Statistical Analysis of Adverse Events in Randomized Clinical Trials Using SAS

Dongsun Cao, ICON Clinical Research, Durham, NC
Xiaomin He, ICON Clinical Research, North Wales, PA

ABSTRACT

Adverse events are generally tabulated and listed in statistical reports in randomized clinical trials. This paper introduces how to apply statistical methods to assess adverse events using SAS. In addition to the crude percentage (rate), adverse events are also adjusted by exposure time or multiple repeat. The paper also introduces the commonly used chi-square test and Fisher's exact test, and other statistical methods such as Cochran-Armitage trend test, nonparametric approach and generalized linear mixed model. The application of statistical methods depends on study designs, data type and investigation purpose. Several clinical trial examples such as dose-response studies, cross-over trials, are analyzed using SAS for illustration purposes.

KEY WORDS

Treatment-emergent adverse events, Chi-square test, Fisher's exact test, Cochran-Armitage trend test, Exposure-adjusted incident rate, Nonparametric approach, Generalized linear mixed model, Cross-over trials.

BACKGROUND

According to ICH guidance for Clinical Safety Data Management: Definitions and Standards for Expedited Reporting, an adverse event (AE) is defined as

> *Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment.*

An AE can therefore be any unfavourable and unintended sign (including an abnormal laboratory finding, for example), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

Existing statistical methodology to evaluate the efficacy of a study drug in clinical trials is well-developed. However, the research on the safety profiles, including AE data, is very limited. In a randomized clinical trial requiring statistical analysis, sample size determination generally requires patient enrollment to meet the efficacy criteria, while safety profiles are only for exploratory purposes. Therefore, the majority of randomized clinical trials simply request AEs to be tabulated and/or listed without any inferential statistical analysis. However, recent research starts to demand rigorous analysis on AE datasets in randomized clinical trials (O'Neil, 1995; Chuang-Stein, 1998). When preparing the statistical reports, clinical investigators and medical writers often ask biostatisticians and SAS programmers to provide statistical analysis on safety profiles, even if these results are not requested for a submission to regulatory authority. It is therefore necessary for practitioners to apply statistical approaches to AE and other safety profiles.

In this paper, we first introduce some basic AE analyses used in randomized clinical trials. Various statistical approaches could be used in AE analysis depending on data type. Depending on the study design and investigation purpose, we also apply several other statistical approaches to explore various clinical trials, including dose-response studies, cross-over trials etc.

CHI-SQUARE TEST AND FISHER'S EXACT TEST

The crude percentage (rate) is the simplest and also the most commonly used statistics for summarizing AEs in clinical trials. Contingency tables are often used to analyze AEs. Chi-square test and Fisher's exact test are two most popular approaches.

A chi-square test (also called chi squared test or χ^2 test) is any statistical hypothesis test in which the sampling distribution of the test statistic is a chi-square distribution when the null hypothesis is true, or asymptotically true. This means that the

sampling distribution (if the null hypothesis is true) can be made to approximate a chi-square distribution as closely as desired when the sample size is large enough. The FREQ procedure in SAS provides several chi-square tests and the default one is Pearson's Chi-Square test for independence (or test for goodness of fit).

There are 3 assumptions of a chi-square test: randomness, independence, and large sample size. We need to ensure that the AE data are randomly sampled from a fixed distribution or population. Since we limit our discussion to randomized clinical trials, this assumption is easily satisfied. Independence is a critical assumption when applying a chi-square test but it is often ignored in practice. There are lots of misuses of this approach in clinical trials. We will use a cross-over trial as an example to illustrate it in a later section. There are two requirements for sample size: one is that the total sample should be sufficiently large to ensure accurate inference; the other is that the expected cell counts are adequately large to avoid erratic boundary extremities performance. A difficulty in programming is to determine how large a sample size is sufficient. A rule of thumb is for the expected cell counts to be at least 5, based on a theorem assuming normal approximation of binomial proportion at $np \geq 5$ and $n(1-p) \geq 5$.

We use a 2 X 2 contingency table as an example here:

		Treatment Group		
		Control	Active	Total
Response	Yes	n_{11}	n_{12}	n_{1+}
	No	n_{21}	n_{22}	n_{2+}
	Total	n_{+1}	n_{+2}	n_{++}

The test statistic is given by:

$$\chi^2 = \sum\sum \frac{(\text{observed cell count} - \text{expected cell count})^2}{\text{expected cell count}} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}},$$

where the expected cell count is given by $\hat{n}_{ij} = (n_{i+} \bullet n_{+j})/n_{++}$. The formula could be extended to a $r \times c$ contingency table where either r or c is greater than 2, and the degree of freedom is $(r-1) \times (c-1)$. Some practitioners tend to apply the Yates' correction of continuation when the sample size or the expected cell counts are too small. The FREQ procedure in SAS offers the continuation adjusted chi-square tests, but limited to 2 X 2 contingency tables.

Fisher's exact test is an alternative statistical test to the chi-square test in the analysis of contingency tables when sample sizes are small. It is named after its inventor, R. A. Fisher, and is one of a class of exact tests, so called because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, or asymptotically. FREQ procedure in SAS calculates and prints Fisher's exact test for 2 X 2 tables by default. For larger tables (either r or c is greater than 2), exact tests must be explicitly requested with the EXACT option on the TABLES statement.

A rule of thumb is, if any of the cells has expected count less than 5 ($\hat{n}_{ij} < 5$), a chi-square test will be inappropriate and we should use the Fisher's exact test instead.

Example 1: Consider a randomized clinical trial with two treatment groups. 40 subjects receive control and 65 subjects receive active treatment. The trial lasts for about 3 months. The AE data are attached in the Appendix section. Note that not every subject has an AE and some subjects may have two or more AEs during the study.

In order to comprehensively assess AE, the following information is generally of interest to the investigators:

- Treatment-Emergent AE (TEAE)

- Drug-related AE

- Serious adverse event (SAE)

- Higher intensity AE

- Death

All definitions are self-explanatory except for TEAE. In general, a TEAE is defined as any event not present prior to the initiation of the study drug or any event already present that worsens in either intensity or frequency following exposure to the study drug. TEAE provides more relevant and useful information than a general AE to the investigators. Therefore, most AE tables in statistical reports should be based on TEAE. In this sense, any unfavorable and unintended sign prior to dosing

but not worsening during the treatment will be automatically excluded from the analysis. The programming practice to code a TEAE is comparing AE's onset date/time and first study drug administration date/time. If the AE has the same intensity before and after the first drug administration, then it is not counted as TEAE. Note that a TEAE is not necessarily a drug-related AE.

Before assessing AEs between groups, we need to present the original AE data with a contingency table. Because not every subject has an AE during the study, the subject level data (e.g. demographic data) should be used to calculate each cell count. The derived dataset, AE_CONTTBL, is also attached in the Appendix section.

In the FREQ procedure, we use CHISQ and FISHER options in the TABLES statement to output two separate datasets with results from chi-square tests and Fisher's exact tests, and also use EXPECTED option to output another table of expected cell counts, which is similar to the contingency table with observed cell counts.

```
proc freq data=ae_conttbl;
  by ord;
  tables outcome*trt / CHISQ FISHER EXPECTED;
  weight freq;
  ods output FishersExact=Exact(where=(Name1='XP2_FISH'));
  ods output ChiSq=ChiSq(where=(Statistic='Chi-Square'));
  ods output CrosstabFreqs=Freqs(where=(Expected >.z));
run;
```

SAS output generally gives a warning as "XX% of the cells have expected counts less than 5. Chi-Square may not be a valid test." The warning message could be used as guidance for choosing between chi-square tests and Fisher's exact tests.one of these two tests. However, it couldn't be used for programming purpose. Instead, we will use the output data by EXPECTED option to investigate whether any cell count is less than 5. We recommend using the chi-square test by default; if any category has a cell count less than 5, then the Fisher's exact test will be used.

```
    Category                              Control      Active      P-value
                                          (N=40)       (N=65)


    Any TEAE                              17 (42.5)    13 (20.0)    0.013
    Drug-Related TEAE                      4 (10.0)     6 ( 9.2)    1.000
    Grade 3 or Higher TEAE                 3 ( 7.5)     1 ( 1.5)    0.153
    Grade 3 or Higher Drug-Related TEAE    2 ( 5.0)     1 ( 1.5)    0.556
    Serious Adverse Event                  3 ( 7.5)     2 ( 3.1)    0.367
    On-study Death                         1 ( 2.5)     0 ( 0.0)    0.381


    Note: P-value is based on a chi-square test. If any cell has expected counts
    less than 5, then the Fisher's exact test is used instead.
```

Note that there are many controversies over the choice between chi-square test (with or without correction of continuation) and Fisher's exact test (Liddell,1976; D'Agostino et al.,1988). In addition, the criteria of expected counts being less than 5 are arbitrary, although it is an often used practice and appears in SAS warning message. We could not make a universal judgment on which method is more appropriate in clinical trails. After all, clinical trials are different in study design and investigation purpose. The choice of the statistical method should be decided by the clinical trial itself and the nature of data. The methods we introduce here are for educational purposes only.

COCHRAN-ARMITAGE TREND TEST IN DOSE-RESPONSE STUDIES

If the sample size is large enough, the chi-square test is appropriate for contingency table analysis; however, it is less powerful for detecting a linear association (if it exists) because its power is dispersed over a greater number of degrees of freedom. For example, in a clinical trial, dose of a treatment can be ordered as 'none' (placebo), 'low', 'medium', and 'high', and we may suspect that adverse effect of the treatment become stronger as the dose increases. In this case, Cochran-Armitage trend test (Cochran, 1954; Armitage, 1955) is a more powerful approach than chi-square test to assess a trend

among adverse event rates. Lots of researches have indicated that the power of the Cochran-Armitage trend test also holds when the sample size is small, in which case the exact permutation method is used.

We use a 2 X 4 contingency table as an example:

		Treatment Group				
		Placebo	50 mg	100 mg	150 mg	Total
Response	Yes	n_{11}	n_{12}	n_{13}	n_{14}	n_{1+}
	No	n_{21}	n_{22}	n_{23}	n_{24}	n_{2+}
	Total	n_{+1}	n_{+2}	n_{+3}	n_{+4}	n_{++}

The test statistic is given by:

$$T = \frac{n_{++}\sum_{j=1}^{4} n_{1j}\left(s_j - \bar{s}\right)}{\sqrt{n_{1+}n_{2+}\sum_{j=1}^{4} n_{+j}(s_j - \bar{s})^2}} \, ,$$

where s_j is the score assigned to each column j and $\bar{s} = \sum_{j=1}^{4} n_{+j}s_j / n_{++}$. The null hypothesis for the Cochran-Armitage

test is no trend, which means that the binomial proportion is the same for all doses of the treatment. Under the null hypothesis, the trend statistic has an asymptotic standard normal distribution. Similar to Fisher's exact test being an alternative to chi-square test, if the data normality assumption is questionable, the Cochran-Armitage trend test can use the exact permutation method (which is not discussed in this paper).

The FREQ, MULTTEST and LOGISTIC procedures in SAS/STAT can implement Cochran-Armitage trend test and give the same p-values (Liu, 2007). Also, these three procedures can provide both asymptotic and exact permutation tests. To simplify, we use the FREQ procedure to illustrate how to assess a trend among adverse event rates.

Example 2: Consider a randomized clinical trial in which approximately 400 patients are randomly assigned into placebo group and new drug groups (50 mg, 100 mg and 150 mg) in a fashion of 1:1:1:1. The investigator is interesting in assessing whether there is a linear increase of risk of adverse effects of the drug when the doses increase.

```
proc freq data=ae_doseresp;
   tables dose*outcome/ TREND;
   weight count;
   ods output TrendTest=TrendTest;
run;
```

In the Cochran-Armitage trend test, the choice of scores is of key importance to the investigator. In this example, we used the dose of treatment as the scores for the trend test, in which SCORES=TABLES is by default in the TABLES statement. Table scores are the values of column levels: 0 (placebo), 50 mg, 100 mg and 150 mg, which is similar to the scores of 0, 1, 2 and 3. We can also use other SCORES options. For example, if the investigator thinks the linear trend is based on logarithm other than of natural base, then dose variables could be formatted using the internal numeric values. The SCORES option also provides nonparametric tests by assigning RANK, RIDIT, or MODRIDIT. The scores have to be meaningful and reflecting the degree of difference among the dose levels. The test results do not change with the location and scale of the scores.

Category	Stats	Placebo	50 mg	100 mg	150 mg
TEAE	n(%)	7 (7.2)	6 (6.3)	13 (12.5)	19 (19.0)
	Test for Trend				-2.7740
	Two-sided P-value				0.006

In this example, it is easy to observe an increasing adverse event rate when the doses increase. Results of the Cochran-Armitage trend test indicate a statistically significant linear trend among doses.

EXPOSURE-ADJUSTED INCIDENT RATE

For some investigators, the crude percentage (rate) may be over-simplified as it ignores the potential occurrence time effects when assessing adverse event difference between treatment groups. The incident rate, which uses the total person-time in a treatment group as the denominator, is often used to take this into consideration.

One of the measures is the exposure-adjusted incidence rate (EAIR), defined as the number of subjects with a specific event divided by the total exposure-time among the subjects in the treatment group and at risk of an initial occurrence of the event (Liu et al., 2006; Siddiqui, 2009). Specifically,

$$EAIR = \frac{n}{T} = \frac{n}{\sum t_i},$$

where n is the number of subjects with events, t_i is a subject's exposure time under a given unit and T is the total exposure time under a given time unit of the subject. If a subject has multiple events, the t_i is the time of the first event. For a subject with no event, the t_i is censored at the last follow-up time for that subject. The definition of EAIR is based on the assumption that the occurrences of a specific event are following an independent Poisson process, so the event occurs with a constant rate over time. Therefore, the EARI is most appropriate when the hazard rate of the specific event is relatively constant over the duration of the study. It would not be an appropriate measure for events that usually occur early in the study (e.g. allergic reactions), events whose hazard rates decrease over time, or events that have a latent period (e.g. cancer).

We still use the same 2 X 2 contingency table as before, the test statistic is give by

$$E = \left(\frac{n_{11}}{T_1} - \frac{n_{12}}{T_2} \right) / \sqrt{\frac{n_{11}}{T_1^2} + \frac{n_{12}}{T_2^2}},$$

which follows the standard normal distribution under the null hypothesis. Therefore, the normal approximation confidence interval for the difference between two groups is

$$\left(\frac{n_{11}}{T_1} - \frac{n_{12}}{T_2} - Z_{\alpha/2} \cdot \sqrt{\frac{n_{11}}{T_1^2} + \frac{n_{12}}{T_2^2}}, \frac{n_{11}}{T_1} - \frac{n_{12}}{T_2} + Z_{\alpha/2} \cdot \sqrt{\frac{n_{11}}{T_1^2} + \frac{n_{12}}{T_2^2}} \right)$$

Aside from normal approximation, other statistical methods, such as two-by-two table method, Miettinen and Nurminen (MN) method and conditional method, are also often used to assess EAIR among treatment groups. Compared to other methods, the normal approximation provides a reasonably good coverage of the difference (Liu et al., 2006).

Example 3: Consider Example 1 again to compare between two group using EAIR. The first event exposure time will be calculated for each subject with adverse events. For other subjects without any events, their exposure times are the last record time in the follow-up. Since this study lasts 3 months, to simplify, we assume those without adverse events have exposure time of 92 days. The EAIR for each treatment group and their difference are calculated as below:

```
    Stats                                    Control      Active       P-value
                                             (N=40)       (N=65)

    Exposure-Adjusted Incidence Rate (EAIR)  0.0070       0.0026       0.008
    95% CI of Treatment Difference                        (0.0008, 0.0080)
```

There are no available procedures in SAS for EAIR calculation and comparison. Since the test statistic follows the standard normal distribution, it would be easy to use DATA steps to derive confidence intervals or p-values by using PROBNORM function. In Example 3, the conclusions based on the crude rate and EAIR are the same; that is, the overall adverse effect on the active treatment group is less than the placebo group.

NONPARAMETRIC APPROACH IN MULTIPLE EVENTS DATA

The assumption of EAIR isn't always realistic at the constant occurrence rate of events. It is also controversial to assume multiple events are independent in a moderate to long-term clinical trial; especially, when the multiple events are repeat events of the same type.

A nonparametric method called the mean cumulative function (MCF) can be used to analyze multiple/repeated occurring AEs to subjects during the entire exposure period in clinical trials. By definition, the recurrent AE data consist of the inter-event times of repetition of the same or different types for each subject. The measurements across subjects are considered to be statistically independent, but the times between AEs with a subject are not necessarily independent. The MCF contains the information of interest in the analysis of recurrent data.

Assume $M(t)$ is the mean cumulative number of AEs up to time t. Its nonparametric estimator $\hat{M}(t)$ is a staircase function that is flat between event times; therefore, $\hat{M}(t)$ involves no assumption about the form of $M(t)$. At time t, a fraction of the subjects has accumulated zero recurrences, another fraction has accumulated one recurrence, another fraction has accumulated two recurrences, and so on. This distribution differs at different time t and has a mean $M(t)$. The estimator $\hat{M}(t)$ and its confidence limits for the recurrence data are analogous to (but different from) the Kaplan–Meier (KM) estimate. The KM analysis includes only the time to first adverse event, while MCF analysis facilitates the inclusion of all recurrent adverse events data. The derivative $\hat{m}(t) = d\hat{M}(t)/dt$ is called the instantaneous recurrence rate or intensity rate of an AE event at time t.

Both RELIABILITY procedure in SAS/QC and PHREG procedure in SAS/STAT can compute and plot $\hat{M}(t)$ and its confidence limits. If there are two or more treatment groups in the study, the pointwise confidence limits in the RELIABILITY help assess the significance of the difference between any two groups, but the PHREG procedure offers direct and flexibility for this assessment. In statistical terminology, the PHREG procedure uses a semiparametric approach, which is similar to the proportional hazards model for lifetime data. The model is, for each observation,

$$M(t) = M_0(t)\exp(X'\beta),$$

where X' is a vector of time invariant covariates and $M_0(t)$ is a baseline MCF. If no regression parameters are specified, the $M_0(t)$ PHREG computes is identical to what MCF PROC RELIABILITY computes (Johnston and So, 2003). Therefore, you can specify the treatment parameter as the unique β so that a test for the treatment effect $H_0 : \beta = 0$ can be conducted.

Example 4: Consider Example 1 again to compare two group using MCF. Different from Example 3, we will calculate recurrence time, censored time and follow-up status for each subject, even if a subject doesn't have any event. That is, a subject with K events contributes $K+1$ observations to the input data set; here K might be 0. The k th observation of the subject identifies the time interval from the $(k-1)$ th event or time 0 (if $k = 1$) to the k th event, $k = 1,..., K$. The $(K+1)$ th observation represents the time interval from the K th event to time of censorship. The derived data set should contain the following variables:
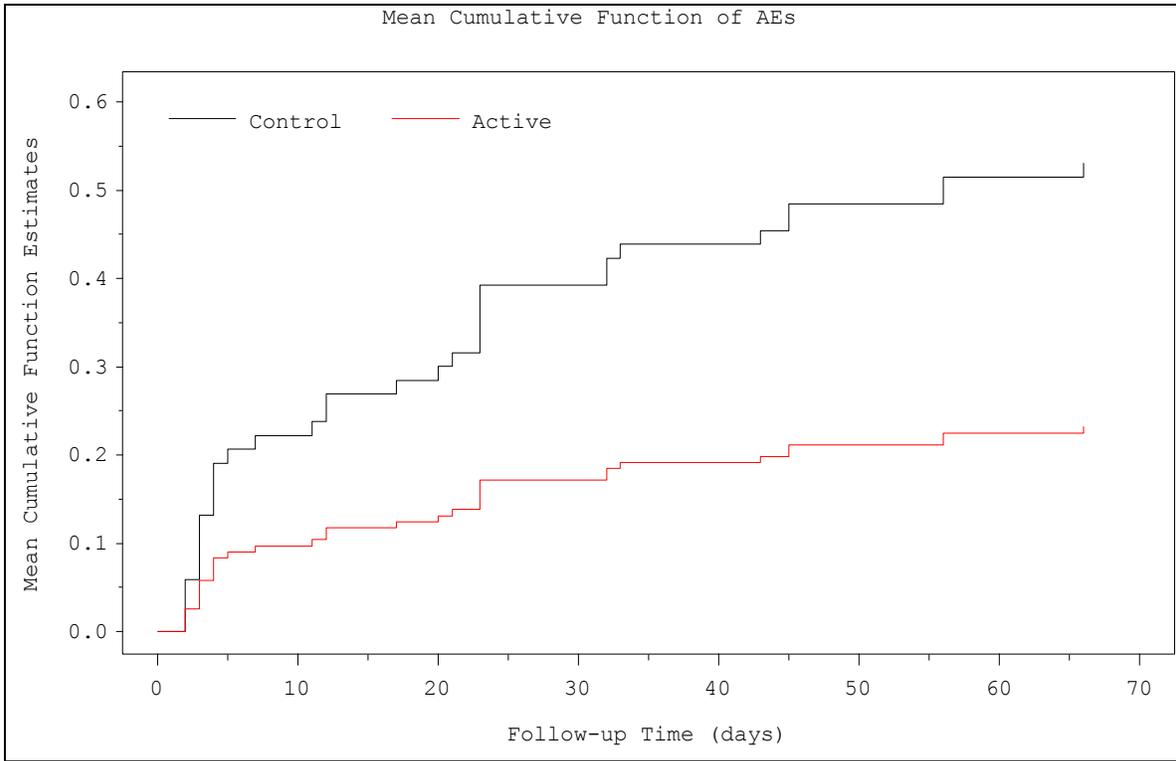
- TStart variable to represent the $(k-1)$ th recurrence time or the value 0 if $k = 1$

- TStop variable to represent the k th recurrence time or the follow-up time if $k = K+1$

- Status variable indicating whether the TStop time is a recurrence time or a censored time; for example, Status=1 for a recurrence time and Status=0 for censored time.

To simplify, aside from the assumption of follow-up (or exposure) day of 92 days for all subjects without AEs, we also assume the recurrence of AEs regardless of body system or preferred term. That is, the repeat events include all events for each subject.

```
data IN2;
  Trt=1; output;
  Trt=2; output;
run;

proc phreg data=aeall covs(aggregate) covm PLOTS(overlay)=MCF;
  model (Tstart,Tstop)*Status(0)=Trt;
  baseline covariates=IN2 out=OUT2 cmf=_all_ / nomean;
  id subjid;
```

```
run;
```

Mean Cumulative Function of AEs


```
            Testing Global Null Hypothesis: BETA=0

        Test                    Chi-Square      DF      Pr > ChiSq

        Likelihood Ratio           5.9360        1        0.0148
        Score (Model-Based)        6.2125        1        0.0127
        Score (Sandwich)           4.1436        1        0.0418
        Wald (Model-Based)         5.8678        1        0.0154
        Wald (Sandwich)            4.9644        1        0.0259

            Analysis of Maximum Likelihood Estimates
                with Sandwich Variance Estimate

                     Parameter   Standard    StdErr
        Parameter   DF  Estimate     Error     Ratio   Chi-Square   Pr > ChiSq

        Trt          1  -0.82759   0.37143     1.087      4.9644       0.0259
```

Results indicate a statistically significant difference between the control and active treatment groups using MCF. However, this is not a coincident conclusion with Example 1 and 3. Because considering repeated occurring AEs for each subject and also incorporating the exposure times, the MCF method is without doubt more sophisticated and accurate than the crude rate and EAIR. However, the difficulty to understand and complexity of using MCF makes it not popular in practice. MCF method would be recommended only in a randomized clinical trial with recurrent AEs of interest to the investigator. Otherwise, the crude percentage (rate) and simple statistical approached would be sufficient to summarize AEs.

GENERALIZED LINEAR RANDOM MODEL IN CROSS-OVER TRIALS

Compared to parallel group trials in which each subject receives only one treatment during the course of the study, cross-over trails have each subject receiving a sequence of treatments. The most common cross-over trial is using the AB-BA design; that is, the two-period two-treatment design. For subjects in the sequence of AB, they receive treatment A first and then, after a washout period, cross-over to treatment B. The remaining subjects receive treatment B first then cross over to treatment A. The main advantage of cross-over trials is that the treatment effects are evaluated "within-subjects"; therefore, cross-over trials could be powerful tools without recruiting more subjects than the parallel group trials. Due to difficulty of accounting for period effects and carry-over effects, cross-over trials are often limited in pre-clinical trials (animal experiments) and early phase clinical trials. More attention should be given in analyzing results from cross-over trials, especially to a trial with the binary outcome.

Mainland-Gart test and Prescott's test are two most commonly used nonparametric-like methods in cross-over trials with binary outcome (Jones and Kenward, 2003). Utilizing explicitly model-based methods, SAS provide two alternative procedures for assessing categorical data in cross-over trials, one is GENMOD procedure using generalized estimating equations (GEE) and the other is GLIMMIX procedure using generalized liner random models. We here focus on the GLIMMIX procedure, in which the linear random effect is subject-specific.

Example 5: Consider a two-period two-treatment study assessing the bioequivalence of a single dose of 2 different formulations of active treatments (Test A and Reference B) on 24 healthy subjects. Subjects are randomly assigned into AB group and BA group in a fashion of 1:1. The data are also in the Appendix section. To assess the adverse events difference between treatment groups, generally all of information in Example 1 should be given, but we here use "Any TEAE" only for the purpose of illustration.

As we have mentioned in the previous section, a direct comparison between treatments using chi-square test or Fisher's exact test is not valid any more because measurements for each subjects are paired or repeated across period. In this case, independence assumption of response evaluations between two groups is violated. As below table indicates, among 11 subjects having any TEAEs in receiving treatment A, they may have also experience TEAEs in receiving Treatment B. A generalized linear mixed model is given by using GLIMMIX:

```
proc glimmix data=final;
  class subjid sequence peroid trt;
  model TEAE = period trt / solution;
  random sequence(subjid);
  lsmeans trt / cl ilink;
run;
```

In the model, fixed effects include both period effects and treatment effects but without carry-over effects. Carry-over effects often appear when an event to a current treatment is affected by the treatment was applied in a previous period. In our example, we assume the carry-over effects do not exist or have been included in period effects. The sequence (nesting subjects) effects are assumed to be random in the model.

	TreatmentA (N=24)		TreatmentB (N=24)	
Sequence	AB	BA	AB	BA
Any TEAE	6	3	5	6
Any TEAE Total	9		11	
P-value of Treatment Difference			0.534	

Results indicate there is a statistically insignificance of AE difference between Test A and Reference B. Not shown here results indicate the period effect is statistically insignificant (P=0.2191) either. As a matter of fact, a chi-square test or Fisher's exact test without adjusting period effect or considering subject-specific random effect will yield a similar p-value and draw the same conclusion as above. However, their fundamental assumptions are invalid in cross-over trials, and we should avoid any incorrect statistical inference in practice.

CONCLUSION

This paper introduces several statistical methods in assessing AEs. Readers may have been familiar with some of these methods in efficacy analysis instead of in safety analysis. As we have mentioned in the beginning, the logic of statistical analysis of AEs is different from that of efficacy analysis. A powerful study is generally designed for evaluating drug's efficacy, while safety profiles, including AEs, are for exploratory purposes only. In other words, the determination of sample size of the trial (if needed) can't guarantee the statistical power for assessing any variables other than efficacy. Therefore, we must be cautious when explaining the p-values or confidence intervals for AE parameters. We always recommend stressing the results of statistical analysis on AEs as of guidance or exploratory nature, rather than statistically confirmative. Most examples given in the paper have small p-values suggesting statistical significance. In reality, large p-values may be observed in clinical trials, which indicate statistically insignificant safety concerns for the study drug, compared with the standard care or other marketing drugs.

In some randomized clinical trials, such as oncology studies, untoward medical events of disease progress or on-study deaths, are the primary or secondary efficacy endpoints. In this sense, these measurements are similar to AEs, and they are assessed using more sophisticated statistical models. All statistical methods introduced in this paper are limited to AEs in safety profiles, and we also recommend using simple methods if possible.

The statistical methods introduced in this paper are definitely not a comprehensive list. The appropriate applications depend on study design and investigating purpose. SAS can provide lots of procedures to easily apply these methods, or we can use DATA steps or macros to implement them.

APPENDIX

Example 1: There are 40 subjects in Treatment=1 (Control) and 65 subjects in Treatment=2 (Active).

Original Data: AE

```
data ae;
label subjid  ='Subject Number'
      trt     ='Treatment, 1=Control, 2=Active'
      aebodsys='Body System of Event'
      aedecod ='Preferred Term'
      aerel   ='Relatedness, Y=Related, N=Not Related'
      aeser   ='Seriousness, Y=Yes, N=No'
      aetoxgrd='AE Toxicity Grade'
      death   ='Death, 1=Yes'
      aestdy  ='Start Day of AE Relative to First Dose'
      aeendy  ='End Day of AE Relative to First Dose'
      lstfdy  ='Day of Last Follow-up'
      ;
input subjid $ 1-3 trt 5 aebodsys $ 7-33 aedecod $ 34-58 aerel $ 59 aeser $ 62
      aetoxgrd 65 death 68 aestdy 70-72 aeendy 73-76 lstfdy 77-79;
cards;
001 2 Nervous system disorders    Dizziness               Y N 1 . 3  4   56
002 1 Cardiac disorders           Cardiomyopathy          N N 2 . 5  6   78
002 1 Nervous system disorders    Headache                N N 2 . 12 13  78
002 1 Vascular disorders          Malignant hypertension  N N 2 . 2  .   78
002 1 Cardiac disorders           Cardiac valve disease   Y N 2 . 2  34  78
002 1 Nervous system disorders    Headache                Y Y 3 . 7  8   78
002 1 Vascular disorders          Hypertension            Y N 2 . 3  4   78
007 2 Infections and infestations Bronchitis              Y N 1 . 4  16  92
012 2 Infections and infestations Viral infection         Y N 1 . 2  9   45
020 2 Eye disorders               Conjunctivitis          N N 1 . 17 21  83
021 2 Infections and infestations Subcutaneous abscess    N N 2 . 45 .   92
021 2 Gastrointestinal disorders  Constipation            N N 2 . 23 24  92
021 2 Renal and urinary disorders Renal failure acute     N N 2 . 43 50  92
021 2 Infections and infestations Sepsis                  Y Y 4 . 56 .   92
031 1 Gastrointestinal disorders  Hypoaesthesia oral      Y N 3 . 3  10  68
032 1 Infections and infestations Tooth abscess           N N 2 . 2  7   78
038 1 Hepatobiliary disorders     Bile duct stone         N Y 2 . 33 45  90
040 2 Nervous system disorders    Headache                N N 1 . 12 13  65
041 2 Nervous system disorders    Dizziness               Y N 1 . 11 12  65
052 1 Infections and infestations Sinusitis               N N 1 . 21 34  96
053 1 Infections and infestations Bronchitis              N N 2 . 4  5   55
053 1 Infections and infestations Sinusitis               N N 2 . 34 .   55
```

```
064 2 Nervous system disorders    Syncope                 N  N  2  . 23  .   87
064 2 Cardiac disorders           Myocardial infarction   N  Y  2  . 45  47  87
065 1 Gastrointestinal disorders  Paraesthesia oral       Y  N  1  1 2   6   20
071 2 Infections and infestations Sinusitis               N  N  2  . 3   16  87
077 1 Nervous system disorders    Headache                N  N  2  . 56  67  83
083 2 Infections and infestations Nasopharyngitis         N  N  1  . 32  .   89
085 1 Cardiac disorders           Atrial fibrillation     N  N  2  . 23  24  65
086 2 Infections and infestations Appendicitis            N  N  2  . 20  27  77
087 1 Eye disorders               Ocular hyperaemia       N  N  2  . 22  56  90
087 1 Eye disorders               Eye swelling            N  N  2  . 12  34  90
088 1 Eye disorders               Eye swelling            Y  N  1  . 56  57  94
091 1 Infections and infestations Bronchitis              N  N  1  . 4   10  89
091 1 Cardiac disorders           Atrial fibrillation     N  Y  3  . 28  30  89
092 1 Infections and infestations Urinary tract infection N  N  1  . 23  26  71
093 1 Infections and infestations Oral herpes             N  N  2  . 12  17  99
093 1 Nervous system disorders    Headache                N  N  1  . 3   5   99
095 2 Infections and infestations Laryngitis              N  N  1  . 66  75  89
095 2 Nervous system disorders    Headache                N  N  1  . 88  89  89
097 1 Infections and infestations Nasopharyngitis         N  N  1  . 45  54  88
097 1 Nervous system disorders    Syncope                 N  N  2  . 23  56  88
099 1 Infections and infestations Diverticulitis          N  N  2  . 4   9   59
099 1 Nervous system disorders    Syncope                 N  N  2  . 21  .   59
103 2 Gastrointestinal disorders  Abdominal pain upper    Y  N  1  . 32  33  78
105 1 Infections and infestations Sinusitis               N  N  2  . 66  .   89
105 1 Infections and infestations Nasopharyngitis         N  N  1  . 45  65  89
;
run;
```

<u>Data Derived :AE_CONTTBL</u>

ORD	AECAT	TRT	OUTCOME	FREQ
1	Any TEAE	1	1	17
1	Any TEAE	1	2	23
1	Any TEAE	2	1	13
1	Any TEAE	2	2	52
2	Drug-Related TEAE	1	1	4
2	Drug-Related TEAE	1	2	36
2	Drug-Related TEAE	2	1	6
2	Drug-Related TEAE	2	2	59
3	Grade 3 or Higher TEAE	1	1	3
3	Grade 3 or Higher TEAE	1	2	37
3	Grade 3 or Higher TEAE	2	1	1
3	Grade 3 or Higher TEAE	2	2	64
4	Grade 3 or Higher Drug-Related TEAE	1	1	2
4	Grade 3 or Higher Drug-Related TEAE	1	2	38
4	Grade 3 or Higher Drug-Related TEAE	2	1	1
4	Grade 3 or Higher Drug-Related TEAE	2	2	64
5	Serious Adverse Event	1	1	3
5	Serious Adverse Event	1	2	37
5	Serious Adverse Event	2	1	2
5	Serious Adverse Event	2	2	63
6	On-study Death	1	1	1
6	On-study Death	1	2	39
6	On-study Death	2	1	0
6	On-study Death	2	2	65

Example 2:

```
data aecount;
label dose    ='Dose Levels'
      outcome ="Response, Y=Yes, N=No"
      count   ='count for Any TEAE'
```

```
             total   ='Total number at a Dose level'
                 ;
input dose outcome $ count total;
datalines;
0   Y     7   97
0   N    90   97
1   Y     6   95
1   N    89   95
2   Y    13  104
2   N    91  104
3   Y    19  100
3   N    86  100
;
run;
```

Example 5: First 12 subjects in the sequence of AB and another 12 subjects in the sequence of BA.

```
data cross;
label subjid  = 'Subject Number'
      sequence = 'Sequence, 1=AB, 2=BA'
      period  = 'Period'
      trt     = 'Treatment, A=Test, B=Reference'
      aebodsys = 'Body System of Event'
      aedecod  = 'Preferred Term'
      ;

input subjid $ 1-3 sequence 5 period 7 trt $ 9 aebodsys $ 11-50 aedecod $ 52-80;
cards;
001 1 1 A Gastrointestinal disorders                 Nausea
001 1 2 B Gastrointestinal disorders                 Nausea
003 1 2 B General disorders                          Chills
004 1 1 A Nervous system disorders                   Headache
004 1 1 A Gastrointestinal disorders                 Nausea
007 1 1 A Gastrointestinal disorders                 Vomiting
009 1 1 A Gastrointestinal disorders                 Vomiting
010 1 2 B Skin and subcutaneous tissue disorders     Skin irritation
011 1 1 A Gastrointestinal disorders                 Vomiting
011 1 2 B Gastrointestinal disorders                 Paraesthesia oral
012 1 1 A Renal and urinary disorders                Renal failure acute
012 1 2 B Renal and urinary disorders                Renal failure acute
013 2 1 B Gastrointestinal disorders                 Nausea
015 2 2 A Renal and urinary disorders                Renal failure acute
017 2 1 B Infections and infestations               Sinusitis
017 2 2 A Infections and infestations               Sinusitis
018 2 1 B Gastrointestinal disorders                 Abdominal pain upper
020 2 1 B Gastrointestinal disorders                 Vomiting
021 2 1 B Gastrointestinal disorders                 Vomiting
024 2 1 B Gastrointestinal disorders                 Nausea
024 2 2 A Gastrointestinal disorders                 Vomiting
;
run;
```

REFERENCES

SAS/STAT® 9.2 User's Guide, The FREQ, PHREG and GLIMMIX Procedures.

O'Neill. R. T. (1995). Statistical concepts in the planning and evaluation of drug safety from clinical trials in drug development: Issues on international harmonisation. Stat. Med. 14:1117–1127.

Chuang-Stein, C. (1998). Safety analysis in controlled clinical trials. Drug Information J.32:1363S–1372S.

Liu, G. F., Wang, J., Liu, K. and Snavely, D. B. (2006). Confidence Intervals for an Exposure Adjusted Incidence Rate Difference with Applications to Clinical Trials. Statistics in Medicine. 25:1275-1286

Siddiqui, O. (2009). Statistical Methods to Analyze Adverse Events Data of Randomized Clinical Trials. Journal of Biopharmaceutical Statistics, 19:889-899.

Johnston, G. and So, Y. (2003). Analysis of Data from Recurrent Events, SUGI 28, Statistics and Data Analysis

Liddell, D. (1976). Practical Tests of 2x2 Contingency Tables. *The Statistician*, 25 (4): 295-304

D'Agostino, R. B., Chase, W., and Belanger, A. (1988). The Appropriateness of Some Common Procedures for Testing Equality of Two Independent Binomial Proportions. *The American Statistician* 42 (3): 198–202.

Cochran, W. G. (1954) Some methods of strengthening the common χ^2 tests. *Biometrics* 10: 417-451.

Armitage, P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics* 11: 375-386.

Liu, H. (2007) Cochran-Armitage Trend Test Using SAS, PharmaSUG 2007, SP05.

Jones, B. and Kenward, M. G. (2003). Design and Analysis of Cross-Over Trials, Second Edition. Chapman & Hall/CRC

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
      Name: Xiaomin He, Ph.D.
      Enterprise: ICON Clinical Research
      Address: 1700 Pennbrook Parkway
      City, State ZIP: North Wales, PA 19454
      Work Phone: 215-616-6406
      Fax: 215-616-8685
      E-mail: Xiaomin.he@iconplc.com
      Web: www.iconplc.com