Joint Evaluation of Benefit and Risk in Ophthalmic Studies

Mourad Atlas[1], Rajesh Nair[1], Chul Ahn[1]
[1]Division of Biostatistics, Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993

Abstract

For ophthalmic devices intended to provide improvement in uncorrected (without glasses or contact lenses) near visual acuity, uncorrected distance visual acuity may be used to assess risk, since subjects may be giving up distance vision for some gain in near vision. In this situation, one joint analysis of the benefit and risk of the device may be the assessment of how much distance vision the subjects give up (risk) for how much improvement in their near vision (benefit). We propose a method for assessing this benefit and risk that involves calculation of a ratio. A limitation is that ratios are commonly reported in the literature as a point estimates without reporting of the confidence intervals around these estimates. We compare the performance of the Delta, Fieller, and Bootstrap methods of constructing confidence intervals for our proposed ratio. A simulation is carried out to assess the performance of each of the three approaches.

Key Words: Ophthalmic device, benefit and risk assessment, visual acuity, Delta, Fieller, Bootstrap

1. Introduction

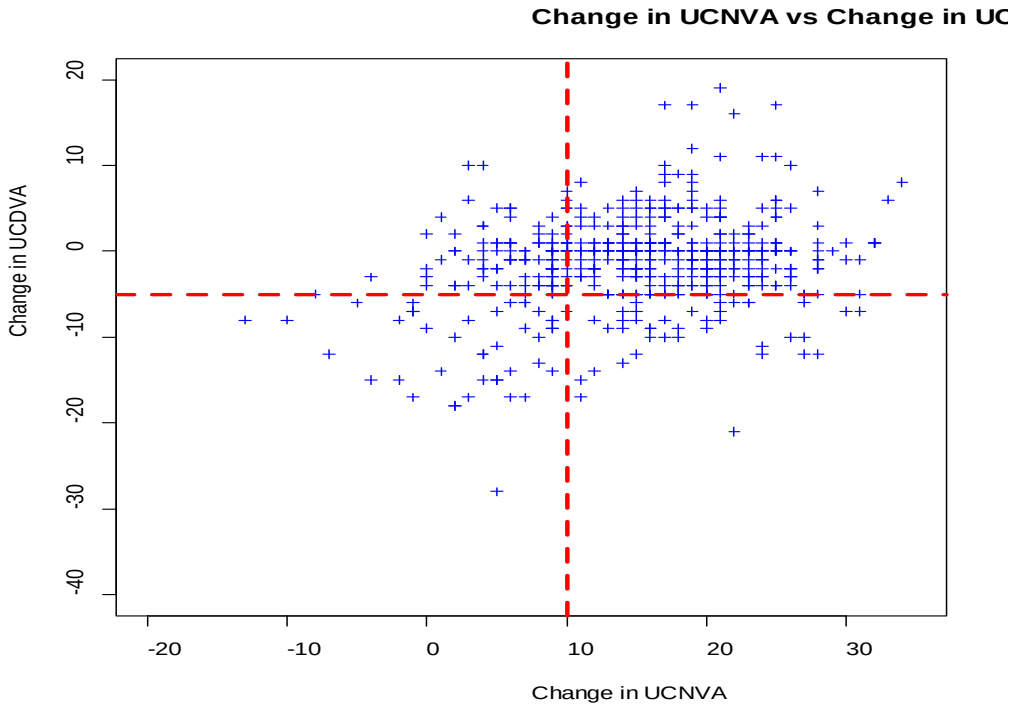
In trials of ophthalmic devices with a main objective of demonstrating improvement in uncorrected (without glasses or contact lenses) near or intermediate visual acuity, uncorrected distance visual acuity (UCDVA) may be used to assess risk, since subjects may be giving up distance vision for some gain in near or intermediate vision. In this case, a joint analysis of the benefit and risk of the device may be the assessment of how much distance vision the subjects give up for how much near or intermediate vision they gain. As a hypothetical example, let's assume that there is a clinical trial investigating a device intended to provide improved near vision in patients with presbyopia, loss of the eye's ability to change focus from distance to near that occurs with the normal aging process. Some of the inclusion criteria may include extremely low refractive error (an error in the focusing of light by the eye) for distance, exceptionally good uncorrected distance visual acuity (20/20 distance vision), and need for reading glasses. In this case, the benefit may be defined as clinically significant improvement of uncorrected near visual acuity (UCNVA). The investigator may want to show the benefit of the device by testing the hypothesis that the percentage of subjects with improvement of UCNVA exceeds a certain target value. Another objective of such a trial may be preservation of UCDVA, since subjects enrolled in the trial have extremely low refractive error and may not require glasses or contact lenses to see well for distance. Therefore, loss of UCDVA may be seen as a risk, defined as a certain degree of loss of letters read on the visual acuity chart. This risk of the device may be assessed by testing the hypothesis that the percentage of subjects with a certain

degree of loss of UCDVA exceeds a certain target value. Alternatively, instead of evaluating the benefit and risk separately, we may consider them jointly.

2. Joint Evaluation of Benefit and Risk

Using our hypothetical example discussed above, we propose a method of evaluating benefit and risk simultaneously. The goal of therapy for the subjects described above may be not to have to wear glasses. Therefore, it is important to know how much UCDVA the subjects give up for how much gain in their UCNVA. To begin, let's look at a hypothetical scatter plot of the within-subject change of UCNVA and UCDVA from baseline to Month 24 in Figure 1 below. Each dot in this figure represents a subject enrolled in the trial. The scales on the X-axis and Y-axis represent the postoperative change in the number of letters of UCNVA and UCDVA from preoperatively, respectively. The vertical reference line is drawn at +10 letters of change in UCNVA. Any dots appearing to the right of this line represent subjects who gained 10 letters or more of UCNVA, which may be considered clinically significant. The horizontal reference line is drawn at a -5-letter change in UCDVA, which may be considered within the measurement error. Any dots appearing below the horizontal reference line represent subjects that lost more than 5 letters of UCDVA from preoperative baseline to 24 months postoperatively.

Figure 1. Change in Uncorrected Near Visual Acuity (UCNVA) vs. Change in Uncorrected Distance Visual Acuity (UCDVA) from Preoperative Baseline at 24 Months Postoperatively (in units of letters on the visual acuity chart)



The first quadrant (Q1), the top right quadrant of Figure 1, is the most favorable benefit/risk quadrant in terms of uncorrected visual acuity outcome, since the subjects in this quadrant have clinically significant gain in UCNVA while not losing more than 5 letters of UCDVA. On the other hand, the third quadrant (Q3), bottom left quadrant, is the least favorable, since subjects in this quadrant experienced a loss of UCDVA of more than 5 letters while not gaining a clinically significant amount of UCNVA.

Let p denote the proportion of subjects in Q1 who have clinically significant gain in UCNVA without losing more than 5 letters of UCDVA. Let q denote the proportion of subjects in Q3 who have no clinically significant gain in UCNVA with loss of UCDVA of more than 5 letters.

We are interested in the ratio of p over q (we let $\theta = p/q$). This ratio of proportions may be a more meaningful description of the effect of the device on the uncorrected vision of the subjects than the difference, even though statistical inference is more complicated for the ratio than the difference of proportions. One of the difficulties in dealing with ratios arises in computing variance estimators.

Despite its popularity as a measure that is used in medical device evaluation, the ratio of benefit and risk is often interpreted and reported primarily as a point estimate without its confidence interval. However, the inferential conclusion one may draw from a sample statistic should reflect uncertainty inherent from the estimation process, and appropriate methods that allow proper interpretation of study findings. One way of proper analysis is to construct confidence interval around the estimate.

3. Calculating the Confidence Interval for the Ratio of Two Proportions

To estimate the confidence interval of the ratio $\theta = p/q$ where p and q are correlated, let $\hat{\theta} = \hat{p}/\hat{q}$ be an estimate of θ. Let's define the variance covariance matrix by

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \quad (1)$$

where v_{11} and v_{22} represent the variance of $\hat{p}$ and $\hat{q}$, respectively.

Using the notation above, we now can describe the three approaches that can be used to construct a confidence interval for the parameter θ, namely the Delta, Fieller, and Bootstrap methods. We selected these three methods, because Delta requires a strong assumption about the distribution of the data, while Bootstrap requires no assumption, and, Fieller requires a weak assumption.

3.1 The Delta Method

The Delta method is a classic technique that is based on a Taylor expansion, yielding an approximate variance for nonlinear function of a random variable. It assumes that the

ratio is normally distributed when the sample size is reasonably large. According to the delta method, the variance estimate of $\hat{\theta}$ is estimated by

$$\hat{\sigma}^2 = \frac{1}{\hat{q}^2}(v_{11} - 2\hat{p}v_{12} + \hat{p}^2 v_{22}) \qquad (2).$$

For a large sample size, we can assume that $\hat{\theta}$ follows a normal distribution. Specifically, we can show that:

$$\hat{\theta} \xrightarrow{d} N(\frac{p}{q}, \frac{p}{nq^3}(p+q)) \text{ as } n \to \infty \qquad (3).$$

3.2 The Fieller Method

Fieller (1954) expressed ratios as linear combination of random variables. He rewrote the ratio $\theta = p/q$ as $p - q\theta = 0$. And, then he considered linear combination, $\hat{p} - \hat{q}\theta = 0$. A familiar statistical theory says that a linear combination of normally distributed random variables is itself normal. Therefore, if $\hat{p}$ and $\hat{q}$ have a bivariate normal distribution with mean vector $(p, q)'$ and variance-covariance matrix as given in equation (2), then

$\hat{p} - \hat{q}\theta \sim N(0, \sigma^2)$, where $\sigma^2 = v_{11} - 2\theta v_{12} + \theta^2 v_{22}$. A $(1-\alpha)\%$ Fieller confidence interval of θ can be then obtained by finding the set of values satisfying the inequality

$$\frac{(\hat{p} + \hat{q}\theta)}{v_{11} + 2\theta v_{12} + \theta^2 v_{22}} \leq t^2_{\alpha/2} \qquad (3)$$

It can be rewritten as $(\hat{p}^2 - t^2 v_{11}) - 2\theta(\hat{p}\hat{q} - t^2 v_{12}) + \theta^2(\hat{q}^2 - t^2 v_{22}) \leq 0 \qquad (4)$

Equation (4) is a quadratic function in the parameter of interest θ, and solving for it leads to the confidence limits.

3.3 The Bootstrap Method

The Bootstrap method has become widely used in statistical inference due to its accessibility. It is shown to be successful in many situations, and is accepted as an alternative to the asymptotic methods. In our case, the sampling distribution of the parameter θ is simulated by sampling from the trinomial distribution and computing $\hat{\theta}$ from the "bootstrapped" sample each time. Below are the steps used for computing bootstrapped confidence limits for θ:

Step 1: Generate n of $u_i's$ from $Uniorm(0,1)$
Step 2: $N_1 = sum(u_i \leq p)$ numbers from $Binomial\ (n, p)$
Step 3: Generate $n - N_1$ number from $Unifrom(0,1)$
Step 4: $N_3 = sum(u_i \leq \frac{q}{1-p})$ numbers from $Binomial\ \left(n - N_1, \frac{q}{1-p}\right)$
Step 5: Calculate and plot the ratio of $\frac{N1}{N3}$

Step 6: Repeat step 1 to 5 B times and compute the 0.025 and 0.975 quantiles in the simulated distribution.

4. Simulation and Results

A simulation study was carried out to investigate the performance of the methods for calculating the confidence limits for the ratio. The coverage probabilities were calculated and histograms were created for 10,000 simulated ratios.The simulation was conducted using the following steps:

1. For each of the following sample sizes (n=50, 100, 300, and 500), 10,000 samples were generated from $(N_1, N_3) \sim Trinomial(n, p = .75, q = .15)$, where N_i represent the number of subject in Q_i $i = 1, 3$.
2. For each sample, the ratio $\frac{N1}{N3}$ and confidence intervals (CIs) by the three methods were calculated.
3. The coverage probability (the number of samples that the "true" value is covered by the CI divided by 10,000) was calculated for each method, and histograms of the coverage probabilities for the CIs were plotted.

Figure 2. Histograms of simulated ratios

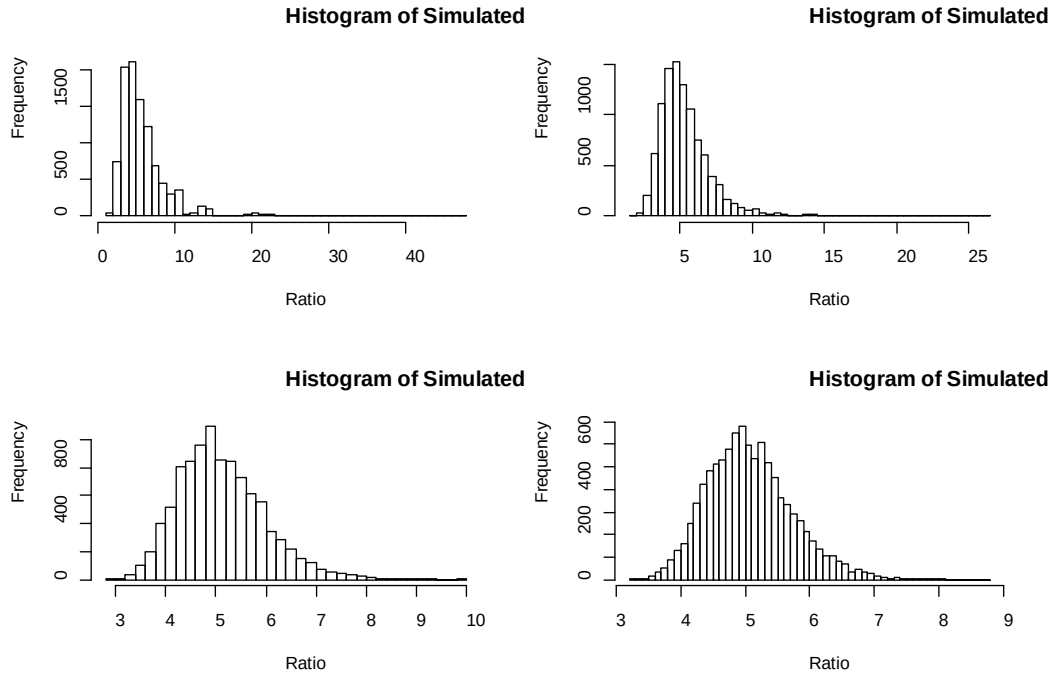


Figure 2 provides the histograms of simulated ratios when the sample size is 50, 100, 300, and 500. We can see that the distribution of ratios is skewed to the right. The skewness is very severe when the sample size is 50, and somewhat severe when it is 100.

When the sample size is 300, the distribution looks very close to the normal even though it is still slightly skewed to the right.

Table 1. Coverage Probabilities

Cov. Prob.	n=50	n=100	n=300	n=500
Delta	92.9%	94.1%	94.8%	94.8%
Fieller	93.5%	93.8%	94.5%	94.7%
Bootstrap	93.9%	94.3%	94.6%	94.6%

Table 2. Confidence Intervals

(LL, UL)	n=50	n=100	n=300	n=500
Delta	(0.99, 9.06)	(2.23, 7.87)	(3.42, 6.64)	(3.77, 6.25)
Fieller	(2.61, 15.73)	(3.09, 10.43)	(3.75, 7.26)	(3.98, 6.58)
Bootstrap	(2.58, 14.33)	(3.09, 9.75)	(3.74, 7.14)	(3.97, 6.52)

Table 1 and 2 represent the simulation results. As expected, the coverage probabilities of the CIs as calculated by the Delta method are smaller than those as calculated by the Fieller and Bootstrap methods when the sample size is 50. This is because the distribution of ratios are skewed to the right when n=50. However, the width of confidence intervals with Delta method is smaller than those from Fieller and Bootstrap when n=50 and n=100. When n=300 or n=500, it appears that the performances of the three methods are similar in both coverage probability and confidence interval width.

5. Conclusions

The joint analysis of benefit and risk as related to uncorrected visual acuity may be useful in some ophthalmic device trials. Based on our simulations, the Delta, Fieller, and Bootstrap methods of calculating the CIs around the ratio of the proportions of subjects with favorable UCVA benefit/risk outcomes to those with unfavorable benefit/risk outcomes, perform well when the sample size is reasonably large (i.e., >300). However, when the sample size is small (i.e., ~50), the CIs calculated bv the Fieller and Bootstrap methods have better coverage probabilities, but their widths are larger than those calculated by the Delta method. One explanation is that Delta method assumes that the ratio is normally distributed, and thus symmetric around its mean. Fieller doesn't assume that the ratio is normally distributed, but it requires that proportions are jointly normally distributed. The confidence intervals from the Fieller method will not be symmetric around the mean in general. Bootstrap doesn't require any assumptions, and seems to perform well even when the sample size is small.

References

Oehlert GW: A note on the delta method. American Statistician 1992, 46:27-29.
Efron B. (1979). Bootstrap methods: another look at the Jackknife. Annals of Statistics 7: 1–26.
Fieller EC. (1940). The biological standardization of Insulin. J. Royal Statististical Society Suppl, 7:1-64.
Fieller EC. (1954). Some problems in interval estimation. Journal of the Royal Statistical Society, Series B 16: 175–185.
Hinkley DV. (1969). On the ratio of two correlated normal variables. Biometrika 56: 635–639.
Luxburg U. and Franz V. (2009). A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations and Bootstrap, Statistica Sinica 19, 1095-1117.