A Retrospective Look at Covid-19 Testing Performance

June 25, 2022

Abstract

In this research paper, we analyze how the sensitivity (the likelihood that those with the disease test positive), specificity (the likelihood that those without the disease test negative), and disease prevalence affect the Covid-19 testing performance. We use a set of different sensitivity and specificity values with a range of prevalence values to evaluate test performance in terms of positive predictive value (PPV, the likelihood that those who test positive have the disease) and negative predictive value (NPV, the likelihood that those who test negative do not have the disease). Our analysis indicates that the disease prevalence has a significant impact on the sensitivity and specificity values necessary to achieve high PPV and NPV. Furthermore, we find that a prevalence of less than 0.01 (1%) would require an unrealistically high specificity value to achieve a strong PPV.

1. **Background and Research Objective**

   Covid-19 testing has become a regular part of life over the past few years and yet, there is still confusion and debate about the results of these tests and how they are presented at large. For example, public concern continues over issues of false positives and negatives. At the same time, a firm understanding of test results would allow hospitals, frontline workers, and healthcare authorities to effectively allocate their resources.

   In this paper, we examine the effectiveness of Covid-19 testing in mid-2020 by computing the likelihoods of true positive and negative results. These values are known as the positive and negative predictive value (PPV and NPV). PPV measures how likely it is that someone who tests positive has the disease (a true positive), and NPV measures the likelihood of someone who tests negative not having the disease (a true negative). The PPV and NPV calculation require the prevalence (p) of Covid-19, sensitivity (the probability that those who have the disease actually test positive), and specificity (the probability that those who do not have the disease actually test negative).

   To obtain the estimated prevalence ($\hat{p}$) of Covid-19 in mid-2020, we examined data from the Pfizer-BioNTech vaccine study [3]. The study employs nucleic acid amplification testing (NAAT), the most sensitive form of testing available to test for Covid-19 contraction [5], providing a more accurate means of assessing prevalence than the publicly available tests. By focusing on the placebo group from the vaccine study ($n = 21728$), we obtained $\hat{p} = 162/21728 = 0.00746$ (0.746%) [3]. Note that this is representative of the general population since vaccines were not widely available in mid-2020.

2. **Methods**

   a. **Predictive Values**

   PPV and NPV are the proportions of positive and negative results that are true positives and true negatives. Let D denote the event of an individual having the disease, and let N denote the event of an individual testing negative for the disease. Then, PPV $= P(D|N')$ and NPV $= P(D'|N)$. Now, let α and β denote assumed sensitivity and specificity, respectively. Then, by [2], estimated PPV and NPV, denoted by $\hat{p}_{ppv}$ and $\hat{p}_{npv}$, respectively, are given by:

$$\hat{p}_{ppv} = \frac{\alpha\hat{p}}{\alpha\hat{p} + (1-\beta)(1-\hat{p})} \quad \text{and} \quad \hat{p}_{npv} = \frac{\beta(1-\hat{p})}{(1-\alpha)\hat{p} + \beta(1-\hat{p})}.$$

   Note that $\hat{p}_{ppv}$ and $\hat{p}_{npv}$ are both functions of estimated disease prevalence $\hat{p}$.

   b. **Confidence Intervals for PPV and NPV**

   Confidence intervals give us a range of plausible values for the parameters of interest, given a specific level of desired confidence. To obtain 95% confidence intervals for PPV and NPV, we apply the delta method. Specifically, for a large n, a 95% confidence interval for a function of p, denoted by $g(p)$, is given by:

$$g(\hat{p}) \pm 1.96\sqrt{\frac{[g'(\hat{p})]^2\hat{p}(1-\hat{p})}{n}},$$

   assuming $g'(p)$ exists and $g'(p) \neq 0$. A direct application of the formula above gives us 95% confidence intervals for PPV and NPV as follows:

$$\hat{p}_{ppv} \pm 1.96\sqrt{\frac{\alpha^2(1-\beta)^2\hat{p}(1-\hat{p})}{n[\alpha\hat{p} + (1-\beta)(1-\hat{p})]^4}} \text{ and } \hat{p}_{npv} \pm 1.96\sqrt{\frac{(1-\alpha)^2\beta^2\hat{p}(1-\hat{p})}{n[(1-\alpha)\hat{p} + \beta(1-\hat{p})]^4}}.$$

By following the recommended values of $\alpha = 0.95$ and $\beta = 0.98$ for a screening indication [1], using the estimated disease prevalence of $\hat{p} = 0.00746$, $\hat{p}_{ppv} = 0.2630$ with its 95% confidence interval $[0.2330, 0.2929]$, and $\hat{p}_{npv} = 0.9996$ with its 95% confidence interval $[0.9996, 0.9997]$ (see Appendix B for more details about the derivation and calculation of confidence intervals). Therefore, while a negative test result almost certainly implies the absence of the disease, only about one out of four positive test results indicates the presence of the disease under the given sensitivity, specificity, and estimated disease prevalence.

3. **Analysis and Insights**

The estimated PPV and NPV ($\hat{p}_{ppv}$ and $\hat{p}_{npv}$) along with their confidence intervals can be used to estimate the true number of positive and negative cases among those tested positive and negative. We specifically utilize weekly Covid-19 testing data from Washington State in the United States between February 29, 2020 and September 25, 2020 [4].

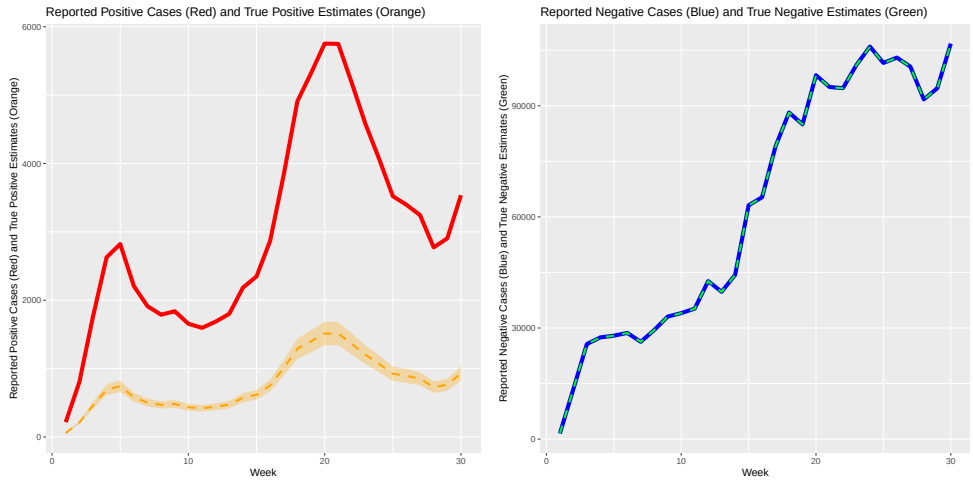

Figure 1: **Left:** Graph of weekly reported positive cases in Washington State (in red) [4] and estimated true positive cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.00746$ (in orange). Pointwise 95% confidence intervals for the true positive cases are also displayed (in shades). **Right:** Graph of weekly reported negative cases in Washington State (in blue) [4] and estimated true negative cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.00746$ (in green). Pointwise 95% confidence intervals for the true negative cases are also displayed (in shades).

Figure 1 compares reported Covid-19 test results to the estimated portions of those tests with true test results. The low $\hat{p}_{ppv}$ (of about 0.2630) implies that the visible peak in the reported positive cases at around Week 20 (mid-July) is very much flattened in the estimated true positive cases. On the other hand, the estimated true negative cases is nearly identical to the reported negative cases due to the high $\hat{p}_{npv}$ (of about 0.9996). Also, while the proportion of false positives is much higher than that of false negatives, there are overall far fewer reported positive cases, as reflected in the difference in the scales in these two graphs of Figure 1.

In summary, the recommended sensitivity and specificity of $\alpha = 0.95$ and $\beta = 0.98$ for a screening indication provides an accurate assignment of reported negative results, but also frequently mislabel true negative results as positive when the disease prevalence is low (e.g., $p = 0.00746$). This is partly a result of the selected sensitivity and selectivity values, and partly a result of the low prevalence of the disease. In the following, we examine how each of these variables affects the overall accuracy of various testing methods, as measured by PPV and NPV.

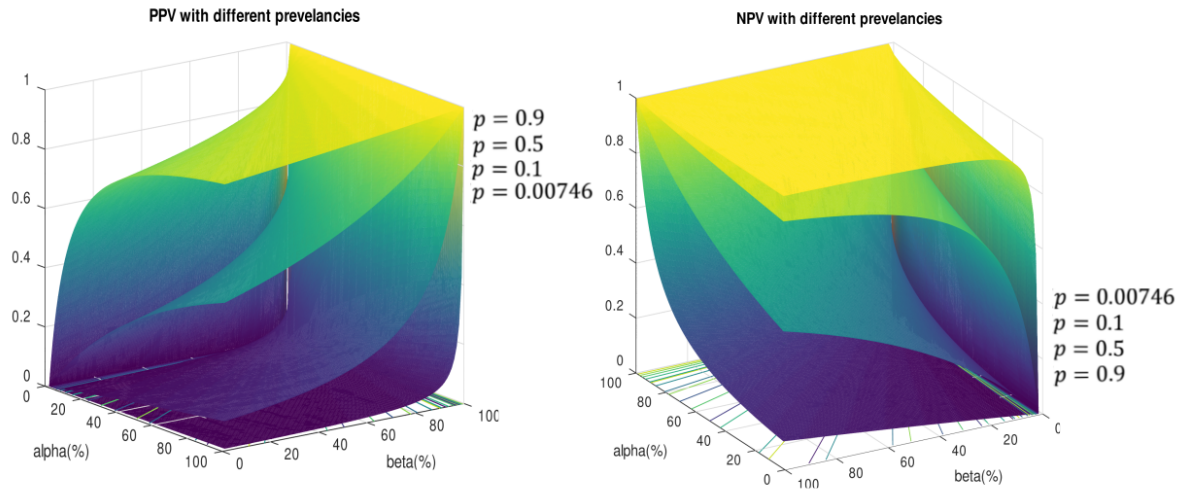

Figure 2: **Left:** PPV as a function of α (along the x-axis) and β (along the y-axis) with surfaces corresponding to different prevalence values, as shown by the legend on the right.
Right: NPV as a function of α (along the x-axis) and β (along the y-axis) with surfaces corresponding to different prevalence values, as shown by the legend on the right.

Figure 2 above clearly shows that as disease prevalence increases, PPV also increases while NPV decreases. Generally, a lower PPV and a higher NPV are associated with a lower prevalence. Moreover, regardless of disease prevalence p and sensitivity α, we observe in the left graph that as $\beta \to 1(100\%)$ (higher values of specificity as we move towards right on the front axis), PPV $\to 1$ as well. Conversely, we see in the right graph that NPV is most sensitive to changes in sensitivity α than specificity β, and as $\alpha \to 1(100\%)$, NPV $\to 1$. Together, these relationships guarantee that increased standards of sensitivity α and specificity β always result in higher accuracy for both PPV and NPV, regardless of disease prevalence. However, for the case that we have examined where $p = 0.00746$, prevalence is so low that specificity β would need to be increased to approximately 0.9999 to constitute near-perfect accuracy in PPV of 0.9862. As such standards are unrealistic in practice, we would have to accept some degree of false positives with our testing results.

4. **Significance and Limitations**

Our analysis implies that most people who tested negative did not have Covid-19 in mid-2020 and, perhaps more importantly, approximately three quarters of the people who tested positive also did not have Covid-19. That shows that Covid-19 testing is not a perfect science even with high sensitivity and specificity when the disease prevalence is low. In other words, a positive result does not necessarily indicate Covid-19 contraction.

However, our results need to be interpreted with caution as the study population may not fully represent the target population and the parameters used may not accurately represent the reality. For example, those who volunteered for the vaccine study [4] are likely biased towards healthier individuals, possibly overlooking at-risk individuals with co-morbidities. On the other hand, the actual sensitivity and specificity may be lower partly due to issues surrounding sample collection, handling, and analysis [1]. Furthermore, as these volunteers were recruited from multiple regions, directly applying the results to the Washington State data may be inappropriate due to the differences in health policies, overall health and well-being of the study and target populations, and demographics in general. For the future work, we are hoping to re-conduct a similar study with different U.S. state populations by carefully avoiding these issues. The follow-up study could be used to assess the vaccine efficacy or to evaluate the impact of different variants.

References

1. Braunstein, G.D., Schwartz, L, Hymel, P., Fielding, J. (2021). False positive results with SARS- CoV-2 RT-PCR tests and how to evaluate a RT-PCR-positive test for the possibility of a false positive result, Journal of Occupational and Environmental Medicine, 63(3), e159-e162.

2. Casselman, B. (2020). Does He Have It? Sensitivity, Specificity, and COVID-19 Testing. American Mathematical Society. https://www.ams.org/publicoutreach/feature-column/fc-2020-09

3. Polack, F.P. et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine, The New England Journal of Medicine, 383(27), 2603-2615.

4. Washington State Department of Health (2022). COVID-19 Data Dashboard. Data retrieved on February 9, 2022. https://www.doh.wa.gov/Emergencies/COVID19/DataDashboard

5. Centers For Disease Control and Prevention (2021). Nucleic Acid Amplification Tests (NAATs). Page accessed on May 31, 2022. https://www.cdc.gov/coronavirus/2019-ncov/lab/naats.html

Appendix A. Secondary Analysis with Varying Estimated Disease Prevalence Values

Figure 1 contrasts reported positive and negative Covid-19 cases with estimated true positive and negative cases along with their 95% confidence intervals. These estimates are calculated using the estimated disease prevalence of $\hat{p} = 162/21728 \approx 0.00746$, which is considered low, with $\alpha = 0.95$, $\beta = 0.98$, and $n = 21728$. The following plots display the same data from the Washington State Department of Health [4], but use different estimated prevalence values ($\hat{p} = 0.1, 0.5, 0.9$) in the calculation of estimated true positive and negative cases along with their 95% confidence intervals while keeping $\alpha = 0.95$, $\beta = 0.98$, and $n = 21728$ as before. These additional case studies allow us to extend our analysis of the effects of various prevalence values on testing results.

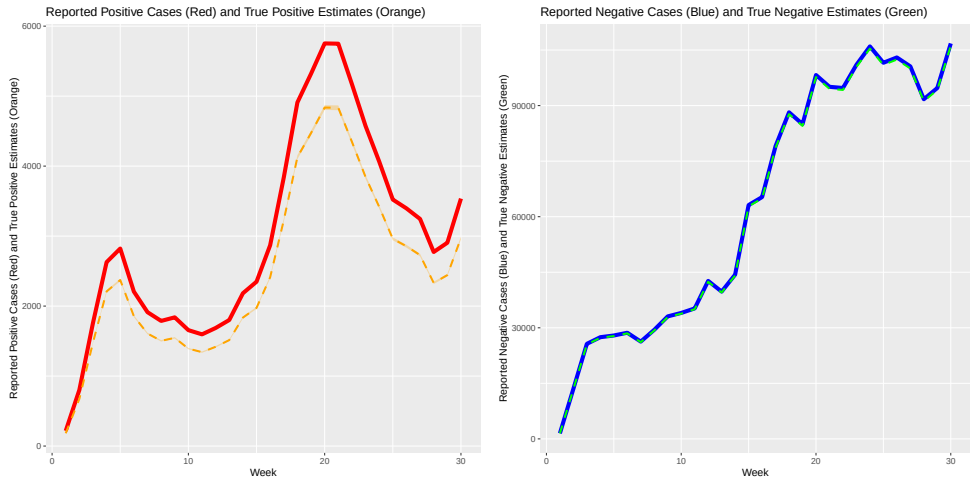

Figure 3: **Left:** Graph of weekly reported positive cases in Washington State (in red) [4] and estimated true positive cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.1$ (in orange). Pointwise 95% confidence intervals for the true positive cases are also displayed (in shades).
Right: Graph of weekly reported negative cases in Washington State (in blue) [4] and estimated true negative cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.1$ (in green). Pointwise 95% confidence intervals for the true negative cases are also displayed (in shades).

Figure 3 shows that even a relatively low estimated prevalence of $\hat{p} = 0.1$ results in a significant jump in the number of estimated true positive cases and a much narrower confidence interval (compared to Figure 1). The gap between the reported positive cases and estimated positive cases continues to decrease as disease prevalence increases. When $\hat{p} \geq 0.5$, nearly all the reported positive cases can be considered true positive cases according to Figures 4–5.

On the other hand, the number of estimated negative cases decreases as disease prevalence increases. Compared to the right plot of Figure 3, the margin of separation between the reported and estimated negative cases gets proportionally larger when $\hat{p} = 0.5$ (see the right plot of Figure 4). The trend becomes very obvious when $\hat{p} = 0.9$ where we see a clear deviation between the reported negative cases and estimated negative cases (see the right plot of Figure 5). This is due to differences between the formulas for PPV and NPV estimation ($\hat{p}_{ppv}$ and $\hat{p}_{npv}$), as discussed in Section 2a, which in turn affect the confidence interval calculations in Section 2b. For a more detailed look at confidence interval derivations, see Appendix B.

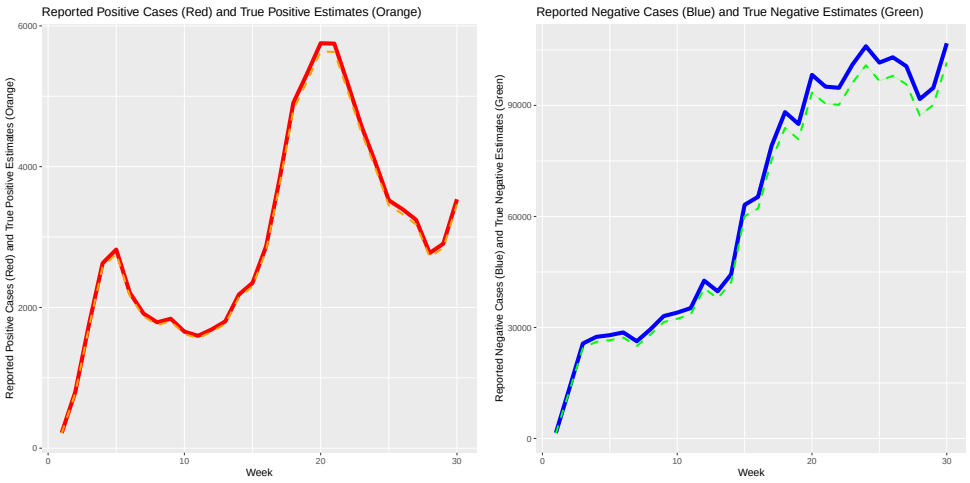

Figure 4: **Left:** Graph of weekly reported positive cases in Washington State (in red) [4] and estimated true positive cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.5$ (in orange). Pointwise 95% confidence intervals for the true positive cases are also displayed (in shades).
Right: Graph of weekly reported negative cases in Washington State (in blue) [4] and estimated true negative cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.5$ (in green). Pointwise 95% confidence intervals for the true negative cases are also displayed (in shades).

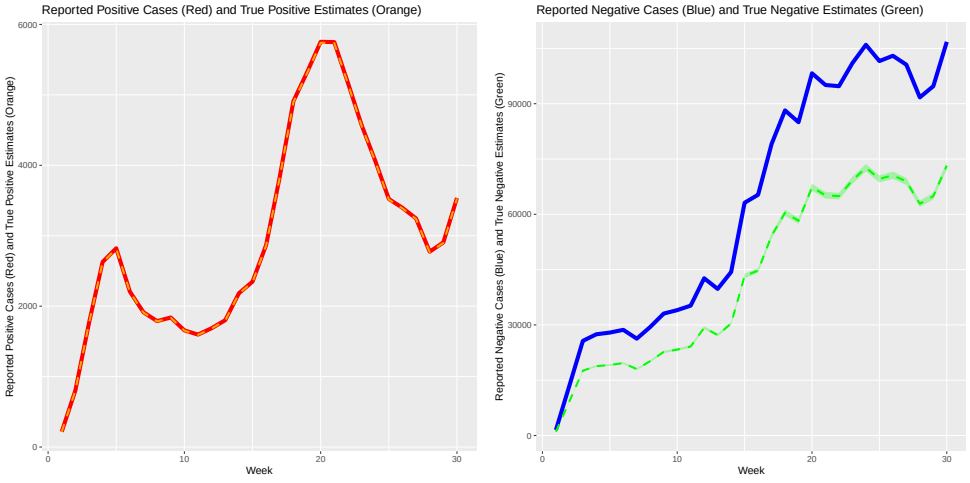

Figure 5: **Left:** Graph of weekly reported positive cases in Washington State (in red) [4] and estimated true positive cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.9$ (in orange). Pointwise 95% confidence intervals for the true positive cases are also displayed (in shades).
Right: Graph of weekly reported negative cases in Washington State (in blue) [4] and estimated true negative cases assuming $\alpha = 0.95$, $\beta = 0.98$, and $\hat{p} = 0.9$ (in green). Pointwise 95% confidence intervals for the true negative cases are also displayed (in shades).

Appendix B. Derivation of PPV and NPV Confidence Intervals

In Section 2b, we apply the delta method to calculate confidence intervals for PPV and NPV. Recall that the delta method requires the derivative of the function of the parameter of interest. Since NPV and PPV in Section 2a are both functions of p, their derivatives with respect to p can be calculated.

B1. Derivative of PPV and NPV with Respect to p

Let

$$p_{ppv} = \frac{\alpha p}{\alpha p + (1-\beta)(1-p)} = \frac{\alpha p}{\alpha p + \beta p - p - \beta + 1}.$$

Then,

$$\begin{aligned}
\frac{dp_{ppv}}{dp} &= \frac{(\alpha p + \beta p - p - \beta + 1)\alpha - \alpha p(\alpha + \beta - 1)}{[\alpha p + (1-\beta)(1-p)]^2} \\
&= \frac{\alpha^2 p + \alpha\beta p - \alpha p - \alpha\beta + \alpha - (\alpha^2 p + \alpha\beta p - \alpha p)}{[\alpha p + (1-\beta)(1-p)]^2} \\
&= \frac{\alpha(1-\beta)}{[\alpha p + (1-\beta)(1-p)]^2}.
\end{aligned}$$

Similarly, let

$$p_{npv} = \frac{\beta(1-p)}{(1-\alpha)p + \beta(1-p)} = \frac{\beta - \beta p}{p - \alpha p + \beta - \beta p}.$$

Then,

$$\begin{aligned}
\frac{dp_{npv}}{dp} &= \frac{(p - \alpha p + \beta - \beta p)(-\beta) - (\beta - \beta p)(1 - \alpha - \beta)}{[(1-\alpha)p + \beta(1-p)]^2} \\
&= \frac{-\beta p - \alpha\beta p - \beta^2 p - (\beta - \beta\alpha - \beta^2 - \beta p + \alpha\beta p - \beta^2 p)}{[(1-\alpha)p + \beta(1-p)]^2} \\
&= -\frac{(1-\alpha)\beta}{[(1-\alpha)p + \beta(1-p)]^2}.
\end{aligned}$$

B2. PPV and NPV Confidence Interval Derivation Using the Delta Method

By the delta method, these derivatives can then be used to construct confidence intervals for PPV and NPV using the formula given in Section 2b:

$$g(\hat{p}) \pm 1.96\sqrt{\frac{[g'(\hat{p})]^2 \hat{p}(1-\hat{p})}{n}}.$$

By letting $g(\hat{p}) = \hat{p}_{ppv}$, we have

$$\hat{p}_{ppv} \pm 1.96\sqrt{\frac{\left[\frac{\alpha(1-\beta)}{(\alpha\hat{p} + (1-\beta)(1-\hat{p}))^2}\right]^2 \hat{p}(1-\hat{p})}{n}}$$

as a 95% confidence interval for PPV. After rearranging and simplifying, we have:

$$\hat{p}_{ppv} \pm 1.96\sqrt{\frac{\alpha^2(1-\beta)^2\hat{p}(1-\hat{p})}{n[\alpha\hat{p} + (1-\beta)(1-\hat{p})]^4}}.$$

Similarly, by letting $g(\hat{p}) = \hat{p}_{npv}$, we have

$$\hat{p}_{npv} \pm 1.96\sqrt{\frac{\left[-\frac{(1-\alpha)\beta}{((1-\alpha)\hat{p}+\beta(1-\hat{p}))^2}\right]^2 \hat{p}(1-\hat{p})}{n}}$$

as a 95% confidence interval for NPV. After rearranging and simplifying, we have:

$$\hat{p}_{npv} \pm 1.96\sqrt{\frac{(1-\alpha)^2\beta^2\hat{p}(1-\hat{p})}{n[(1-\alpha)\hat{p}+\beta(1-\hat{p})]^4}}.$$

Using $\hat{p} = 165/21728$ which we calculated previously, we can compute 95% confidence intervals for PPV and NPV for various α and β values. For $\alpha = 0.95$ and $\beta = 0.98$, we obtain the following PPV and NPV estimates:

$$\hat{p}_{ppv} = \frac{0.95(\frac{162}{21728})}{0.95(\frac{162}{21728}) + (1-0.98)(1-\frac{162}{21728})} \approx 0.26298$$

and

$$\hat{p}_{npv} = \frac{0.98(1-\frac{162}{21728})}{(1-0.95)(\frac{162}{21728}) + 0.98(1-\frac{162}{21728})} \approx 0.99962.$$

Additionally, we have:

$$\frac{dp_{ppv}}{dp}\Big|_{p=\hat{p}} = \frac{0.95(1-0.98)}{[0.95(\frac{162}{21728}) + (1-0.98)(1-\frac{162}{21728})]^2} \approx 26.19116$$

and

$$\frac{dp_{npv}}{dp}\Big|_{p=\hat{p}} = -\frac{(1-0.95)0.98}{[(1-0.95)(\frac{162}{21728}) + 0.98(1-\frac{162}{21728})]^2} \approx -0.05175.$$

Thus, the 95% confidence interval for PPV is

$$\begin{aligned} & 0.26298 \pm 1.96\sqrt{\frac{(26.19116)^2(\frac{162}{21728}(1-\frac{162}{21728}))}{21728}} \\ = \; & 0.26298 \pm 1.96\sqrt{\frac{685.977(0.00740)}{21728}} \\ = \; & 0.26298 \pm 0.02999 \\ = \; & [0.23302, 0.29294]. \end{aligned}$$

Similarly, the 95% confidence interval for NPV is

$$\begin{aligned} & 0.99962 \pm 1.96\sqrt{\frac{(-0.05175)^2(\frac{162}{21728}(1-\frac{162}{21728}))}{21728}} \\ = \; & 0.99962 \pm 1.96\sqrt{\frac{0.00268(0.00740)}{21728}} \\ = \; & 0.99962 \pm 0.0000592 \\ = \; & [0.99956, 0.99968]. \end{aligned}$$