PharmaSUG 2023 - Paper SA-285

Guidelines for the Statistical Analysis in German Dossier Submissions

Meiling Gao, Xi Qian, and Hui Mao, BioPier Inc.

ABSTRACT

German dossier submission is the regulatory submission process required by the German federal institute for drugs and medical devices for the approval of new drugs, generic drugs, and biosimilars in Germany. The submission contains comprehensive information on the quality, efficacy, and safety of the drug, and must follow specific guidelines. Pharmaceutical companies must submit such information to a German dossier before launching their products in Germany. However, the submission process can be complicated and challenging, as it requires a thorough review and evaluation of drugs with complex statistical method. This paper aims to fill the gap by providing guidelines for conducting statistical analysis in German dossier submissions. We will cover the statistical strategies and common pitfalls when preparing the dossier submission. The paper focuses on a comprehensive discussion of the most used statistical methods for analyzing a broad range of data and outcomes, including dichotomous, continuous, time-to-event data, *etc*. For each category of analysis, we will begin with an introduction to the relevant statistical basics, followed by a description of the sample data. We will also provide sample SAS® codes and guidelines for interpreting the results.

INTRODUCTION

The submission of German dossier for a new drug requires a detailed analysis of clinical trial data to demonstrate the drug's efficacy, safety, and quality. This process often involves preparation of hundreds of tables with complex statistical analysis, which must be conducted in a comprehensive and robust manner to ensure that the new drug's safety and effectiveness.

The evaluation of clinical interventions often involves the analysis of dichotomous, continuous, and time-to-event outcomes. In the context of German dossier submissions, the dichotomous outcomes typically include disease response rate and adverse event occurrence, which can be quantified using effect measures such as relative risk, odds ratio, and risk difference. Common continuous outcomes may include quality-of-life scores, and their corresponding effect measures may include mean difference in change from baseline or standardized mean difference such as Hedge's g. Time-to-event outcomes may include overall survival, time to relapse, time to onset of adverse event and other similar metrics. The commonly used outcomes, descriptive statistics, effect measures, and statistical methods are summarized in Table 1. This paper aims to provide an overview of basic statistical concepts and methods for analyzing these three types of outcomes. We will also include mock shells, sample SAS codes, and analysis results to illustrate the practical application of these statistical techniques.

Outcomes	Descriptive statistics	Measure of effect	Statistical tests/models
Dichotomous • Response rate • Adverse event (AE) occurrence	• n/N (%)	• Relative risk (RR) • Odds ratio (OR) • Risk difference (RD)	• Stratified Cochran-Mantel-Haenszel (CMH) Chi-square test • Stratified logistic regression
Continuous • Quality of life (QoL) score • Change in QoL score from baseline	• Mean • Median • Standard deviation • Min • Max	• Mean difference • Mean difference in change from baseline • Hedges's g	• Linear regression • Mixed model for repeated measures (MMRM) model
Time-to-event • Overall survival • Time to relapse • Time to onset of AE	• Events (n, %) • Quartiles • Kaplan-Meier estimates	• Hazards ratio (HR)	• Log-rank test • Cox proportional hazards (PH) model

Table 1. Outcomes, common descriptive statistics, effect measures, and statistical methods used in German dossier.

CASE STUDIES

DICHOTOMOUS OUTCOME ANALYSIS

Basic concepts related to dichotomous outcome analysis such as risk, odds, relative risk, odds ratio, risk difference, and their statistical inferences are fundamental to understanding dichotomous data analysis and interpreting research results.

- Risk refers to the probability of an event occurring within a specific time frame or population. It is commonly used in clinical studies to assess the likelihood of a certain outcome, such as the occurrence of a disease or side effect.
- Odds represent the likelihood of an event occurring relative to the likelihood of it not occurring. Odds can be expressed as the ratio of the probability of an event occurring to the probability of it not occurring.
- Relative risk (RR) also can be named as risk ratio is a statistical concept that compares the risk of an event occurring in two or more groups. RR is often used to determine the relative risk of an outcome associated with a certain exposure or intervention.
- Odds ratio (OR) is another statistical concept that is calculated as the ratio of the odds of an event occurring in one group compared to the odds of the same event occurring in another group.
- Risk difference (RD) is the absolute difference in the risk of an event occurring between two groups. Risk difference is often used to determine the effectiveness of different treatments or interventions. By comparing the risk difference between two groups, researchers can determine the absolute reduction or increase in the risk of an event occurring.
- Exact confidence intervals (CI) are commonly used in German dossier. Exact CI for a dichotomous outcome refers to a method of constructing a CI that takes into account the exact distribution of the data, rather than relying on an approximation. For dichotomous outcomes, such as the occurrence of a certain event or response to treatment, the exact CI can be calculated using various methods such as the Clopper-Pearson method, the Wilson score method, and the Agresti-Coull method. The Clopper-Pearson method is one of the most commonly used methods for constructing exact CI for dichotomous outcomes. It is based on the binomial distribution and provides a lower and upper bound for the true proportion or probability of the outcome.
- Statistical inference of dichotomous outcomes in clinical trials involves analyzing data from two groups, typically a treatment group and a control group, to determine if there is a significant difference in the proportion of individuals experiencing a particular outcome, such as a disease or adverse event. This is typically done using hypothesis testing and confidence intervals. Hypothesis testing involves comparing the observed difference in proportions to a null hypothesis, while confidence intervals provide a range of plausible values for the true difference in proportions. These methods allow researchers to draw conclusions about the effectiveness of a treatment and the likelihood of chance differences between groups. Logistic regression model and Cochran-Mantel-Haenszel method with stratification factors are often used.

Sample Data and Results Interpretation

The common outcomes of dichotomous data analysis in the submission of German dossier include clinical response rate and adverse event occurrence. Table 2.1 is an example output of clinical response rate.

This table presents results from a mock clinical trial comparing a control group (N=58) to a treatment group (N=66) in terms of responders and non-responders. Responders are defined as subjects who achieved clinical response, and non-responders are subjects who did not achieve clinical response or randomized subjects with no efficacy data available. The table presents the number and percentage of responders and non-responders in each group, along with 95% confidence intervals (CI) for the percentage of responders. The unadjusted difference and risk ratio of responders between the treatment and control groups, along with their corresponding 95% CIs, are also computed by the normal

approximation method. The table further presents the adjusted difference, risk ratio and odds ratio of responders between the treatment and control groups, which are computed using the Cochran-Mantel-Haenszel (CMH) weighted average approach, along with their corresponding 95% CIs and p-values after adjusting for the stratification factors if homogeneity is met. The risk differences are larger than 0, risk ratios and odds ratios are larger than 1, suggesting higher clinical response in treatment group. However, the p-values for the adjusted difference and general association are 0.436 and 0.437, which are larger than 0.05 indicates there is no statistically significant difference between the treatment and control groups in terms of the proportion of responders.

Table 2.1
Proportion of Patients Who Achieved Clinical Response
(Randomized Set)

	Control (N=58)		Treatment (N=66)	
	n (%)	95% CI	n (%)	95% CI
Responders	46 (79.3)	(68.89, 89.74)	55 (83.3)	(74.34, 92.32)
Non-responders	12 (20.7)		11 (16.7)	
Unadjusted Difference in Proportion of Responders			4.02	(-9.74, 17.79)
Adjusted Difference in Proportion of Responders			5.51	(-8.37, 19.39)
Unadjusted Risk Ratio of Responders			1.05	(0.89, 1.25)
Adjusted Risk Ratio of Responders			1.07	(0.90, 1.27)
Adjusted Odds Ratio of Responders			1.44	(0.58, 3.59)
p-value: Adjusted Difference			0.436	
p-value: General Association			0.437	

Table 2.2. Output of dichotomous outcome analysis.

Sample Code

- Frequency, percentage, and corresponding 95% CI

```
proc freq data=indata order=data;
  by grp;
  tables aval/binomial alpha=0.05;
  weight count/zeros;
  exact binomial;
  output out=ci binomial;
run;
```

- Cochran-Mantel-Haenszel (CMH) weighted average approach

```
ods output commonRelRisks=crr_mh CommonPdiff=cpd_mh breslowDaytest=pv_bd
cmh=pv_cmh commonpdifftests=pv_mh;
proc freq data=indata order=formatted;
    tables &stratum*grp*avaln/ relrisk CMH missing commonriskdiff(TEST=MH
    CL=MH);
run;
```

- Minimum risk weight method

```
ods output CommonPdiff=cpd_mr CommonPdiffTests=pv_mr;
proc freq data=indata order=formatted;
  tables  &stratum*grp*avaln/missing commonriskdiff(TEST=MR CL=MR);
```

```
run;
```

- Unadjusted relative risk

```
ods output RelativeRisks=crr_un RiskDiffCol1=cpd_un;
proc freq data = indata;
  table grp*avaln / riskdiff(CL=(WALD)) RelRisk alpha=0.05 chisq CMH;
run;
```

CONTINUOUS OUTCOME ANALYSIS

In German dossier submissions with continuous outcomes, several basic statistical concepts are commonly used to analyze the data and draw conclusions, including observed value and their change from baseline, Hedges's g, and least squares mean. The continuous outcome is often analyzed by linear regression modeling, such as Mixed-Effects Models for Repeated Measures.

- Observed value and their change from baseline are simple statistical concepts used to describe the numerical values obtained for a specific endpoint at different time points during the clinical trial. Baseline values are typically obtained before the start of the intervention, and changes from baseline are often used to evaluate the efficacy of the treatment.
- Hedges's g is a standardized measure of effect size that is commonly used to compare the means of two groups. It takes into account the sample size and standard deviation of the groups being compared. It can be used to determine the magnitude of the effect of a treatment. A value of 0.2 is considered a small effect size, value of 0.5 is considered a medium effect size, and value of 0.8 or higher is considered a large effect size.
- Least squares mean is a statistical concept used to estimate the mean value of an endpoint while controlling for other factors, such as baseline values and treatment group. It is often used in linear regression models and can provide a more accurate estimate of the mean value than simply calculating the arithmetic mean.
- Linear regression is a statistical method that analyzes the relationship between one or more independent variables and a dependent variable. In clinical trials, linear regression is often used to analyze the relationship between the treatment group and the continuous endpoint of interest, while controlling for other factors such as baseline values and potential confounding variables. The analysis provides an estimate of the treatment effect, as well as a confidence interval and p-value to indicate the statistical significance of the observed effect.
- Mixed-Effects Models for Repeated Measures (MMRM) is a type of linear mixed-effects model that is specifically designed for analyzing longitudinal data with repeated measures. MMRM models account for within-subject correlations and missing data. It can be used to estimate the mean trajectory of the continuous endpoint over time while adjusting for the treatment effect and other covariates. MMRM models can also be used to estimate the treatment effect at specific time points. It can provide a more accurate estimate of the treatment effect compared to traditional linear regression models.

Sample Data and Results Interpretation

One of the common continuous outcomes analyzed in German dossier is the quality of life (QoL) scores. Table 2.2 is an example output of QoL score over time. Descriptive statistics, n, mean, standard deviation, min and max, Hedges's g, least squares means for change from baseline, differences in LS means in change from baseline, 95% CIs and P-values are provided.

This Table provides the results of a clinical trial with continuous outcomes. The trial compares control and treatment, with respect to their effect on the observed value and change from baseline. At baseline, control has a mean value of 65.9 with a standard deviation of 18.26, while treatment has a mean value of 71.3 with a standard deviation of 20.66. The mean difference between the two groups is 5.1, but the p-value of 0.194 indicates that this difference is not statistically significant. The Hedges's g value of 0.26

with a confidence interval of -0.1 to 0.6 also suggests a small effect size. The actual values and change values at week 4 can be interpreted in the same manner.

Table 2.2
Summary of QoL Scores by Study Week and Treatment Group
(Modified Randomized Set)

	Actual Value			Change from Baseline		
	Control N=58	Treatment N=66	Treatment vs Control LS Mean Difference	Control N=58	Treatment N=66	Treatment vs Control LS Mean Difference
Baseline						
n	49	57				
Mean (std dev)	65.9 (18.26)	71.3 (20.66)	5.1			
Median	68.0	75.0				
Min, Max	13, 96	0, 100				
95% CI for Mean	(60.6, 71.1)	(65.8, 76.8)	(-2.5, 12.7)			
P-value			0.194			
Hedges's g			0.26 (-0.1, 0.6)			
Week 4						
n	38	50		36	45	
Mean (std dev)	73.1 (18.30)	66.3 (24.31)	-8.2	6.8 (20.68)	-6.9 (21.70)	-9.7
Median	78.0	75.0		6.5	0.0	
Min, Max	20, 100	9, 100		-35, 79	-83, 32	
95% CI for Mean	(67.1, 79.1)	(59.4, 73.2)	(-16.3, -0.2)	(-0.2, 13.8)	(-13.4, -0.4)	(-18.2, -1.2)
P-value			0.058			0.033
Hedges's g			-0.42 (-0.9, 0.0)			-0.5 (-0.9, -0.0)

Table 2.2. Output of continuous outcome analysis.

Sample Code

- Least squares (LS) means, LS mean difference, 95% CIs and p-values

```
ods output lsmeans=lsm_aval diffs=dif_aval;
proc mixed data=indata;
  class usubjid grp(ref=first) avisitn;
  model aval=grp avisitn grp*avisitn base;
  random intercept / subject=usubjid;
  repeated avisitn / subject=usubjid type=ar(1) ri;
  lsmeans grp*avisitn / cl pdiff;
run;
```

- Hedge's g

```
sdp1=(n1 - 1)*(se1*sqrt(n1))**2;
sdp2=(n2 - 1)*(se2*sqrt(n2))**2;
ssd=sdp1 + sdp2;
sdp=sqrt(ssd/(n1 + n2 - 2));
df= n1 + n2 - 2;
g=estimate/sdp;
stdg=sqrt(((n1 + n2)/n1*n2) + (g*g)/(2* (n1 + n2)));
tcrit=tinv(1-0.05/2, df);
tvalue=g/stdg;
glo=g - tcrit*stdg;
gup=g + tcrit*stdg;
pvalue=(1-probt(abs(tvalue), df))*2;
```

TIME-TO-EVENT DATA ANALYSIS

German dossier submissions often involve time-to-event outcome, such as overall survival, all-cause mortality, time to relapse, and time to onset of adverse events, *etc*. Overall survival refers to the time from the start of a clinical trial until the patient's death from any cause. It is a measure of how long patients survive after they have been diagnosed with a disease or received a particular treatment. All-cause mortality, on the other hand, refers to the number of deaths from any cause that occur during the course of the clinical trial, regardless of whether the deaths are directly related to the disease being studied or not. It is a broader measure that takes into account all factors that may contribute to a patient's death. Time to relapse refers to the time from treatment initiation until a patient experiences a relapse of their disease. The time to onset of adverse events refers to the duration between the initiation of treatment and the first occurrence of any adverse event in a patient. Analyzing such data requires specialized statistical techniques, collectively known as survival analysis. The goal of survival analysis is to estimate the probability of an event occurring over time and to investigate the factors that influence the time to event.

- The Kaplan-Meier (KM) analysis is one of the most commonly used techniques in survival analysis. It provides a non-parametric estimate of the survival function. The KM estimates take into account the censoring of data, which occurs when the event of interest has not occurred by the end of the study. The number of patients still at risk at each time point is often reported using a "number at risk" table. The quantiles from the survival function are also important in survival analysis. For example, the median survival time represents the time at which 50% of patients have experienced an event of interest. Other quantiles, such as the 25th and 75th percentiles, can provide additional information about the distribution of survival times. The log-rank test is a commonly used non-parametric test for comparing survival curves between two or more groups, and the resulting p-value can indicate whether the observed differences are statistically significant.
- The Cox proportional hazards model is a commonly used parametric model for survival analysis. It allows for the estimation of the hazard ratio (HR) and associated 95% confidence interval (CI), which can provide information on the relative risk of experiencing the event of interest between two groups.

Sample Data and Results Interpretation

Table 2.3 is an example output of time to event analysis. This table provides a summary of the all-cause mortality data from a mock clinical trial comparing two groups, referred to as control and treatment. The trial includes 58 patients in the control group and 66 patients in the treatment group. The number of subjects who died is 3 (5.2%) in the control group and 7 (10.6%) in the treatment group. The number of subjects censored, which means they are still alive at the end of the study, is 55 (94.8%) in the control group and 59 (89.4%) in the treatment group. The observed event time for those who died is 143.0 days (median) in the control group, with a range of 54 to 149 days, and 134.0 days (median) in the treatment group, with a range of 3 to 158 days. The Kaplan-Meier estimates of time to death are calculated, and the results show that the median time to death is not reached in either group (NE means not estimated). The log-rank p-value is 0.311, indicating that there is no significant difference in the time to death between the two groups. The hazard ratio (95% CI) comparing the treatment group to the control group is 1.81 (0.460, 7.146), which suggests that the treatment group has a slightly higher risk of death than the control group, but the difference is not statistically significant. The median (95% CI) of follow-up is the same in both groups at 141.0 days. The Kaplan-Meier estimates of time to death at various time points are also presented, with the number at risk for each group. The estimates suggest that the treatment group has a slightly lower survival rate than the control group at some time points. It is important to note that the number of patients at risk decreased over time due to patient deaths and study withdrawals.

Table 2.3
Time To All-cause Mortality
(Randomized Set)

	Control (N=58)	Treatment (N=66)
Number of Subjects Who Died [n (%)]	3 (5.2)	7 (10.6)
Number of Subjects Censored [n (%)]	55 (94.8)	59 (89.4)
Observed Event Time for Those Who Died (days) [Median (min, max)]	143.0 (54, 149)	134.0 (3, 158)
Kaplan-Meier Estimates of Time to Death (days) [95% CI]		
25th	NE (143.0, NE)	158.0 (156.0, NE)
50th (Median)	NE (149.0, NE)	NE (156.0, NE)
75th	NE (NE, NE)	NE (158.0, NE)
Log-rank p-value		0.311
Hazard Ratio (95% CI)		
Treatment Group: Treatment vs Control		1.81 (0.460, 7.146)
Median (95% CI) of Follow up (days)	141.0 (140.0,142.0)	141.0 (140.0,142.0)
Kaplan-Meier Estimate (95% CI) [No. at risk]		
28 days	100.0 (100.0, 100.0) [n= 55]	98.5 (89.7, 99.8) [n= 63]
56 days	98.1 (87.1, 99.7) [n= 51]	98.5 (89.7, 99.8) [n= 62]
84 days	98.1 (87.1, 99.7) [n= 51]	95.3 (86.1, 98.5) [n= 60]
112 days	98.1 (87.1, 99.7) [n= 50]	95.3 (86.1, 98.5) [n= 59]
140 days	98.1 (87.1, 99.7) [n= 39]	91.4 (80.2, 96.4) [n= 41]
168 days	79.9 (39.2, 94.7) [n= 2]	57.1 (15.5, 84.5) [n= 2]
196 days	NE (NE, NE) [n= 0]	57.1 (15.5, 84.5) [n= 1]
224 days	NE (NE, NE) [n= 0]	NE (NE, NE) [n= 0]

Table 2.3. Output of time to event analysis.

Sample Code

- Kaplan-Meier (KM) analysis

```
proc lifetest data=indata atrisk method=km timelist=&tlist reduceout
outsurv=sci;
  time aval*cnsr(1);
  strata grp/test=logrank;
  ods output Quartiles=qt;
  ods output HomTests=logrankpv;
  ods output ProductLimitEstimates=kmest;
run;
```

- Median (95% CI) of follow up

```
proc lifetest data=indata method=km;
   time aval*cnsr(0);
   strata grp;
   ods output quartiles=qtfu;
run;
```

- Hazard ratio

```
proc phreg data=indata alpha=0.05;
  class grp(ref=first);
```

```
    model aval*cnsr(1)=grp/rl;
    strata &stratum;
    hazardratio grp/diff=ref;
    ods output parameterestimates=hr;
  run;
```

CONCLUSION

German dossier submission process is a rigorous and detailed process that requires a thorough statistical analysis of clinical trial data to demonstrate the safety, efficacy, and quality of new drugs. This paper provides a comprehensive summary of these statistical methods, including basic concepts, sample data, analysis codes, and interpretation of results. Overall, a robust statistical analysis is a critical component of the German dossier submission process and is essential in ensuring that new drugs are safe and effective for patients.

REFERENCES

Walker, Glenn A., and Jack Shostak. 2010. Common Statistical Methods for Clinical Research with SAS® Examples, Third Edition. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

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

Meiling Gao
BioPier Inc.
meiling@biopier.com

Xi Qian
BioPier Inc.
xqian@biopier.com

Hui Mao
BioPier Inc.
hmao@biopier.com