

Confidence Intervals in Analysis and Reporting of Clinical Trials

Guangbin Peng, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

Regulatory agencies around the world have recommended reporting confidence intervals for treatment differences along with the results of significance tests. SAS provides easy and convenient ways to produce confidence intervals using procedures such as PROC GLM and PROC UNIVARIATE in conjunction with ODS (output delivery system). In this paper, I will discuss the relationship between significance tests and confidence intervals, summarize the types of confidence intervals used in clinical study reports, and provide examples from clinical trials to illustrate the computation of distribution-dependent confidence intervals for the mean treatment difference and distribution-free confidence intervals for the median response within each treatment group using SAS.

INTRODUCTION

Confidence interval estimation and significance testing (hypothesis testing) are the two most commonly used statistical inference methods for clinical trials (Walker, 1997). Because p-values have been widely presented and, in the past, more easily obtained from standard statistical software than have confidence intervals, significance tests have also been more widely accepted by the medical community than have confidence intervals. The extensive use of significance tests with clinical trial data has further increased their popularity and made confidence intervals less popular. However, the overuse and misinterpretation of significance tests have led to advocates

for more frequent use of confidence intervals (Simon, 1993).

There is a close relationship between confidence intervals and significance tests (Hahn and Meeker, 1991); in fact, a confidence interval can often be used to test a hypothesis. If the $100(1-\alpha)\%$ confidence interval for the mean treatment difference in a clinical trial does not contain zero, there is evidence to indicate a treatment difference at the $100\alpha\%$ significance level. This strategy is equivalent to the hypothesis test that rejects the null hypothesis of no mean treatment difference at the level of α .

Compared to p-values, confidence intervals are generally more informative. They provide quantitative bounds that express the uncertainty inherent in estimation, instead of merely an accept or reject statement. The length of a confidence interval depends on the sample size; this influence of sample size is evident from observing the length of the interval, while this is not the case for a significance test. So confidence intervals are usually more meaningful than statistical hypothesis tests alone. Moreover, they are easier to explain to those with no formal training in statistics (Hahn and Meeker, 1991).

Regulatory agencies around the world have emphasized in recent years the importance of reporting confidence intervals in clinical study reports. The ICH Harmonized Tripartite Guideline on statistical principles for clinical trials states, "Estimates of treatment effects should be accompanied by confidence intervals, whenever possible, and the way in which these will be calculated should be identified. . . . it is important to bear in mind the need to provide

statistical estimates of the size of treatment effects together with confidence intervals (in addition to significance tests).”

In this paper, I will discuss the types of confidence intervals and the different methods for constructing them. I will also illustrate the calculation of confidence intervals for clinical trial data by providing sample SAS code and output. Finally, I will discuss the advantages of presenting confidence intervals along with p-values in clinical study reports.

CHARACTERISTICS OF CONFIDENCE INTERVALS

Because there exist a variety of confidence intervals, the analyst must determine which type of interval to use depending on the application (Hahn and Meeker, 1991). Two commonly used types are confidence intervals for population parameters and confidence intervals for distribution percentiles. The most frequently used type of confidence interval attempts to capture the population mean. Sometimes, however, the analyst may construct confidence intervals for the standard deviation or other shape parameters for a distribution to satisfy his needs. In the case that the assumed distribution parameters are not suitable to describe the sampled population, the analyst may focus on one or more percentiles of the sampled distribution and construct confidence intervals for them (for example, for the median or the quartiles). Sometimes one-sided confidence bounds are desired in situations where the major interest is restricted to the lower limit or the upper limit alone.

All statistical intervals have an associated confidence level. The analyst must determine the confidence level

based on what seems to be an acceptable degree of assurance for the specific application (Hahn and Meeker, 1991). For instance, the analyst may construct a 95% confidence interval for the mean. This indicates that the method of construction guarantees that 95% of all such intervals will contain the (true) population mean. (Of course, this also means that 5% of them will not.) One can request a higher level of confidence, which will reduce the chances of obtaining an interval that does not contain the population mean. However, increasing the confidence level results in a wider (that is, less precise) interval for a fixed sample size. On the other hand, when there is a fixed confidence level, the length of intervals becomes shorter as the sample size increases. So, the analyst may choose higher confidence levels with large samples and lower confidence levels with small samples. In some cases, obtaining meaningful confidence intervals becomes impractical because of the small sample size or complexity of analysis.

CONFIDENCE INTERVALS FOR CLINICAL TRIALS

The most frequently used confidence interval for clinical trial data is the 95% confidence interval for the mean treatment difference. The selection of 95% for the confidence level is common across disciplines. The reason for this selection seems quite obvious for clinical trials, especially for confirmatory trials as a result of the study design, and the 95% confidence level provides reasonable assurance along with adequate precision for most trials. The methods for calculating these confidence intervals can be generally put into two categories: distribution-dependent and distribution-free methods.

Construction of distribution-dependent confidence intervals requires one to assume a particular distribution, such as the normal distribution. From experience, the normal assumption appears to be valid for many clinical trial data analyses. However, it may be inappropriate to calculate distribution-dependent confidence intervals when the assumed distribution does not fit the data well. In such cases, distribution-free confidence intervals should be constructed. A distribution-free interval sometimes may not exist, and its length is generally longer than the corresponding distribution-dependent interval for a particular distribution. This is the price that one pays for not making the distribution assumption (Hahn and Meeker, 1991). So, a distribution-dependent confidence interval should be chosen whenever there is solid evidence that the data follows a tractable distribution.

DISTRIBUTION-DEPENDENT CONFIDENCE INTERVAL

If the assumption that the data are normally distributed is valid, one can construct confidence intervals for the mean treatment difference. The general form of a confidence interval for the mean difference between two treatment groups (Group A and Group B) is

$$\bar{Y}_a - \bar{Y}_b \pm t_{1 - \alpha / 2, df} * S_{(\bar{Y}_a - \bar{Y}_b)} \quad (1)$$

where \bar{Y} is the mean or least squares mean, and $S_{(\bar{Y}_a - \bar{Y}_b)}$ is the standard error of the estimate of $(\bar{Y}_a - \bar{Y}_b)$, and $t_{1 - \alpha / 2, df}$ is the 100(1- α /2) percentile from student's t distribution with 'df' degrees of freedom.

In practice, one can calculate a 95% confidence interval for the mean

difference through analysis of variance, although it is critical to obtain the degrees of freedom and estimate the standard error (or mean square error) correctly. Before the release of SAS version 8, the analyst had to extract these values from the SAS procedures and then calculate the confidence intervals using the formula above (or one of its variations) with custom written SAS code. The following code using SAS version 6.09 represents one way to obtain the 95% confidence interval for the mean treatment difference from an ANOVA model.

```

*-----*;
* The following statements get output *;
* datasets containing statistics *;
* needed for calculation of confidence*;
* interval *;
*-----*;
proc glm data=final outstat=glmdt
  noprint;
  class &trt &str &invcd;
  model &dep=&indp;
  lsmeans &trt/pdiff stderr tdiff
    out=lsmeandt;

*-----*;
* The following statements get the *;
* degree of freedom from the model *;
*-----*;
data _null_; set glmdt;
  if _type_='ERROR';
  call symput('df', df);

*-----*;
* The following statements get the *;
* LSMEANS from the model *;
*-----*;
data _null_; set lsmeandt;
  if &trt='&control' then call
    symput('pLSM', lsmean);
  if &trt='&trt1' then call
    symput('tLSM', lsmean);

*-----*;
*The following statements get the 95% *;
*confidence interval *;
*-----*;
data mgmean; set lsmeandt;
  lb=(&tLSM -&pLSM)-tinv(.975,
    &df)*abs(&tLSM -&pLSM)/sqrt(f);
  ub=(&tLSM -&pLSM)+tinv(.975,
    &df)*abs(&tLSM -&pLSM)/sqrt(f);

```

In this example, the estimated standard error was obtained using the relationship between the t-distribution and F-

distribution. Since we have the following two equations:

$$|t| = \left| \frac{\bar{Y}_a - \bar{Y}_b}{S_{(\bar{y}_a - \bar{y}_b)}} \right| \quad (2)$$

$$F = t^2 \quad (3)$$

Then $S_{(\bar{y}_a - \bar{y}_b)}$ can be estimated using the F -value from the model.

With the release of SAS version 8 and ODS, obtaining confidence intervals has become an easier task from the programming point of view. The following code illustrates the way to get the same confidence interval using ODS in SAS version 8.

```

*-----*;
* The following statements use ods to *;
* obtain the 95% confidence interval *;
* for the mean treatment difference *;
*-----*;
ods listing close;
ods output LSMeanDiffCL=lmci;

proc glm data=final;
class &trt &invcd;
model &dep=&indp ;
lsmeans &trt/pdiff cl ;
run;

ods output close;
ods listing ;

*-----*;
* The following statements put the 95% *;
* confidence interval for treatment *;
* difference into macro variables *;
*-----*;
data _null_ ; set lmci ;
call symput ('LCL',put (LowerCL,
5.2)); **** Lower CI ;
data _null_ ; set lmci ;
call symput ('UCL',put (UpperCL,
5.2)); **** Upper CI ;

```

In some cases, the analyst was asked to construct confidence intervals for proportions or percentages. For instance, researchers may want to estimate the difference between treatments in the percentage of patients who respond to the therapy. Rather than providing only the chi-square p -value, presenting the confidence interval for the treatment

difference in the percentage of responders will be more informative. Using the normal approximation to the binomial distribution, the asymptotic confidence interval is as follows:

$$(P_a - P_b) \pm Z_{1-\alpha/2} * S_{(P_a - P_b)} \quad (4)$$

where P_a , P_b are the percentage of responders in each treatment group and $S_{(P_a - P_b)}$ is the standard error of the difference in percentages.

The task will be very simple if one can estimate $S_{(P_a - P_b)}$. In fact, it can be estimated by the quantity,

$$\sqrt{P_a(1 - P_a) / n_a + P_b(1 - P_b) / n_b} .$$

One can extract the necessary components from PROC FREQ to calculate the confidence interval in this case or use the new option 'RISKDIFF' in SAS version 8 to complete the job.

```

*-----*;
* The following statements put the 95% *;
* confidence interval for treatment *;
* difference into macro variables *;
*-----*;
proc freq data=&ind noprint;
table &trt*&var /chisq riskdiff;
output out=pval PCHI rdif2;

data pval; set pval; length categ $55.;
categ="&cat";
pv=P_PCHI;
ll=100*L_RDIF2;
ul=100*U_RDIF2;
keep pv ul ll categ;

```

DISTRIBUTION-FREE CONFIDENCE INTERVAL

Without requiring that the data follow a particular distribution, one can construct distribution-free confidence intervals using order statistics. However, it is generally impossible to obtain an interval with precisely the desired confidence level, which has severely limited their use. In order to obtain an interval with the desired confidence level or with an acceptable length, relatively large samples are often

required (Hahn and Meeker, 1991). After specifying the confidence level, one has to determine the order statistics from the sample and select those order statistics as the endpoints of the interval with at least the specified confidence level.

Generally, the two-sided distribution-free $100(1-\alpha)\%$ confidence interval for Y_p , the 100pth percentile of the sample of size n is $[x_{(l)}, x_{(u)}]$, where $x_{(j)}$ is the j th order statistic. The lower rank and upper rank are integers that are symmetric or nearly symmetric around $i = [np] + 1$, where $[np]$ is the integer part of np (the non-parametric estimate of the 100pth percentile lies between $x_{([np])}$ and $x_{([np]+1)}$). Determination of l and u requires a stepwise process:

1. Let l and u be the two integers closest to $p(n+1)$. (If $p(n+1)$ is itself an integer, then l and u will be the same.)
2. Decrement l and increment u by one until the following constraint is met:

$$Q_b(u-1; n, p) - Q_b(l-1; n, p) \geq 1-\alpha, \quad (5)$$

where Q_b is the cumulative binomial probability, $0 < l < u \leq n$, and $0 < p < 1$. The coverage probability is sometimes less than $1-\alpha$ in the tails of distribution when the sample size is small. To avoid this problem, one can choose asymmetric values of l and u . PROC UNIVARIATE in SAS version 8 provides a convenient way to obtain the distribution-free confidence interval for the quartiles. The following code shows how to get a 95% distribution-free confidence interval for the median.

```
*-----*;  
* The following statements get the 95% *;  
* confidence interval for each *;  
* treatment group based on order *;  
* statistics *;  
*-----*;  
ods listing close;  
ods output quantiles=pctldf ;
```

```
proc univariate data=final loccount  
  modes cibasic(alpha=.05)  
  cipctldf(TYPE=ASYMMETRIC  
  alpha=.05);  
  var P&var ;  
  by &trt ;  
run;  
  
data pct; set pctldf;  
  if quantile='50% Median' ;  
run;  
ods output close;  
ods listing ;  
  
*-----*;  
* The following statements put the 95% *;  
* confidence interval into macro *;  
* variables *;  
*-----*;  
data _null_ ; set pct; if &trt= "&trt1";  
  call symput ('lclddf',  
  put(LCLDistFree, 7.2));  
  
data _null_ ; set pct; if &trt= "&trt1";  
  call symput ('uclddf',  
  put(UCLDistFree, 7.2));
```

In clinical trials, the primary objective usually is to demonstrate a difference between treatments. It may be hard to draw meaningful inference from the significance test alone when using non-parametric or even more complicated approaches. Providing a confidence interval can help to interpret the significance of the finding and quantify the treatment effect to a certain degree. Table 1 shows the results from an actual clinical study. The p-value is obtained using van Elteren's test, a form of Wilcoxon rank sum test with consideration of strata (based on patients' baseline severities). The null hypothesis of this significance test is that the distributions of data for each treatment group are the same. The significant p-value shows indirectly the strong treatment effect. More importantly, the 95% confidence intervals for the median of each treatment group are not overlapped, providing quantitative information about the treatment difference.

The process for constructing confidence intervals should be consistent with that for the associated significance test.

Usually, one designs a trial to detect a treatment difference using a pre-specified statistical test at α level of 5% (two-sided test), so that a 95% confidence interval is the equivalent choice and will provide inference consistent with the significance test. This close relationship is based upon the common statistical model used for the analysis and the confidence interval. If the significance test is based on least squares means under the generalized linear model, the inference from the confidence interval for the difference of LSMEANs will agree with the conclusion drawn from the significance test. However, when the primary analysis takes a non-parametric approach, the analyst has to be cautious in construction of the distribution-free confidence interval. When the treatment effect is strong, there will likely be agreement in inference no matter which method is used. But ambiguity may appear between several approaches when the treatment effect is not clearly identifiable. In table 2, the p-value from van Elteren's test is near the nominal significance level. The overlapping of the confidence intervals for median percent change has provided the additional information about this uncertainty that the significance test alone cannot.

CONCLUSION

The reporting of p-values alone is insufficient and can sometimes be misleading. Confidence intervals are generally more informative than p-values. They provide quantitative bounds that express the uncertainty inherent in estimation, instead of merely

an accept or reject statement. On the other hand, inference from confidence intervals is equivalent to that obtained from the associated (same distributional assumption) hypothesis test. Therefore, reporting of confidence intervals along with significance tests will improve the understanding of the findings from clinical trials. This paper has presented the types of confidence intervals and their properties and demonstrated how to construct both distribution-dependent and distribution-free confidence intervals using SAS. Although these processes were challenging prior to the release of SAS version 8, they are now very straightforward and can be implemented easily with the code presented here.

Table 1 LOCF Analysis Summary
All Randomized Subjects

Treatment Group	Timepoint	Results						95% CI {a}	p-Value {b}
		n	Mean	Median	SD	Min	Max		
Group 1 (N =339)	Baseline	322	18.97	15.00	14.62	0.00	103.00		
	Endpoint		14.75	10.31	14.46	0.00	89.00		
	Change		-4.22	-3.00	13.03	-87.00	36.00		
	PercentChange		-9.17	-27.48	93.34	-100.00	700.00	(-34.69, -18.18)	
Group 2 (N =344)	Baseline	286	18.17	14.00	14.26	0.00	87.00		
	Endpoint		9.77	7.00	9.92	0.00	53.67		
	Change		-8.40	-6.94	11.39	-62.00	35.25		
	PercentChange		-34.31	-50.00	78.28	-100.00	550.00	(-56.04, -43.75)	<.001

{a} 95% confidence interval for the median percent change was obtained using non-parametric methods for order statistics.

{b} P-value is obtained from Van Elteran's test for the analysis of the percent changes

Table 2 LOCF Analysis Summary
All Randomized Subjects

Treatment Group	Timepoint	Results						95% CI {a}	p-Value {b}
		n	Mean	Median	SD	Min	Max		
Group1 (N =231)	Baseline	229	18.33	14.70	15.50	1.00	82.00		
	Endpoint		11.55	7.00	12.43	0.00	66.00		
	Change		-6.79	-5.50	10.71	-72.00	31.70		
	PercentChange		-31.94	-40.00	59.53	-100.00	316.67	(-50.00, -33.33)	
Group2 (N =227)	Baseline	200	18.49	14.00	15.24	1.00	92.00		
	Endpoint		10.00	6.00	11.32	0.00	57.56		
	Change		-8.49	-7.00	12.45	-88.50	26.25		
	PercentChange		-41.27	-53.55	53.11	-100.00	240.00	(-60.00, -42.42)	.050

{a} 95% confidence interval for the median percent change was obtained using non-parametric methods for order statistics.

{b} P-value is obtained from Van Elteren's test for analysis of the percent changes

REFERENCES

Hahn, G. J. and Meeker, W. Q. (1991). Statistical Intervals: A Guide for Practitioners, New York: John Wiley & Sons, Inc.

ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials.

Walker, G. A. (1997). Common Statistical Methods for Clinical Research with SAS Examples, SAS Institute Inc.

Simon, R. (1993). Why Confidence Intervals are Useful Tools in Clinical Therapeutics. Journal of Biopharmaceutical Statistics, 3(2), 243-248.

ACKNOWLEDGEMENT

I owe special thanks to Scott Beattie, who reviewed this paper in depth and provided the great comments. I would also like to thank Yan Zhao for his valuable input.

CONTACT INFORMATION

Guangbin Peng
Statistical Analyst
Primary Care Product Team
Lilly Corporate Center
Indianapolis, Indiana 46285
U.S.A
Phone: 317 433 8445
Fax: 317 276 4789
Email: guangbin_peng@lilly.com