Numeracy
Advancing Education in Quantitative Literacy

Volume 14 | Issue 1 Article 7

2021

Using COVID-19 Vaccine Efficacy Data to Teach One-Sample Hypothesis Testing

Frank Wang
LaGuardia Community College, CUNY, fwang@lagcc.cuny.edu

Recommended Citation

Wang, Frank. "Using COVID-19 Vaccine Efficacy Data to Teach One-Sample Hypothesis Testing." *Numeracy* 14, Iss. 1 (2021): Article 7. DOI: https://doi.org/10.5038/1936-4660.14.1.1383

Using COVID-19 Vaccine Efficacy Data to Teach One-Sample Hypothesis Testing

Abstract

In late November 2020, there was a flurry of media coverage of two companies' claims of 95% efficacy rates of newly developed COVID-19 vaccines, but information about the confidence interval was not reported. This paper presents a way of teaching the concept of hypothesis testing and the construction of confidence intervals using numbers announced by the drug makers Pfizer and Moderna publicized by the media. Instead of a two-sample test or more complicated statistical models, we use the elementary one-proportion *z*-test to analyze the data. The method is designed to be accessible for students who have only taken a one-semester elementary statistics course. We will justify the use of a *z*-distribution as an approximation for the confidence interval of the efficacy rate. Bayes's rule will be applied to relate the probability of being in the vaccine group among the volunteers who were infected by COVID-19 to the more consequential probability of being infected by COVID-19 given that the person is vaccinated.

Keywords

COVID-19, hypothesis testing, Bayes's rule, quantitative reasoning

Cover Page Footnote

Frank Wang is Professor of Mathematics at LaGuardia Community College of the City University of New York. His research interests include general relativity and dynamic systems.

Introduction

On November 21, 2020, the *New York Times* featured the article "What Does It Mean if 2 Companies Report 95% Efficacy Rates?" Carl Zimmer wrote, "In the case of Pfizer, for example, the company recruited 43,661 volunteers and waited for 170 people to come down with symptoms of COVID-19 and then get a positive test. Out of these 170, 162 had received a placebo shot, and just eight had received the real vaccine" (2020).

On November 30, 2020, the drug maker Moderna issued a press release containing the following statement:

> Today's primary analysis was based on 196 cases, of which 185 cases of COVID-19 were observed in the placebo group versus 11 cases observed in the mRNA-1273 group, resulting in a point estimate of vaccine efficacy of 94.1%. . . . The Phase 3 COVE trial is a randomized, 1:1 placebo-controlled study testing mRNA-1273 at the 100 μg dose level in 30,000 participants in the U.S., ages 18 and older (Moderna 2020).

The COVID-19 pandemic has produced a deluge of news coverage of quantitative concepts (Ancker 2020). Madison et al. (2012) have written a textbook to illustrate the value of using media articles for quantitative reasoning. COVID vaccine efficacy is arguably the most important statistical news in students' lifetimes, and information given in the above quotes offers a uniquely powerful teaching opportunity to demonstrate how numbers are used in shaping public policy. The objective of this paper is to elucidate the meaning of these numbers in the context of hypothesis testing and confidence intervals for the efficacy rate. Without the confidence interval, a crucial piece of information is missing.

Because the vaccine development typically involves an experimental group and a placebo-controlled group, a two-sample hypothesis testing is a suitable statistical procedure to analyze data. Such a topic is usually presented in a second statistics course. For example, at the author's institution, an urban public university, MAT 120 Elementary Statistics I covers up to one-sample z- and t-tests, and two-sample methods are taught in MAT 121 Elementary Statistics II. Currently, there are 39 sections of MAT 120, but only 1 section of MAT 121, and our situation is quite common, meaning that most students do not have the mathematical tools to analyze news stories concerning vaccine trials. The impetus behind this note is to provide a path forward for educators of first-semester statistics through reframing the data in the news as a one-sample hypothesis test.

In the next section we introduce the principle of hypothesis testing by comparing vaccine trials with coin flipping to make the concept comprehensible for a lay person. Then we use both the p-value and confidence-interval approaches to test hypotheses for one parameter presented in standard textbooks. The following section is concerned with the efficacy rate, its dependence on the number of sick people who are vaccinated, and its linear approximation. The rest of the sections

are mathematical justification of our approximation to allow practitioners to verify our methodology. Further technical details, including a simulation for the confidence interval, are presented in the Appendix.

Principle of Hypothesis Testing

To transform the two-sample problem into a single-sample problem appropriate for many introductory courses, assume that half of the 43,661 volunteers in the Pfizer trial mentioned in the *New York Times* article received the real vaccine, and the other half received a placebo shot. (The Moderna press release explicitly states a 1:1 ratio in their trial.) Out of the 170 people who later became COVID-positive, if the vaccine is ineffective, which is always the null hypothesis in a medical experiment (Ancker and Begg 2017), we expect $0.5 \times 170 = 85$ people to be in the placebo group and 85 people in the vaccine group. But common sense informs us that the actual outcome may not be 85 exactly. It is analogous to tossing a fair coin 170 times and counting the number of heads. One should not be surprised to see 86 or 83 heads.

Figure 1 shows the theoretical probability of getting x heads after tossing a fair coin 170 times, and it is the same probability of observing x volunteers in the placebo group among the 170 people who were infected by the coronavirus, under the null hypothesis that the vaccine is doing nothing. The figure is produced using the open-source software R, and the command is shown in the figure caption so that educators can change the numbers to produce a new plot for a different situation. We see that the probability attains a maximum when $x = 85$ and decreases as x deviates from 85.

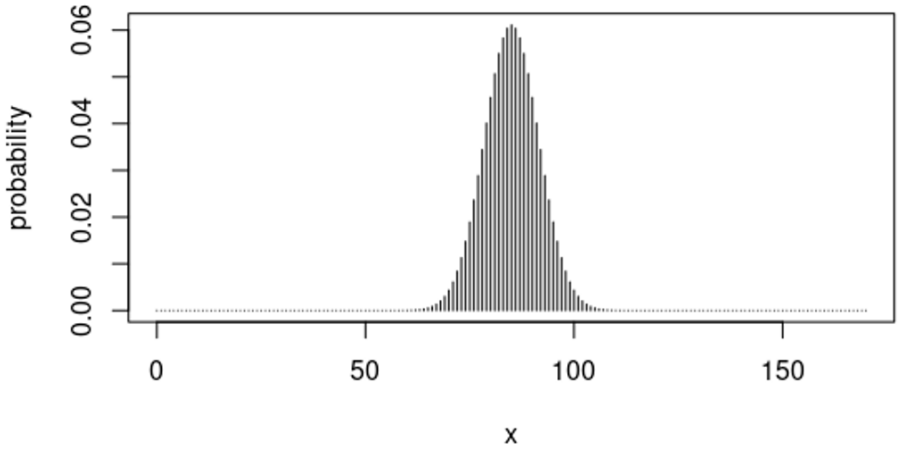

Figure 1. The figure is produced using R with the command plot(dbinom(0:170, 170, 0.5)). For the Moderna trial, we replace 170 with 196.

Let us imagine an outcome that among the 170 people who were COVID positive, 80 were in the vaccine group, and 90 in the placebo group. Although it appears that the infection rate is smaller for the vaccine group, we cannot be sure whether this outcome occurs by chance or is due to the vaccine. Examining Figure 1, there is a noticeable probability for $x = 90$, so it could be by luck that more people in the placebo group and fewer people in the vaccine group got infected. However, Pfizer's actual data show $x = 162$, and from Figure 1 we see that the probability is visually indistinguishable from zero for $x > 105$ or $x < 65$. It will be unreasonable to claim that Pfizer's number, $x = 162$, happens by chance, and this leads scientists to conclude that the vaccine is making a difference. Such a probabilistic consideration is the essence of hypothesis testing.

According to a *New York Times* report (LaFraniere et al. 2020), on November 8, 2020, Pfizer's chief executive, Dr. Albert Bourla, was told that 90 out of 94 people who had gotten sick were in the placebo group. The CEO demanded the scientist to "repeat it, did you say 19 or 90?" It appeared that the CEO was astonished by the unexpected high efficacy. Statisticians use a number called the *p*-value to measure how surprising the data are, given the null hypothesis. We will find the *p*-value shortly.

Students might want to know how the probabilities in Figure 1 are obtained. Returning to the coin-tossing example, if we toss a coin twice, there is one way of getting two tails, two ways of getting one heads and one tails, and one way of getting two heads, so the probability of getting 0, 1, and 2 heads is proportional to 1:2:1. Applying the same consideration, if we toss a coin three times, the probability of getting 0, 1, 2, and 3 heads is proportional to 1:3:3:1. Mathematicians recognized a pattern, and proved that the probability for x heads in n coin tosses is proportional to the nth line of Pascal's triangle. It goes without saying that writing 170 lines of Pascal's triangle will be at least tedious if at all practical, and the need to enumerate efficiently is the driving force for developing mathematical theorems to simplify the calculation for a large number.

While statisticians reject the null hypothesis that Pfizer's result occurs purely by chance, there are still unanswered questions. We need to recognize that no vaccine offers 100% protection. It is possible that a person in the vaccine group gets infected with COVID-19, and it is also possible that a person in the placebo group avoids the virus. We need to quantify such occurrences when developing public policy. The estimation of the vaccine efficacy rate requires further statistical techniques, which will be discussed in the subsequent sections.

Although some students find the topic of hypothesis testing alien, we have shown that the underlying idea is just counting and evaluating plausibility. We remark that psychologists have found evidence that it is easier for the human mind to consider the count instead of percentage (Gigerenzer and Hoffrage 1995;

Cosmides and Tooby 1996). For example, instead of saying 8%, when the same information is presented as 8 out of 100 people, research subjects vastly improved their estimation of probability. The author has replicated some psychological studies with community college students and reached a similar conclusion (Wang 2015). Our demonstration of hypothesis testing using the count instead of percentage is guided by psychological principles.

One-Proportion *z*-Test

Textbooks usually present two ways of answering the question about the null hypothesis: confidence intervals and hypothesis tests. They look at the same problem from two perspectives. The confidence interval is data-centric, and a hypothesis test is model-centric (De Veaux et al. 2020). We use both approaches to analyze the Pfizer data.

We have seen that the probability of getting x heads after tossing a fair coin n times is proportional to the nth line of Pascal's triangle. More formally, the probability follows a binomial distribution, and the probability of observing x heads is written as

$$P(x) = C_x^n \, p^x \, (1-p)^{n-x}$$

where $p = 1/2$ for a fair coin but can be other values between 0 and 1, and

$$C_x^n = \frac{n!}{x!\,(n-x)!}$$

is the binomial coefficient. The algorithm of Pascal's triangle reflects the recurrence relation $C_x^n = C_x^{n-1} + C_{x-1}^{n-1}$.

Abraham de Moivre (1667–1754), a mathematician and consultant to gamblers, noted that the shape of the binomial distribution approaches a very smooth curve if n is large. Students might have already noted this in Figure 1. The bell-shaped curve is the probability density function of a normal distribution. For a normal distribution with mean 0 and standard deviation 1, it is called the z-distribution. While many students are intimidated by the formulas, we emphasize that the z-distribution is rooted in elementary counting. If students understand the principle, they can use computer programs such as R to implement calculations.

Instead of counting the number of heads, in statistics literature the proportion is used. (The change from counts to proportions is perhaps one reason that students find statistics hard, based on the psychological studies mentioned earlier.) Let the true probability of getting heads when flipping a coin be p, the sampling distribution

for the proportion of heads after n tosses is a normal distribution with mean p and standard deviation

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

We point out that the inverse square-root dependence of the standard deviation on n is the most important result in statistics. Howard Wainer's 2007 essay "The Most Dangerous Equation" refers to this equation's critical importance, and he cites past havoc because of ignorance of how sample size affects statistical variation. This equation explains the need for drug makers to wait for a certain number of volunteers to be COVID-positive before their result becomes statistically significant. Rebecca Robbins (2020) reported in the *New York Times* that the surge in coronavirus cases worldwide is actually helping researchers measure more quickly how well their vaccines protect against COVID-19. Specifically,

> In late-stage vaccine trials, the faster that participants get sick, the faster that drug developers gain enough data to know whether their vaccines are effective. . . . The trial ends after a certain number of cases—around 150 to 170—have accrued. That number is chosen to make sure the results have sufficient statistical power to tell how well the vaccine works.

For a hypothesized ineffective vaccine, the proportion of volunteers who received a placebo shot among sick people is $p = 1/2$, based on the assumption that the number of people in the vaccine and placebo groups is the same, just like flipping a fair coin. In the Pfizer trial, $n = 170$ and $\sigma_{\hat{p}} = \sqrt{0.5\ (0.5)/170} = 0.038$. We expect the distribution of sample proportions to be the bell-shaped curve in Figure 2.

The observed sample proportion in the Pfizer trial is

$$\hat{p} = \frac{162}{170} = 0.953.$$

It is apparent that the observed sample proportion is very far from the null hypothesis $p = 0.5$. The test statistic measures the deviation in terms of standard deviation,

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p\ (1-p)}{n}}} = \frac{0.953 - 0.5}{0.038} = 11.81.$$

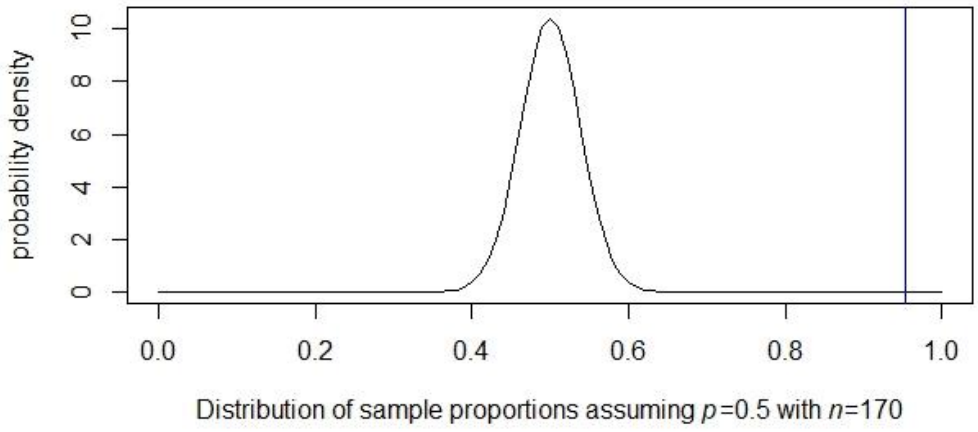

Figure 2. We use a normal distribution to approximate the binomial distribution in Figure 1, and the horizontal axis here is the proportion $x/n = x/170$. The vertical line indicates the observed sample proportion measured in the Pfizer vaccine trial.

The probability for an observation to have $z \geq 11.81$, or the area under the bell-shaped curve to the right of the vertical line, is 1.73×10^{-32} (using R's pnorm(-11.81)). This is the *p*-value, and in this case it means the probability of observing $\hat{p} \geq 0.953$, if the null hypothesis is true. Because the *p*-value is essentially zero, we reject the null hypothesis that the vaccine is ineffective. Comparing Figures 1 and 2, it should be clear that this formal approach also reaches the same conclusion. The main difference is that we use a normal distribution as an approximation of probability, and we use the proportion x/n as the variable instead of the count x.

There have been prevalent misconceptions about statistical significance, and all students should be encouraged to read the statement on *p*-values by the American Statistical Association (Wasserstein and Lazar 2016). Principle 3 says "scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold."

We introduce the confidence-interval approach to go beyond the *p*-value. To estimate the confidence interval for the proportion of people in the placebo group among the 170 COVID-positive cases, we use

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

where $z_{\alpha/2}$ is the critical value. For the most commonly used 95% confidence level, $z_{\alpha/2} = 1.96$. One should pay attention to the use of the estimated sample proportion $\hat{p}$ for the standard error, instead of the hypothesized population parameter p. The numbers provided by Pfizer give

$$\left[0.953 \pm 1.96\sqrt{\frac{0.953\ (1-0.953)}{170}}\right] = [0.921,0.985].$$

Because $p = 0.5$ is outside the 95% confidence interval, we reject the null hypothesis, the same conclusion as before. In canned language, students are typically instructed by textbooks to say, "I am 95% confident that the true proportion of people who did not get the vaccine, among people who are COVID-positive, is between 92.1% and 98.5%." While the statement captures some aspects of the idea, we remind the readers that the interpretation of confidence intervals is subtle. Further insight on this important topic is provided by Gelman and Nolan (2011).

In the next section we will show that Pfizer's confidence interval, 92.1% to 98.5%, serves as an approximation for the vaccine efficacy rate. In the Moderna trial, the efficacy rate is approximated by 185/196 = 94.4% based on the company's press release, and the 95% confidence interval is estimated to be

$$\left[0.944 \pm 1.96\sqrt{\frac{0.944\ (1-0.944)}{196}}\right] = [0.926, 0.962].$$

The confidence interval provides much more information than a single number, and the above calculations serve as an example that we can extract such crucial information from news articles if we scrutinize the press releases.

Vaccine Efficacy Rate and Its Linear Approximation

While news reports of "statistical significance" can be discussed without the addition of new terminology, the concept of vaccine "efficacy" lies outside traditional introductions to statistics. But with a little additional work, this too is fertile ground for instruction. The definition of efficacy can be found in Zimmer's *New York Times* article:

> From these numbers, Pfizer's researchers calculated the fraction of volunteers in each group who got sick. Both fractions were small, but the fraction of unvaccinated volunteers who got sick was much bigger than the fraction of vaccinated ones. The scientists then

> determined the relative difference between those two fractions. Scientists express that difference with a value they call efficacy. If there's no difference between the vaccine and placebo groups, the efficacy is zero. If none of the sick people had been vaccinated, the efficacy is 100 percent (2020).

Let the number of volunteers 43,661 be N, and let the number of COVID-19 cases in the control group among the 170 sick people be x. Pfizer reported $x = 162$, but we keep x as a variable. The sample proportion for people who got sick in the vaccine group is

$$\hat{p}_v = \frac{170 - x}{N/2},$$

and that in the control group is

$$\hat{p}_c = \frac{x}{N/2}.$$

The efficacy rate is

$$e = -\frac{\hat{p}_v - \hat{p}_c}{\hat{p}_c} = 1 - \frac{170 - x}{x}.$$

We plot the efficacy rate as a function of x in Figure 3. As stated in the above quote, if there is no difference between the vaccine and placebo group, or $x = 170/2 = 85$, the efficacy is zero. If none of the sick people had been vaccinated, or $x = 170$, the efficacy is 100 percent. For Pfizer's trial, $x = 162$, and $e = 1 - 8/162 = 0.951$, which is the basis of the company's claim of 95% efficacy rate. For Moderna, the efficacy rate is $1 - 11/185 = 94.1\%$ from the numbers given in the press release.

To help the general public understand the meaning of Pfizer's 95% efficacy rate, let us imagine two otherwise identical communities, but one is vaccinated and the other is not. For every 100 people who become COVID-positive in the unvaccinated community, there will only be 5 infections in the vaccinated community. Consider an inferior vaccine, say with an efficacy rate of 75%, the number of cases in a vaccinated community will be 25, compared with 100 in an unvaccinated community. Because any medical system has a limited capacity, the efficacy rate clearly plays a crucial role in public health planning.

Although it is a calculus topic, it is not difficult to understand that one can approximate a curve by a straight line within a small interval (see the dashed line in Figure 3). If x is near 170, we can put 170 in the denominator of e and obtain a linear approximation:

$$e \approx 1 - \frac{170 - x}{170} = \frac{x}{170}.$$

This linear relation between e and x is the reason that we could use the z-distribution to construct the confidence interval for the efficacy rate (when it is greater than 90%) in the preceding section.

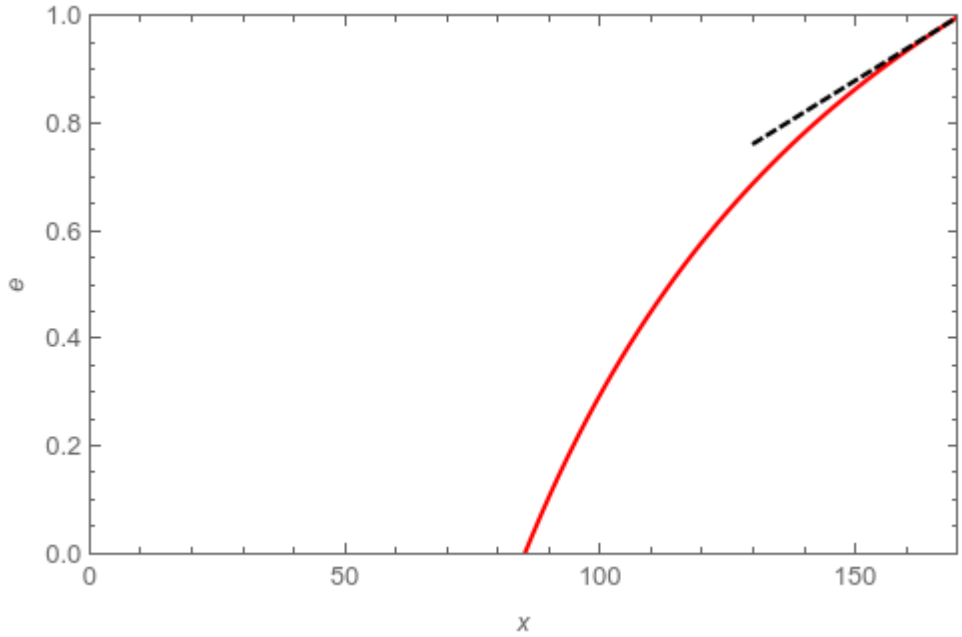

Figure 3. The curve is the efficacy rate as a function of number of x volunteers in the control group who got infected among the 170 people who got infected. The dashed line is a linear approximation of the function near $x = 170$.

Dr. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, said before trial results were published that he would be satisfied with a 75% effective vaccine (Grady 2020). At such a rate, we can solve the equation

$$1 - \frac{170 - x}{x} = 0.75$$

to find $x = 136$. The Food and Drug Administration even said it would consider granting emergency approval for vaccines that showed just 50 percent efficacy (Zimmer 2020), and in this case $x \approx 113$. (Instructors can ask students to find x for other efficacy rates.) As seen in Figure 3, the discrepancy between e and its linear approximation is noticeable when x is less than about 150. In the Appendix we will show how to obtain the confidence interval for any x value using a simulation method.

Two-Proportion *z*-Test

As mentioned in the Introduction, it is natural to perform a hypothesis test for two-sample proportions for a medical experiment. For courses that teach a two-sample test, the vaccine trials offer relevant applications. Referring to the notation in the preceding section, we want to determine whether the infection rates for the placebo group and the vaccine group are different, in other words, whether $p_c - p_v$ differs from zero. The test statistic is

$$z = \frac{\hat{p}_c - \hat{p}_v}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{\frac{N}{2}} + \frac{1}{\frac{N}{2}}\right)}},$$

where $\hat{p} = 170/N$ in Pfizer's trial. Simplifying it, we have

$$z = \frac{2x - 170}{\sqrt{170}}\,\frac{\sqrt{N}}{\sqrt{N-170}} \approx \frac{2x-170}{\sqrt{170}} \quad \text{for } N \gg 170.$$

The approximation is based on the fact that $\lim_{N\to\infty} \frac{\sqrt{N}}{\sqrt{N-170}} = 1$. In Pfizer's trial, N = 43,661, and the factor $\frac{\sqrt{43661}}{\sqrt{43661-170}} = 1.002$ plays a minor role in calculating z. As N tends to infinity, it gets canceled out. As stated previously, the sample size is the most important consideration in inferential statistics. We see that the number of volunteers N = 43,661 is less important, but the number of sick people (170) plays the role of the sample size because of z's inverse square-root dependence on it. With x = 162, $z = \frac{2\times162-170}{\sqrt{170}}\,1.02 = 11.83$ for the two-sample test, not too different from the one-sample z-test, 11.81. If we drop the factor of 1.02, the test statistics for the one-sample and two-sample tests are exactly the same. We will use R to double check these numbers in the Appendix.

Bayes's Rule

Finally, publicly reported data from vaccine trials can reinforce lessons from Bayes's Rule. The ratio $x/170$ we have been using estimates the probability that a person is in the placebo group, given that the person is COVID-positive. In conditional probability notation, we write it as P(placebo|COVID+). Similarly, $(170 - x)/170$ is the probability that a person is in the vaccine group, given that the

person is COVID-positive denoted by *P*(vaccine|COVID+). However, the more relevant conditional probability is the inverse: *P*(COVID+|vaccine), the probability of getting infected by coronavirus, given that the person is vaccinated. The confusion between conditional probability and its inverse is part of Daniel Kahneman's Nobel Prize-winning research summarized in his popular book *Thinking, Fast and Slow* (Kahneman 2011). Lewis (2021) discusses probabilities in the context of the pandemic in this *Journal*.

As seen in previous sections, we have been using numbers from newspapers, so what we had was *P*(vaccine|COVID+). To obtain the inverse probability, we need Bayes's rule. The simplest form of Bayes's rule is in odds form, posterior odds = prior odds × likelihood ratio (Kahneman 2011; Gelman et al. 2013). Textbooks usually present Bayes's rule in probabilistic form, and in the present case

$$P(\text{vaccine}|\text{COVID}+) = \frac{P(\text{COVID}+|\text{vaccine})\,P(\text{vaccine})}{P(\text{COVID}+)}$$

and

$$P(\text{placebo}|\text{COVID}+) = \frac{P(\text{COVID}+|\text{placebo})\,P(\text{placebo})}{P(\text{COVID}+)}.$$

The odds form is the ratio of these two probabilities,

$$\frac{P(\text{vaccine}|\text{COVID}+)}{P(\text{placebo}|\text{COVID}+)} = \frac{P(\text{vaccine})}{P(\text{placebo})} \times \frac{P(\text{COVID}+|\text{vaccine})}{P(\text{COVID}+|\text{placebo})}.$$

The description "posterior odds = prior odds × likelihood ratio" mentioned earlier is more explicit now.

Kahneman and other psychologists found that people tend to neglect the prior odds, and such an oversight leads them to confuse between conditional probability and its inverse. However, for a 1:1 placebo-controlled study,

$$P(\text{vaccine}) = P(\text{placebo})$$

and the prior odds can be neglected. We obtain

$$\frac{P(\text{vaccine}|\text{COVID}+)}{P(\text{placebo}|\text{COVID}+)} = \frac{P(\text{COVID}+|\text{vaccine})}{P(\text{COVID}+|\text{placebo})}.$$

Our inference of the probability of getting coronavirus among people who receive the vaccine, based on the inverse conditional probability from the newspaper articles, is justified.

Conclusion

COVID-19 presents a major challenge to citizenship, and it is important that people understand the meaning of numbers so that they can use that knowledge to guide their behavior (Best 2020). We have demonstrated that much information can be obtained from numbers related to vaccine trials reported in the media with great effect in courses that teach introductory statistics. With the fortuitous high vaccine efficacy rate, we have simplified the hypothesis testing so that students with basic statistical training can understand the results and undertake the analysis by themselves. In an era of misinformation and disinformation, numeracy skills empower citizens and help them make better personal and public decisions.

Acknowledgments

I thank Prabha Betne, Kristen Day, Andrew Gelman, Anthony Giangrasso, Esther I. Wilder, Yun Ye, and Shenglan Yuan for their comments on my note. I thank the editors Nathan D. Grawe, H. L. Vacher, and anonymous reviewers, for their suggestions for improving my manuscript. This work is partly funded by the National Science Foundation, awards #1121844 and #1644975. The author has no connection, financial or otherwise, with any pharmaceutical companies.

References

Ancker, J. S. 2020. “The COVID-19 Pandemic and the Power of Numbers.” *Numeracy* 13(2): Article 2. https://doi.org/10.5038/1936-4660.13.2.1358

Ancker, J. S., and M. D. Begg. 2017. “Using Visual Analogies to Teach Introductory Statistical Concepts.” *Numeracy* 10(2): Article 7. https://doi.org/10.5038/1936-4660.10.2.7

Best, J. 2020. “COVID-19 and Numeracy: How about Them Numbers?” *Numeracy* 13(2): Article 4. https://doi.org/10.5038/1936-4660.13.2.1361

Cosmides, L., and J. Tooby. 1996. “Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty.” *Cognition* 58: 1–73. https://doi.org/10.1016/0010-0277(95)00664-8

De Veaux, R. D., P. F. Velleman, and D. E. Bock. 2020. *Stats: Data & Models*, 5th edition. Hoboken, NJ: Pearson.

Gelman, A., D. Nolan. 2011. *Teaching Statistics: A Bag of Tricks*. New York: Oxford University Press. https://doi.org/10.1201/b16018

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*, 3rd edition. Boca Raton, FL: CRC Press.

Gigerenzer, G., and U. Hoffrage. 1995. "How to Improve Bayesian Reasoning without Instruction: Frequency Formats." *Psychological Review* 102(4): 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Grady, D. 2020. "Early Data Show Moderna's Coronavirus Vaccine is 94.5% Effective." *New York Times*, November 16.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

LaFraniere, S., K. Thomas, N. Weiland, D. Gelles, S. G. Stolberg, and D. Grady. 2020. "Politics, Science and the Remarkable Race for a Viable Vaccine." *New York Times*, November 22.

Lewis, M. A. 2021. "COVID-19 and Quantitative Literacy: Focusing on Probability." *Numeracy* 14(1): Article 5. https://doi.org/10.5038/1936-4660.14.1.1362

Madison, B. L., S. Boersma, C. L. Diefenderfer, and S. W. Dingman. 2021. *Case Studies for Quantitative Reasoning: A Casebook of Media Articles*. 3rd edition. Boston: Pearson.

Moderna. 2020. "Moderna Announces Primary Efficacy Analysis in Phase 3 COVE Study for Its COVID-19 Vaccine Candidate and Filing Today with U.S. FDA for Emergency Use Authorization." https://investors.modernatx.com/news-releases/news-release-details/moderna-announces-primary-efficacy-analysis-phase-3-cove-study 17

Robbins, R. 2020. "One Benefit of Surge in Cases: A Faster Evaluation of Vaccines." *New York Times*, November 19.

Wainer, H. 2007. "The Most Dangerous Equation." *American Scientist* May–June. https://doi.org/10.1511/2007.65.1026

Wasserstein, R. L., and N. A. Lazar. 2016. "The ASA Statement on *P* values: Context, Process, and Purpose." *The American Statistician* 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wang, F. 2015. "Numeracy Infusion Course for Higher Education (NICHE), 2: Development of Students' Bayesian Reasoning Skill." *Numeracy* 8(2): Article 7. https://doi.org/10.5038/1936-4660.8.2.7

Zimmer, C. 2020. "What Does It Mean if 2 Companies Report 95% Efficacy Rates?" *New York Times*, November 21.

Appendix

Most practitioners use software to perform statistical analysis. Below is the R command and output for a two-sample proportion test based on Carl Zimmer's article (2020).

```
prop.test(c(162, 8), c(43661/2, 43661/2))
```

```
        2-sample test for equality of proportions

data:  c(162, 8) out of c(43661/2, 43661/2)
X-squared = 138.24, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.005842099 0.008266602
sample estimates:
      prop 1       prop 2
0.0074208103 0.0003664598
```

To conduct a one-sample proportion test, we have the following. The continuity correction is turned off so that we can compare χ^2 with z^2.

```
prop.test(162, 170, correct = FALSE)
```

```
        1-sample proportions test without continuity correction

data:  162 out of 170, null probability 0.5
X-squared = 139.51, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.9099002 0.9759645
sample estimates:
        p
0.9529412
```

The χ^2 values for the two-sample and one-sample tests are 138.24 and 139.51, respectively, and they are not too different from each other. The χ^2 value for the one-sample test should be the square of the test statistic, and indeed the test statistic in the main text is the same as the software's output: $11.81^2 = 139.51$. We confirm that our one-sample approximation is satisfactory.

Without relying on the linear approximation, we generate 1,000,000 random numbers from a binomial distribution with size 170 and probabilities 8/170 and 162/170, and calculate 1,000,000 efficacy rates; see the R output shown Figure A.1. The 95% confidence interval for the Pfizer trial based on our simulation is [0.914, 0.981]. The *z*-distribution confidence interval [0.921, 0.985] in the main text is up to the mark with the simulation. The discrepancy can be explained by the use of the

normal distribution to approximate the binomial distribution, and the linear approximation of the efficacy rate discussed in the section on efficacy. We conclude that for a high efficacy rate (> 90%), the method of using the one-proportion *z*-test gives a reasonable estimate. If the efficacy is less than 90%, we need to use other methods such as the simulation introduced here to obtain the confidence interval.

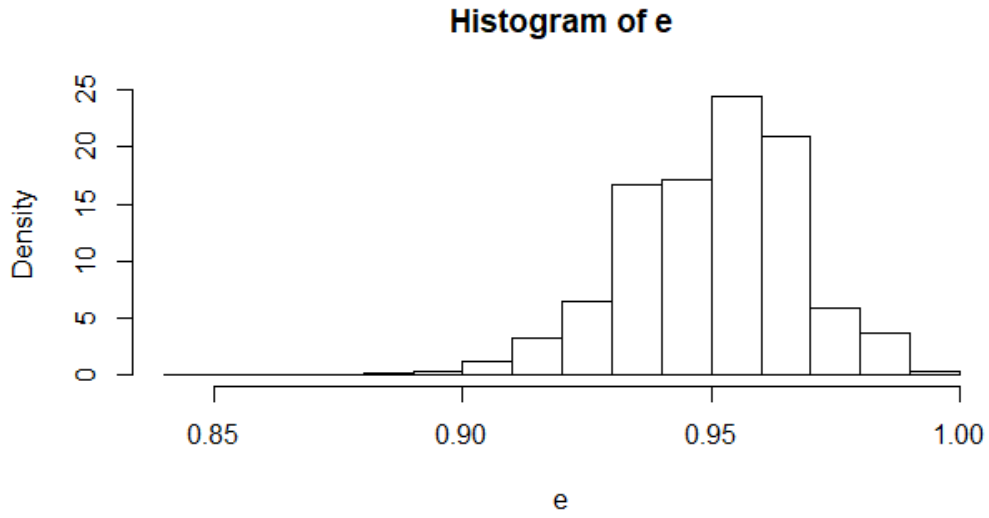

Figure A.1. The histogram of simulated 1,000,000 vaccine efficacy rates produced by R.

```
set.seed(2)
y <- rbinom(1e6, 170, 8/170)
x <- rbinom(1e6, 170, 162/170)
e <- 1 - y/x
hist(e, probability = TRUE)
quantile(e, c(.025, .975))
     2.5%     97.5%
0.9141104 0.9814815
```