because the precision is otherwise overestimated (Dohoo et al., 2003; Kish, 1965).

Effective sample size. The FPC and the design effect both act on the variance of the estimate, $\text{Var}(\hat{p}) = D_{eff} \hat{p}(1 - \hat{p})/n \cdot (N - n)/(N - 1)$. All four methods therefore use the effective sample size

$$n_{\text{eff}} = \frac{n(N - 1)}{D_{eff}(N - n)} \quad (\text{with FPC}), \quad n_{\text{eff}} = \frac{n}{D_{eff}} \quad (\text{without FPC}),$$

which is the exact inverse of the planning formula used in Tools A and B of the framework. Here $n_{\text{eff}} = 464.3$ for simple random sampling.

1.3 Reading Guide for the Tables

The following tables show the simulated evaluation for the fixed sample size. The column “**Resulting MoE**” gives the precision actually achieved with this sample size at the respective observed prevalence, expressed as half the width of the interval in percentage points (pp).

Note: **Red values** indicate that the MoE defined during planning (± 2 ,pp) has been exceeded. In these scenarios the study no longer delivers results with the originally required precision.

1.4 Brief Description of the Methods

Wald / Cochran (normal approximation): The Wald interval, $\hat{p} \pm z \sqrt{\hat{p}(1 - \hat{p})/n}$, is the most widely used method for sample size and interval calculation (Cochran, 1977; Thrusfield et al., 2018). It goes back to Laplace (1812) and is named after the Wald test (Wald, 1943), from whose inversion the interval results (Agresti & Coull, 1998). It relies on a normal approximation that assumes a symmetric sampling distribution of the estimated proportion. For small samples and proportions close to 0 or 1, this simplification leads to coverage that is considerably below the nominal level and to bounds outside $[0, 1]$; even for moderate proportions the actual coverage fluctuates erratically (Brown et al., 2001; Newcombe, 1998). The method is therefore not recommended for prevalences near the boundaries.

Wilson score interval: The Wilson score interval is derived by inverting the score test instead of the Wald test (Wilson, 1927). It shifts the centre of the interval slightly towards 50 %, yields asymmetric bounds near the boundaries and never produces limits below 0 % or above 100 %. Comparative studies have shown that its actual coverage is close to the nominal level across the whole range of proportions, which is why it is recommended as the default approximate method (Agresti, 2007; Brown et al., 2001; Newcombe, 1998).

Clopper-Pearson (exact method): The Clopper-Pearson interval avoids any approximation and derives the bounds directly from the binomial distribution, expressed through quantiles of the Beta distribution (Clopper & Pearson, 1934). It guarantees a coverage of at least the nominal level for every true proportion and is therefore strictly conservative. In practice, this guarantee results in wider intervals than the approximate methods, so that planning with this method requires the largest sample sizes (Brown et al., 2001; Newcombe, 1998).

Bayesian credible interval: The Bayesian approach computes credible intervals and allows existing knowledge to be integrated in a structured way (Gelman et al., 2013; Kruschke & Liddell, 2018). With a Beta prior $\text{Beta}(\alpha_0, \beta_0)$ and x positive animals among n , the posterior is $\text{Beta}(\alpha_0 + x, \beta_0 + n - x)$, from whose quantiles the equal-tailed interval is obtained; α_0 and β_0 can be interpreted as the number of positive and negative animals in previous studies. With a flat, non-informative prior ($\text{Beta}(1, 1)$) the interval for moderate sample sizes behaves very similarly to the frequentist methods (Gelman et al., 2013). An informative prior pulls the interval towards the prior mean and narrows it, so that the interval is no longer centred on the observed prevalence; the resulting MoE is then reported as half the interval width. Bayesian approaches to sample size planning for binomial proportions are described by Joseph et al. (1995) and Joseph et al. (2008).

2 Wald / Cochran (Normal Approximation)

Applied study design:

n = 140 animals (of N = 200)

Assumed prevalence: 5 % | Planning target: 95 % CI with ± 2 pp MoE

Observed prev. (%)	CI lower bound (%)	CI upper bound (%)	Resulting MoE ($\pm$ pp)
1.00	0.09	1.91	0.91
2.50	1.08	3.92	1.42
5.00	3.02	6.98	1.98
10.00	7.27	12.73	2.73
15.00	11.75	18.25	3.25
20.00	16.36	23.64	3.64
25.00	21.06	28.94	3.94
30.00	25.83	34.17	4.17
35.00	30.66	39.34	4.34
40.00	35.54	44.46	4.46
45.00	40.47	49.53	4.53
50.00	45.45	54.55	4.55

3 Wilson Score Interval

Applied study design:

n = 140 animals (of N = 200)

Assumed prevalence: 5 % | Planning target: 95 % CI with ± 2 pp MoE

Observed prev. (%)	CI lower bound (%)	CI upper bound (%)	Resulting MoE ($\pm$ pp)
1.00	0.42	2.39	0.99
2.50	1.42	4.36	1.47
5.00	3.36	7.38	2.01
10.00	7.59	13.07	2.74
15.00	12.04	18.53	3.25
20.00	16.61	23.88	3.63
25.00	21.28	29.13	3.93
30.00	26.01	34.32	4.15
35.00	30.80	39.45	4.32
40.00	35.64	44.52	4.44
45.00	40.53	49.55	4.51
50.00	45.47	54.53	4.53

4 Clopper-Pearson (Exact Method)

Applied study design:

n = 140 animals (of N = 200)

Assumed prevalence: 5 % | Planning target: 95 % CI with ± 2 pp MoE

Observed prev. (%)	CI lower bound (%)	CI upper bound (%)	Resulting MoE ($\pm$ pp)
1.00	0.35	2.49	1.07
2.50	1.34	4.47	1.56
5.00	3.17	7.34	2.09
10.00	7.34	12.99	2.82
15.00	11.94	18.66	3.36
20.00	16.48	23.96	3.74
25.00	21.11	29.18	4.04
30.00	25.80	34.33	4.26
35.00	30.76	39.64	4.44
40.00	35.57	44.67	4.55
45.00	40.42	49.66	4.62
50.00	45.32	54.61	4.64

5 Bayesian Credible Interval

Applied study design:

n = 140 animals (of N = 200)

Assumed prevalence: 5 % | Planning target: 95 % CI with ± 2 pp MoE

Prior used: Beta(25, 75)

Observed prev. (%)	CI lower bound (%)	CI upper bound (%)	Resulting MoE ($\pm$ pp)
1.00	3.62	7.31	1.84
2.50	4.67	8.74	2.04
5.00	6.35	10.94	2.30
10.00	9.98	15.44	2.73
15.00	13.86	20.03	3.08
20.00	17.66	24.36	3.35
25.00	21.50	28.64	3.57
30.00	25.39	32.87	3.74
35.00	29.49	37.25	3.88
40.00	33.44	41.42	3.99
45.00	37.43	45.55	4.06
50.00	41.45	49.66	4.10

6 References

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed., p. 372). John Wiley; Sons. <https://doi.org/10.1002/0470114754>
- Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119–126. <https://doi.org/10.2307/2685469>
- Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2). <https://doi.org/10.1214/ss/1009213286>
- Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), 404. <https://doi.org/10.2307/2331986>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley; Sons, Inc.
- Dohoo, I. Robert., Martin, S. Wayne., & Stryhn, Henrik. (2003). *Veterinary epidemiologic research* (p. 706). University of Prince Edward Island. <https://islandscholar.ca/islandora/object/ir%3Air-batch6-2657>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman; Hall/CRC. <https://doi.org/10.1201/b16018>
- Humphry, R. W., Cameron, A., & Gunn, G. J. (2004). A practical approach to calculate sample size for herd prevalence surveys. *Preventive Veterinary Medicine*, 65(3-4), 173–188. <https://doi.org/10.1016/j.prevetmed.2004.07.003>
- Joseph, L., M’Lan, C. E., & Wolfson, D. B. (2008). Bayesian sample size determination for binomial proportions. *Bayesian Analysis*, 3(2). <https://doi.org/10.1214/08-BA310>
- Joseph, L., Wolfson, D. B., & Du Berger, R. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *The Statistician*, 44(2), 143. <https://doi.org/10.2307/2348439>
- Kish, L. (1965). *Survey sampling* (p. 643). John Wiley; Sons, Inc.
- Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin and Review*, 25(1), 155–177. <https://doi.org/10.3758/s13423-017-1272-1>
- Kulig, B., Schäfer, B., Wilczek, U., Lange, A., Hensel, O., & Jung, L. (2026). *Statistical Framework not only for Precision Livestock Management and Ethology, but for performing Sample Size Calculation for Binomially Distributed Traits, Incorporating Simple Sample Size Planning, Post-Hoc Analysis, Confidence Interval Inflation, Baye*. DaKS, Universität Kassel.
- Laplace, P.-S. (1812). *Théorie analytique des probabilités*. Courcier.
- Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3), 187–193. <https://doi.org/10.1198/000313001317098149>
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17(8), 857–872. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8%3C857::AID-SIM777%3E3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8%3C857::AID-SIM777%3E3.0.CO;2-E)
- Thrusfield, M., Christley, R., Brown, H., Diggle, P. J., French, N., Howe, K., Kelly, L., O’Connor, A., Wood, H., & Sargeant, J. (2018). *Veterinary epidemiology* (4th ed.). Wiley-Blackwell.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426–482. <https://doi.org/10.1090/S0002-9947-1943-0012401-3>
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209. <https://doi.org/10.2307/2276774>