PharmaSUG 2014 - Paper IB05

Attain 100% Confidence in Your 95% Confidence Interval

Indu Nair, United BioSource Corporation, Blue Bell, Pennsylvania
Binal Patel, United BioSource Corporation, Blue Bell, Pennsylvania

ABSTRACT

A very common mistake in the calculation of a confidence interval occurs when there are no qualifying subjects in a by group for a category being tested. It is very tempting to assume that the confidence interval will be missing when the count is zero, which is incorrect. This tends to get overlooked since the usual methods of calculating confidence intervals, such as using a simple PROC FREQ in SAS®, will not take care of the situation without manipulating the code further. This paper will present the different methods in PROC FREQ that allow you to calculate the confidence intervals and discuss which methods are more appropriate to use. This paper will also explain how to use a formula instead of PROC FREQ to calculate confidence intervals correctly and with confidence.

INTRODUCTION

A confidence interval (CI) is an estimated range of values which is likely to include a population parameter of interest. The estimated range is calculated from a given set of sample data. The confidence level is used to indicate the reliability of an estimate. Common choices for the confidence level are 90%, 95%, and 99%.

In clinical trials, responder analyses offer crucial evidence about the effect of investigational treatments. The values in the parameter are classified as a 'Responder' and 'Non-responder'. The response values are often coded as zeros and ones in the data, where zeros represent 'Non-responder' and ones represent 'Responder'. The proportion of responders is often used to provide evidence of the effectiveness of the investigational treatments, thus is a common parameter used for clinical trials. This point estimate (i.e., the percentage of the responder) and the corresponding confidence interval are often provided to describe an estimated proportion of the responders and its reliability calculated from the current sample in the trial.

Methods of constructing confidence intervals depend on the underlying distribution of the estimate. For the most part, this paper will focus on the confidence interval for the binomial proportion using the exact method, also known as the Clopper-Pearson method (Clopper and Pearson, 1934). In special cases, such as difference in proportions, another method will also be explained (Method 5). The confidence interval for the Clopper-Pearson method can be written as

$$\left\{\theta | P[\mathrm{Bin}(n,\theta) \leq X > \frac{\propto}{2}\right\} \bigcap \left\{\theta | P[\mathrm{Bin}(n,\theta) \geq X > \frac{\propto}{2}\right\}$$

where *X* is the number of successes and the Bin(n,θ) represents a binomial variable with n being the number of trials and θ being the probability of success. The Clopper-Pearson method can also be written using the F distribution as

$$pL = \left(1 + \frac{n - n_1 + 1}{n_1 + F(1 - \frac{\alpha}{2}, 2n_1, 2(n - n_1 + 1))}\right)^{-1} \quad pU = \left(1 + \frac{n - n_1}{(n_1 + 1) + F(\propto/2, 2(n_1 + 1), 2(n - n_1))}\right)^{-1}$$

where n is the number of trials, n_1 is the number of successes, and *F(∝,b,c)* is the ∝th percentile of the *F* distribution with *b* and *c* as the degrees of freedom. The FREQ procedure uses this equation to calculate exact confidence limits.

Table 1 shows an example of two different treatment groups, with the number of subjects in each treatment group noted by N=XX. The number and percentages of responders along with their 95% CI is also shown. This paper will discuss different methods to obtain these 95% CIs, why some methods may not give accurate results, and how different scenarios in the data affect calculations of the CI.

	Treatment Group	
Parameter	**Treatment 1 (N=20)**	**Treatment 2 (N=20)**
Responders,n (%)	0 (0.00)	20 (100.0)
[95% CI]	[0.0 - 16.8]	[83.2 - 100.0]

Table 1. Number of Responders and 95% CI by Treatment Groups

DATA SET DESCRIPTION

For this paper, a data set that contains 40 subjects in total assigned to two different treatment groups (TRTAN) will be used. All subjects in TRTAN=1 will be considered as responders (RESP=1), and all subjects in TRTAN=2 will be considered as non-responders (RESP=0). Output 1 below is produced when PROC FREQ is used on the data set described.

```
                        The FREQ Procedure

                        Table of trtan by resp

              trtan     resp

              Frequency|
              Row Pct  |
              Col Pct  |       0|       1|  Total
              ---------+--------+--------+
                     1 |     20 |      0 |     20
                       | 100.00 |   0.00 |
                       | 100.00 |   0.00 |
              ---------+--------+--------+
                     2 |      0 |     20 |     20
                       |   0.00 | 100.00 |
                       |   0.00 | 100.00 |
              ---------+--------+--------+
              Total          20       20       40
```

Output 1. PROC FREQ Output for Values of Responders (RESP) by Treatment Group (TRTAN)

DERIVING CONFIDENCE INTERVALS

METHOD 1 - PROC FREQ WITHOUT ANY ADJUSTMENTS

In addition to being used for getting counts, PROC FREQ can also compute different statistics, including confidence intervals. In this example, PROC FREQ with the option BINOMIAL will be used to produce the Clopper-Pearson exact binomial confidence limits for each treatment group. By default, the confidence limits derived will be 95% confidence limits (ALPHA=0.05); however, other confidence levels can be specified through the ALPHA option. PROC FREQ also produces the exact confidence limits by default, and you can use the EXACT option to specify it as well.

Apply the code below to obtain the confidence limits.

```
proc freq data = testdata;
         by trtan;
         tables resp / binomial alpha = 0.05 ;
         output out = varci binomial ;
run ;
```

Output 2 displays the statistics that would be produced from the PROC FREQ code above. As shown below, TRTAN=1 has 20 non-responder values. The binomial proportion calculated is for RESP=0 and the confidence limits are shown. Assuming PROC FREQ will give the confidence limits for the responders is a mistake that many people succumb to. The value of interest in this scenario is the percentage of responders (RESP=1).

```
------------------------------------ trtan=1 ---------------------------------
                                The FREQ Procedure
                                             Cumulative    Cumulative
                  resp    Frequency     Percent     Frequency      Percent
                  ---------------------------------------------------------
                     0          20      100.00            20       100.00
                               Binomial Proportion for resp = 0
                               -------------------------------
                               Proportion                1.0000
                               ASE                       0.0000
                               95% Lower Conf Limit      1.0000
                               95% Upper Conf Limit      1.0000

                               Exact Conf Limits
                               95% Lower Conf Limit      0.8316
                               95% Upper Conf Limit      1.0000

                                 Test of H0: Proportion = 0.5

                               ASE under H0              0.1118
                               Z                         4.4721
                               One-sided Pr >  Z         <.0001
                               Two-sided Pr > |Z|        <.0001
```

Output 2. PROC FREQ for Values of Responders (RESP) for Treatment Group 1 (TRTAN)

Output 3 is the PROC FREQ output for treatment group two. This looks very similar to Output 2; however in this treatment group, there are 20 responders instead of 20 non-responders. The binomial proportion is getting calculated for RESP=1 and the confidence limits are shown. Since responders are our parameter of interest, the confidence limits given in the output shown below are for the accurate percentage.

A word of caution is that if this treatment has a mix of responders and non-responders, PROC FREQ, by default, will calculate the confidence limits for the lowest level of responders, RESP=0 and the confidence limits will not be accurate.

```
------------------------------------------ trtan=2 ----------------------------------------
                                The FREQ Procedure
                                             Cumulative    Cumulative
                  resp    Frequency     Percent     Frequency      Percent
                  ---------------------------------------------------------
                     1          20      100.00            20       100.00
                               Binomial Proportion for resp = 1
                               -------------------------------
                               Proportion                1.0000
                               ASE                       0.0000
                               95% Lower Conf Limit      1.0000
                               95% Upper Conf Limit      1.0000

                               Exact Conf Limits
                               95% Lower Conf Limit      0.8316
                               95% Upper Conf Limit      1.0000

                                 Test of H0: Proportion = 0.5

                               ASE under H0              0.1118
                               Z                         4.4721
                               One-sided Pr >  Z         <.0001
                               Two-sided Pr > |Z|        <.0001
```

Output 3. PROC FREQ for Values of Responders (RESP) for Treatment Group 2 (TRTAN)

The above output includes various statistics for the binomial proportion including asymptotic upper and lower confidence limits for binomial proportions, p-values for hypothesis testing of p=0.5, in addition to upper and lower exact confidence limits for binomial proportions.

The exact confidence lower and upper limits that we are interested in are variables XL_BIN (lower limit) and XU_BIN (upper limit) which will output to data set VARCI (as shown in PROC FREQ code earlier). Each of these limits needs to be multiplied by 100 to provide the 95% confidence interval. The resulting final table containing the confidence interval is shown in Table 2.

	Treatment Group	
Parameter	**Treatment 1 (N=20)**	**Treatment 2 (N=20)**
Responders, n	0	20
[95% CI]	[83.2 - 100.0]	[83.2 - 100.0]

Table 2. 95% CI by Treatment Group Using METHOD 1

Looking at the 95% CIs in Table 2, you may not immediately realize that there are any issues. However, if the 95% CIs along with the number and percentage of responders per treatment group are presented as shown in Table 3 below, the confidence interval for treatment group one is clearly incorrect. The number of responders is 0; however, the 95% CI is [83.2-100.0] which does not cover the point estimate of 0. In other words the percentage value should always fall within the confidence interval range.

	Treatment Group	
Parameter	**Treatment 1 (N=20)**	**Treatment 2 (N=20)**
Responders,n (%)	0 (0.00)	20 (100.0)
[95% CI]	[83.2 - 100.0]	[83.2 - 100.0]

Table 3. Number of Responders and 95% CI by Treatment Using METHOD 1

Using this PROC FREQ approach will not give the correct confidence interval when there is a population of only non-responders.

METHOD 2 – PROC FREQ – TAKING THE INVERSE

The statistics obtained from method 1 can be manipulated to produce the correct CI. In order to obtain the estimate for the responders, subtracting the estimate from 1 will provide the correct limit. The following code will be used to get the inverse of both confidence limits.

```
data fix ;
    set varci ;
    ** TAKE THE INVERSE **;
    if _bin_ = 1 then
       do ;
           nxl_bin = 1 - xl_bin ;
           nxu_bin = 1 - xu_bin ;
           CI = "[" || put(nxu_bin*100,5.1) || "-" || put(nxl_bin*101,5.1) || "]" ;
       end ;
    else CI = "[" || put(xl_bin*100,5.1) || "-" || put(xu_bin*101,5.1) || "]";
run;
```

Note that as shown in the above code, only the record corresponding to treatment group one is required to take the inverse of the confidence limit. To obtain the confidence interval, the lower and upper limits are swapped after being subtracted from 1.

Parameter	Treatment Group	
	Treatment 1 (N=20)	Treatment 2 (N=20)
Responders,n (%)	0 (0.00)	20 (100.0)
[95% CI]	[0.0 - 16.8]	[0.0 - 16.8]

Table 4. Number of Responders and 95% CI by Treatment Using METHOD 2

In Table 4, it's shown that treatment group one has the correct values. However, the above method has also changed the values in treatment group two, which is not expected.

Treatment group two inadvertently changes when _BIN_=1 is specified in the code to take the inverse of confidence limit. This is because treatment group two only has responders and no non-responders; therefore the binomial proportion is equal to 1 in the PROC FREQ output.

Although the inverse method seems like a reasonable method, it will not be an accurate method to use in all situations. The above method would have worked well if treatment group two had a mix of responders and non-responders.

Another option which should give the correct confidence limits with the addition of a simple option will be discussed next.

METHOD 3 – PROC FREQ WITH WEIGHT OPTION

The option WEIGHT in PROC FREQ will be used and how this can be manipulated for our calculations is explained below.

For this option, a dummy data set will be generated as shown below with both responder values and zero counts.

```
data dummy ;
    do trtan = 1 to 2 ;
       resp = 1 ; count = 0 ; output ;
       resp = 0 ; count = 0 ; output ;
    end ;
run ;
```

TRTAN	RESP	COUNT
1	1	0
1	0	0
2	1	0
2	0	0

Table 5. Dummy data set

Table 5 shows what the resulting dummy data set will look like. The dummy set will then be merged with the frequency count of responders from the sample data set. Table 6 shows the data set of the frequency count of responder values.

TRTAN	RESP	COUNT
1	0	20
2	1	20

Table 6. Data set with Frequency Counts of Responder Values

After the above process, it is ensured that for each treatment group, there will be two records corresponding to both responder and non-responder counts. The data set will look like Output 4 below.

trtan	resp	count
1	0	20
1	1	0
2	0	0
2	1	20

Output 4. Data set after merging with dummy data set

With this new data set created, using PROC FREQ with the WEIGHT statement and the ZERO option will provide accurate estimates and CIs. The LEVEL=2 option must also be used. This option tells PROC FREQ to calculate the confidence limits for the second level of the responder variable (RESP=1), which is the value for responders. This option is applied when all treatments are set up to have a record for each RESP variable.

```
proc freq data = testdata noprint ;
    by trtan ;
    tables resp / binomial (level=2) alpha = 0.05 ;
    weight count / zero ;
    output out = varci binomial ;
run ;
```

The ZERO option specified above is an option that allows any values of zeros as true counts. This code allows the responder value to be forced into the data when the confidence limits are being calculated, allowing for the correct confidence limits to be produced. Table 7 shows the correct 95% CIs produced using this method.

	Treatment Group	
Parameter	**Treatment 1 (N=20)**	**Treatment 2 (N=20)**
Responders,n (%)	0 (0.00)	20 (100.0)
[95% CI]	[0.0 - 16.8]	[83.2 - 100.0]

Table 7. Number of Responders and 95% CI by Treatment Using METHOD 3

METHOD 4 - FORMULA

Besides using PROC FREQ as a method to calculate confidence intervals as shown in method 2 and 3, a manual equation can be used to calculate the interval as well. As stated earlier, SAS uses the F distribution equation to calculate the exact confidence limits. The confidence limits can be calculated manually by converting the equation into a formula to be used in SAS. However before using this formula, a dummy data set needs to be created to represent all responses in all treatment groups as shown earlier in Table 5.

The dummy data set is merged with the frequency of responses by treatment groups and the total population of the treatment groups. The data set that is produced is shown in Output 5 below. The COUNT represents the frequency of response variable by treatment, the DENOM represents the total population count in each treatment group and the PERCENT is the calculation of COUNT/DENOM.

Once this data set is created, this can be used in a formula based on the equation.

trtan	resp	count	denom	percent
1	0	20	20	100
1	1	0	20	0
2	0	0	20	0
2	1	20	20	100

Output 5. Data set after merging with dummy data set

The equation is turned into a SAS formula that can be applied to the data set for the responder values (RESP = 1). The following SAS code will be used.

```
** LOWER CONFIDENCE BOUND  **;
if percent ne 0 then
     lcb=100*count/(count+(denom-count+1)*finv(.975,2*(denom-count+1),2*count));
else lcb=0 ;

** UPPER CONFIDENCE BOUND  **;
if percent ne 100 then
     ucb=100*(1-(denom-count)/((denom-count)+(denom-(denom-count)+1)*
         finv(.975,2*(denom-(denom-count)+1),2*(denom-count)))) ;
else ucb = 100 ;
```

The results shown in Table 7 below have the same results as in method 3 discussed above.

	Treatment Group	
Parameter	**Treatment 1 (N=20)**	**Treatment 2 (N=20)**
Responders,n (%)	0(0.0)	20(100.0)
[95% CI]	[0.0- 16.8]	[83.2- 100.0]

Table 7. Number of Responders and 95% CI by Treatment Using METHOD 4

METHOD 5 – FORMULA FOR DIFFERENCE IN PROPORTION

Often in clinical trials, a comparison of two treatment groups needs to be conducted. Thus the difference in proportion of the responders is of interest.

A sample data set of 90 subjects with treatment groups one and two, containing 60 and 30 subjects respectively, will be used in this method. There will be 30 responders in treatment group one and 15 responders in treatment group two.

As shown in Table 8, the point estimate of the difference is 0.

	Treatment Group 1 (N=60)	**Treatment Group 2 (N=30)**
Responders, n (%)	30 (50.0)	15 (50.0)

Table 8. Difference in proportion of two treatment groups

Confidence limits for the point estimate of zero can be produced using PROC FREQ with the RISKDIFF option. By default, the asymptotic 95% confidence limits will be the only limits produced for the difference. Shown below is an example of the code with this option and Output 6 shows the results.

```
proc freq data = testdata ;
  tables trtan*resp / riskdiff ;
run ;
```

Statistics for Table of trtan by resp
Column 2 Risk Estimates

	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.5000	0.0645	0.3735	0.6265	0.3681	0.6319
Row 2	0.5000	0.0913	0.3211	0.6789	0.3130	0.6870
Total	0.5000	0.0527	0.3967	0.6033	0.3927	0.6073
Difference	0.0000	0.1118	-0.2191	0.2191		

Difference is (Row 1 - Row 2)

Output 6. PROC FREQ Output for RISKDIFF

The Newcombe method based on the Wilson score is a widely used method to calculate the confidence interval for the difference in proportion (Newcombe, 1998). This method avoids aberrations and has better coverage properties. In SAS version 9.1, there is no readily available option in PROC FREQ to calculate this interval except using the formula approach.

The confidence interval based on the Wilson score for each individual proportion is represented as

$$\left(\hat{p}_i + z_{\alpha/2}^2/2n_i \pm z_{\alpha/2}\sqrt{(\hat{p}_i(1-\hat{p}_i) + z_{\alpha/2}^2/4n_i)/n_i}\right) / 1 + z_{\alpha/2}^2/n_i$$

where i represents one of the proportions. Once the Wilson score confidence limits are obtained for both proportions, it can be used in the following equation to obtain the difference in proportion confidence limits.

$$d_L = (\hat{p}_1 - \hat{p}_2) - \sqrt{(\hat{p}_1 - L_1)^2 + (U_2 - \hat{p}_2)^2}$$

$$d_U = (\hat{p}_1 - \hat{p}_2) + \sqrt{(U_2 - \hat{p}_1)^2 + (\hat{p}_2 - L_2)^2}$$

The U and L represent the upper and lower limit, respectively, of the corresponding proportion. Using this equation, a formula can be written in SAS to calculate the confidence interval of the difference of proportion.

```
upper1 = (p1+(z**2)/(2*n1)+z * sqrt ((p1*(1-p1) + (z**2)/(4*n1))/n1))/(1+(z**2)/n1) ;
lower1 = (p1+(z**2)/(2*n1)-z * sqrt ((p1*(1-p1) + (z**2)/(4*n1))/n1))/(1+(z**2)/n1) ;
upper2 = (p2+(z**2)/(2*n2)+z * sqrt ((p2*(1-p2) + (z**2)/(4*n2))/n2))/(1+(z**2)/n2) ;
lower2 = (p2+(z**2)/(2*n2)-z * sqrt ((p2*(1-p2) + (z**2)/(4*n2))/n2))/(1+(z**2)/n2) ;

du = (p1 - p2) + sqrt ( (upper1 - p1)**2 + (p2 - lower2)**2 ) ;
dl = (p1 - p2) - sqrt ( (p1 - lower1)**2 + (upper2 - p2)**2 ) ;
```

In the SAS code above, p1 and p2 are the proportions of responders in each treatment group, n1 and n2 are the number of subjects in each treatment group, and z is the 100(1 – α/2)th percentile of the standard normal distribution. The 95% confidence interval of the zero difference in the proportion of the two treatment groups is shown in Table 9 below.

	Treatment Group 1 (N=60)	**Treatment Group 2 (N=30)**
Responders, n (%)	30 (50.0)	15 (50.0)
Difference in Proportion (T1 – T2), 95% CI	**0.00 [-.2084, .2084]**	

Table 9. Number of Responders and 95% CI for the Difference in Proportion

In higher versions of SAS (such as SAS 9.2, SAS 9.3), additional options provided for RISKDIFF can be used to produce the Newcombe intervals directly. Following is the code.

```
proc freq data = testdata ;
  tables trtan*resp / riskdiff  (column=2 cl=(newcombe));
run ;
```

COLUMN is used to specify which level of the variable is of interest and the CL option allows you to specify which method to use to generate the 95% confidence limits. Output 7 shows what is produced using the code above, which not only includes the 95% asymptotic confidence limits but the confidence limits based on the Newcombe method as well.

```
                          Column 2 Risk Estimates

                                        (Asymptotic) 95%        (Exact) 95%
                    Risk        ASE     Confidence Limits     Confidence Limits
-------------------------------------------------------------------------------
Row 1           0.5000      0.0645      0.3735      0.6265      0.3681      0.6319
Row 2           0.5000      0.0913      0.3211      0.6789      0.3130      0.6870
Total           0.5000      0.0527      0.3967      0.6033      0.3927      0.6073

Difference      0.0000      0.1118     -0.2191      0.2191

                         Difference is (Row 1 - Row 2)

            Confidence Limits for the Proportion (Risk) Difference
                              Column 2 (resp = 1)
                          Proportion Difference = 0.0000
                    Type              95% Confidence Limits
                    Newcombe Score    -0.2084        0.2084

                              Sample Size = 90
```

Output 7. 95% CI for the Difference in Proportion Using SAS

CONCLUSION

Although PROC FREQ gives the tools to calculate the confidence interval, there are many things to be aware of. The code needs to be accurate and efficient to correctly calculate the confidence interval, even when there are only non responders in a given treatment group. It is also important to know that it is always possible to calculate confidence intervals by using formulas. By applying the methods described above, you can walk away confidently with a confidence interval that works for all data situations.

REFERENCES

Clopper, C.J. and Pearson, E.S. 1934. The Use of Confidence of Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika.* 404-413.

He, Xiaomen; Wu, Shwu-Jen. 2009. Confidence Intervals for the Binomial Proportion with Zero Frequency. *Proceedings for PharmaSUG Conference 2009*. Available at http://www.pharmasug.org/download/papers/SP10.pdf

Motulsky, Harvey. 2010. *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking.* 23-55.New York, New York. Oxford University Press.

Newcombe, Robert G. 1998." Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods." *Statistics in Medicine*. 873-890. New York, New York. John Wiley & Sons, Ltd.

Rice, John A. 2007. *Mathematical Statistics and Data Analysis.* 217-220. Belmont, CA. Brooks/Cole. SAS/STAT® 9.22 User's Guide, The FREQ Procedure.

Shekar, Tulin. 2013. How to Handle Binomial Proportion with Zero Frequency. *Proceedings for the Northeast SAS Users Group Conference 2013.* Available at http://www.nesug.org/Proceedings/nesug13/41_Final_Paper.pdf

ACKNOWLEDGMENTS

We would like to thank Dr. Wenjiong Zhou from United BioSource Corporation for her review and valuable input for this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Indu Nair
Enterprise: United BioSource Corporation
Address: 920 Harvest Drive, Suite 200
City, State ZIP: Blue Bell, Pennsylvania 19422 USA
Work Phone: +1 267-470-1611
E-mail: indu.nair@ubc.com

Name: Binal Patel
Enterprise: United BioSource Corporation
Address: 920 Harvest Drive, Suite 200
City, State ZIP: Blue Bell, Pennsylvania 19422 USA
Work Phone: +1 215 390 2219
E-mail: binal.patel@ubc.com