Paper AS11

Analysis of composite endpoints using the Win-Ratio method in SAS

Murat Ipek, Parexel International, Berlin, Germany

ABSTRACT

In clinical trials the evaluation of composite endpoints is usually defined as the time from randomization to the very first event. This method is missing the clinical importance of endpoints while often the first outcome is of less importance from a medical point of view. To overcome this problem, Pocock et al. published the concept of the Win-Ratio for the analysis of composite endpoints. Subjects from two different treatment groups are divided into pairs based on a risk profile and winners and losers are determined within the pairs. This method underlines the clinical context of two or more endpoints and supports the importance of a certain endpoint, e.g. cardiovascular death as the top priority, and by sequential ordering of the remaining endpoints, e.g. myocardial infarction or stroke.

INTRODUCTION

Cardiovascular trials typically examine efficacy of treatment by using composite endpoints, e.g. cardiovascular death (CV Death), myocardial infarction or stroke. The analysis focuses on time to the very first event – regardless of the clinical importance – since randomization. In most trials, data is analyzed by the Cox-Model, log-rank test, and/or Kaplan-Meier plots to calculate appropriate statistical measurements, like hazard ratios, confidence intervals, p-values and graphical presentations.

This approach has limitations. All endpoints are handled with the same priority or importance since only the first event is considered. It ignores whether a patient dies as a result of a heart attack. Previous events, which are usually non-fatal, are given (indirectly) greater or lesser weight than, for example, severe, fatal events. It should also be remembered that non-fatal events can occur multiple times, whereas death, as the final event, can only be observed once.

To address the importance of individual endpoints and the importance of the clinical context and clinical priority of two or more endpoints, Pocock et al. [1] published in 2011 the concept of the win ratio for analyzing composite endpoints. Subjects from two different treatment groups are divided into pairs based on a risk profile, and within the pairs the winners and losers are determined in relation to the 'new' treatment. By determining clinical priorities, fatal events can thus be weighted more heavily and reflecting the event to be avoided more “realistic”.

This paper shows you the authors' thoughts behind the new, yet very simple and descriptive methodology, and how you can utilize the win ratio method in SAS®. The present work is closely based on the publication by Pocock et al. [1]. To note, in the published article of Pocock et al. [1] there is also a possibility of the calculation of the win ratio without be formation of pairs. In this paper, only the methodology “*The matched pairs approach*” by the addition of a risk profile will be discussed.

THE IDEA / THE MOTIVATION

For the calculation of the win ratio a total of four steps are required. The first step is the definition of the composite endpoint, including the definition of a risk profile. In a clinical trial setting those are pre-defined in the clinical trial protocol and will be considered as “defined” in the following. The remaining three steps are the creation of pairs based on a risk profile, the comparison of the endpoints within the pairs and the calculation of the win ratio. The individual steps are explained in more detail below.

STEP 1: SELECTION OF ENDPOINTS AND RISK PROFILE VARIABLES

This is usually defined in the clinical study protocol.

STEP 2: CREATION OF PAIRS BASED ON A RISK PROFILE

In the second step, subjects from the test treatment group and subjects from the standard treatment group (or control group) are combined into pairs based on the pre-identified risk profile(s). Risk profiles may include the age of the subject, whether any comorbidities, or disease stages such as the New York Health Association Classification. Thus, a pair generally consists or two subjects from different treatment groups who are "equal" in terms of a clinical profile.

In the following example the characteristics of diabetes mellitus, hypertension and the age of subject were used for pairing. This approach ensures that "similar" subjects form a pair from a medical point of view.

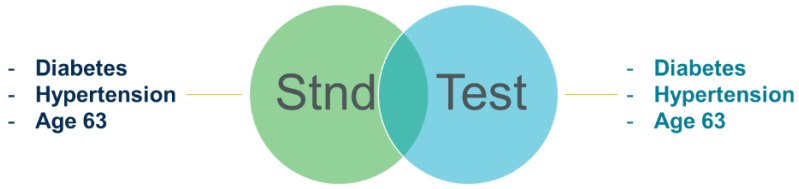


Figure 1: Example of a pair based on risk factors for diabetes, hypertension and age

One requirement for the formation of pairs is to have equal sized treatment groups. In a clinical trial setting this is usually given by the randomization, so that the difference in size of treatment groups would be small. However, before the formation of pairs, the size of the treatment groups must be identical for the given risk profile variables. To achieve a balance, the authors suggest excluding subjects randomly from the treatment with the higher number of subjects.

STEP 3: COMPARISON OF THE ENDPOINTS WITHIN THE PAIRS

In the third step, the comparison of the subjects within each pair is performed. The comparison starts with the most severe event. In clinical trials of cardiovascular diseases, the prevention of the fatal event has the top priority. Taken this as an example, the comparison within each pair would start by looking into the fatal outcome of the subjects. If both subjects have died within a pair, the day of death will be key to assess "winner" or "loser" from point of the new treatment. If they died on different study days, the comparison is straight forward by looking at the subject who died earlier. If only one of the subjects died, a comparison is only possible if the observation period of the subject alive is at least as long as the date of death of the subject died. Otherwise the subjects will not be comparable within the pair for the event "death", and the second endpoint, must be considered. In a cardiovascular trial the second endpoint could be the presence of a myocardial infarction. The comparison as done for death will now be repeated. If both subjects recorded a myocardial infarction, the date of the event will be compared to assess "winner" or "loser", again from point of the new treatment. If only one of the subjects recorded the event, the observation period will be checked first as done for the fatal outcome. If subjects are eligible for a comparison, the assessment of "winner" or "loser" will be performed. If a comparison within a pair is not possible for all endpoints, the pair will be considered as "not comparable".

Once all pairs are evaluated for all endpoints, the pairs are then assigned to one of the following categories, where a) and b) are outcomes of the first endpoint, and c) and d) are outcomes of the second endpoint:

a) Subject under New Treatment is winner; e.g. has a positive outcome for the first event
b) Subject under Standard Treatment is winner; e.g. has a positive outcome for the first event
c) Subject under New Treatment is winner; e.g. has a positive outcome for the second event
d) Subject under Standard Treatment is winner; e.g. has a positive outcome for the second event
e) None of the above

Using the cardiovascular trial above, the pairs would be assigned into one of the following categories:

a) Subject under New Treatment died first
b) Subject under Standard Treatment died first
c) Subject under New Treatment had myocardial infarction first
d) Subject under Standard Treatment had myocardial infarction first
e) None of the above

The interesting part of this method is the clinical priority which is seen by the above categories. Hence, the categories a) and b) consider the event "death" and have a higher priority over categories c) and d). The latter apply only if it is

not known which of the two subjects within a pair died first (no assignment into categories (a) and (b) possible). A pair is counted in the category e) if it is not counted in a) – d).

In the first example, treatment A is winner to treatment B in terms of the death endpoint. Within the pair, the event with the highest priority (death) is always first observed in the patient under treatment B. In the second pair, and because the endpoint with highest priority is evaluable, the second endpoint (hospitalization) is ignored.

In the second example, the event “death” is not observed in any patient, except in the second pair, which is not considered because the observation time of the subject under treatment B is too short compared to the time until death for the other subject of the pair. Thus, the endpoint 'hospitalization' is compared, which is always observed first in the subject under treatment B. Within the third pair the endpoint “hospitalization” can be evaluated as the observation period of the subject under treatment B is at least as long as the one for the subject under treatment A.

The third example describes cases where an assignment is not possible. In both pairs, the observation time is too short to evaluate the endpoint “death” (first pair) or the endpoint “hospitalization” (second pair), so that no treatment is “winner” or “loser” to both endpoints.

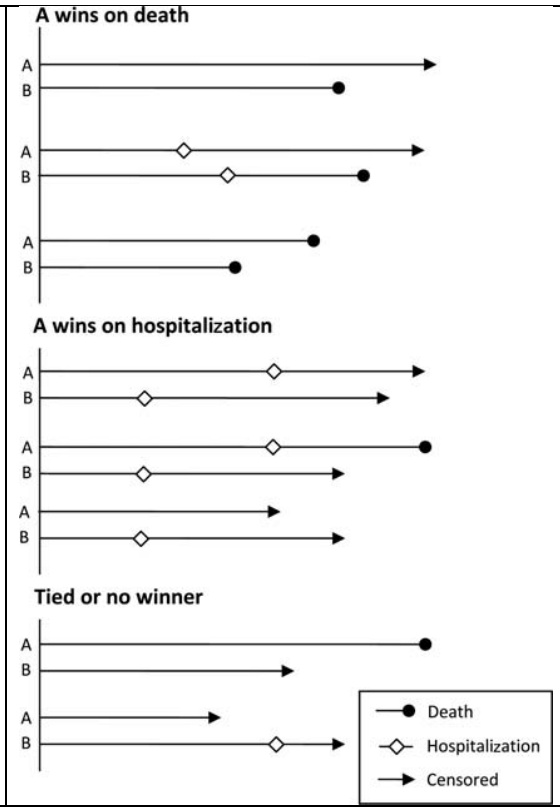


Figure 2: Various scenarios in which pairs can be assigned to categories (a) to (e) with respect to treatments A and B. Events considered are "death" and "hospitalization"; example is copied over from [1]

STEP 4: CALCULATION OF THE WIN RATIO

In the final step, the assignments of the pairs in the categories a) to e) are evaluated. Let N_a, N_b, N_c, N_d, and N_e be the number of evaluated pairs in terms of categories a) to e). Let $N_b + N_d = N_W$ describe the number of 'winners' from the viewpoint of the new treatment, whereas $N_a + N_c = N_L$ describe the number of losers. The ratio between 'winner' and 'loser' defines the win ratio as below:

$$R_W = \frac{N_W}{N_L} = \frac{p_W}{1 - p_W}$$

with the proportion of “winner” given as:

$$p_W = \frac{N_W}{N_W + N_L}$$

and 95% confidence intervals:

$$[p_L; p_U] = p_W \pm 1.96 \left[\frac{p_W(1 - p_W)}{N_W + N_L}\right]^{1/2}$$

The confidence interval for the win ratio can be calculated as:

$$\left[\frac{p_L}{1-p_L};\frac{p_U}{1-p_U}\right]$$

IMPLEMENTATION IN SAS

In this section the implementation of the win ratio will be shown using SAS code. As support, consider the following fictious setting of a clinical trial and the selected risk factors and endpoints:

Risk factors for pairing subjects:

- Multi-Drug Resistance (MDR) status of a subject. This variable is given as a binary variable
- Acute Physiology and Chronic Health Evaluation Score, in short APACHE-II Score (APC). This variable is a discrete variable ranging from 0 to 71

Endpoints for identification of winner or loser:

- Death, defined as the time from randomization to death
- Ventilator-free days which is the number of days without artificial respiration (VF)

BUILD PAIRS

Before building pairs, the treatment group size needs to be checked. In case of unequal size of treatment groups, it is important to identify the risk variable(s) which may cause an imbalance. In the example above, APC is defined in the range from 0 to 71 and difficult to use for pairing. To overcome the issue of a discrete pairing, ranks were calculated and used instead. This means that an imbalance between the two treatment groups can only be caused by the variable MDR. It should be noted here that the challenge of pairing is not in merging records. Rather, it is important to first identify a possibly existing imbalance and then eliminate it by selecting random subjects through the right risk profile or variable.

For getting the same number of subjects in both treatment groups, the idea was not to exclude subjects randomly, but to determine the minimum number of subjects in both groups. After identification of the minimum sample size, subjects were pulled randomly from both treatment groups without replacement. With the procedure FREQ, the number of subjects with MDR="Yes" and MDR="No" were counted, in order to determine the minimum number of subjects within each characteristic. These were filed as macro variables (`'keepYes'`, `'keepNo'`). An example SAS code is given below:

```
PROC FREQ DATA=wr15 NOPRINT;
  TABLE trt01p * mdr / OUT=wr35freq;
RUN;

PROC SORT DATA=wr35freq;
  BY mdr;
RUN;

PROC TRANSPOSE DATA=wr35freq OUT=wr36trans;
  BY mdr;
  VAR count;
  ID trt01p;
RUN;

DATA wr40num;
  SET wr36trans;
  keep  =MIN(standard,test);
  IF mdr="Yes" THEN CALL SYMPUT('keepYes',STRIP(PUT(keep,BEST.)));
  IF mdr="No"  THEN CALL SYMPUT('keepNo',STRIP(PUT(keep,BEST.)));
RUN;
```

Subsequently, the minimum number of subjects was used to randomly select the required sample. Selection was done using the procedure SURVEYSELECT as shown below:

```
PROC SURVEYSELECT DATA=wr15 (WHERE=(mdr="Yes")) SEED=953
    METHOD=SRS SAMPSIZE=&keepYes. OUT=wr50yes OUTALL NOPRINT;
  STRATA trt01p;
  ID _all_;
RUN;

PROC SURVEYSELECT DATA=wr15 (WHERE=(mdr="No")) SEED=953
    METHOD=SRS SAMPSIZE=&keepNo. OUT=wr51no OUTALL NOPRINT;
  STRATA trt01p;
  ID _all_;
RUN;
```

The Simple Random Sampling method (METHOD = SRS) was used to perform random drag without replacement. The required sample size was set with the SAMPSIZE option. The STRATA statement was used to consider the distribution of subjects within the two treatment groups. Subjects were thus assigned to the groups by random pulling over the determined minimum number. For example, if the minimum number of subjects in both treatment groups is 20 for the characteristic "Yes", then exactly 20 unique subject IDs per treatment group were drawn. By merging the two resulting data sets `wr50yes` and `wr51no` together, the same number of subjects within the expressions "`Yes`" and "`No`" of the risk variable MDR were retained. The two treatment groups are now balanced. The next step is to calculate the ranks of the risk variable APC. SAS provides the ability to calculate the ranks by using the RANK procedure. However, for reasons of increased control in the formation of subgroup ranks, the ranks were calculated in a DATA step, as illustrated by the following SAS code:

```
DATA &outdataset.;
  SET &indataset.;
  BY mdr;
  RETAIN &rankVarname. rankVar ;
  IF FIRST.mdr THEN DO;
    rankVar=&var2Consider.;
    &rankVarname.=1;
  END;
  ELSE DO;
    IF rankVar NE &var2Consider. THEN &rankVarname.=&rankVarname.+1;
    rankVar=&var2Consider.;
  END;
  DROP rankVar;
RUN;
```

The solution presented assign ranks for the "Standard" and "Test" treatment groups separately, controlled by the by-variable MDR. The ranks for APC have to be unique within each MDR characteristics. It may well happen that no clear ranking for APC is possible, e.g. if more than one patient has an APC of 33. In this case, a third risk variable can be used. Since the ranking can be repeated several times, this was realized within a small macro. Once the ranks for APC are unique to both treatment groups, they are merged horizontally. This can be done using the SQL procedure as shown below:

```
PROC SQL;
  CREATE TABLE wr60merge AS
    SELECT l.usubjid      AS test_usubjid
          ,l.rank3        AS test_rank3
          ,l.mdr          AS test_mdr
          ,l.died         AS test_died
          ,l.dayDied      AS test_dayDied
          ,l.dayVF        AS test_dayVF
```

```
          ,r.usubjid      AS stnd_usubjid
          ,r.rank3        AS stnd_rank3
          ,r.mdr          AS stnd_mdr
          ,r.died         AS stnd_died
          ,r.dayDied      AS stnd_dayDied
          ,r.dayVF        AS stnd_dayVF
      FROM           test05rank03     AS l
        FULL JOIN    stnd05rank03     AS r
          ON l.mdr        = r.mdr
         AND l.rank3      = r.rank3
  ;
QUIT;
```

IDENTIFICATION OF WINNER AND LOSER

At this stage, the pairing is complete. Now, for each pair, the endpoints death and VF need to be compared, starting with the top priority event first, which is the fatal outcome. The goal is to assign each pair to one of the categories (a) to (e):

a) Subject under New Treatment died first
b) Subject under Standard Treatment died first
c) Subject under New Treatment has less VF days
d) Subject under Standard Treatment has less VF days
e) None of the above

The implementation in SAS is quick and easy. With the RETAIN statement, the winners are cumulated per treatment group. The following SAS code is a sample from the SAS program only. The procedure in case of further endpoints (here VF) is analogous and will not be discussed here.

```
RETAIN Na 0 Nb 0 Nc 0 Nd 0 Ne 0;
* Only one died;
IF test_died NE stnd_died AND comparable THEN DO;
  IF      test_died THEN DO; Na=Na+1; winner="STND"; END;
  ELSE IF stnd_died THEN DO; Nb=Nb+1; winner="TEST"; END;
END;
* Both died;
ELSE IF test_died AND stnd_died AND test_dayDied NE stnd_dayDied THEN DO;
  IF test_dayDied < stnd_dayDied THEN DO;
    Na=Na+1;
    winner="STND";
  END;
  ELSE IF test_dayDied > stnd_dayDied THEN DO;
    Nb=Nb+1;
    winner="TEST";
  END;
END;
...
```

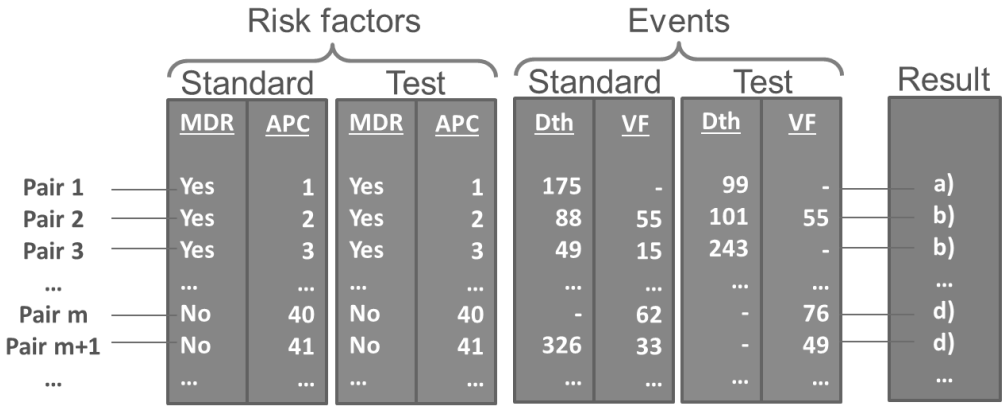
At the end, the number of pairs assigned to the respective categories a) to e) is of interest. For this purpose, RETAIN variables were initially created in order to count the pairs with each iteration for the defined variables N_a, N_b, N_c, N_d and N_e. The first condition above describes the pairs where only one subject died. An indicator variable is used, in which the necessary condition of the "same" observation period is recorded (`comparable`). The second constraint above describes the pairs where both subjects died, but on different study days. A comparison for the outcome "death" is only possible if subjects died at different days.

CALCULATION OF THE WIN RATIO

All pairs based on the endpoints are evaluated. The last step is to calculate the win ratio as given below:

```
/* Nw is the number of 'winners' for treatment with Test, i.e. those matched pairs
where STND fared worse */
Nw=Nb+Nc; /*Test died after Standard; Test shows higher VF*/
Nl=Na+Nd; /*Test died before Standard; Standard shows higher VF*/
/*--- CALCULATION OF SUBJECTS EXCLUDED ---*/
Nexclud=STRIP(PUT(&_overall.-2*(&keepNo.+&keepYes.),BEST.));
/*--- CALCULATION OF PROPORTION FOR WINNERS ---*/
pw=Nw/(Nw+Nl);
pl=pw-1.96*SQRT( (pw*(1-pw) )/(Nw+Nl) );
pu=pw+1.96*SQRT( (pw*(1-pw) )/(Nw+Nl) );
/*--- CALCULATION OF WIN RATIO AND THE 95% CONFIDENCE INTERVAL ---*/
rw_=nw/nl;
rw=pw/(1-pw);
rl=pl/(1-pl);
ru=pu/(1-pu);
/*--- CALCULATION OF Z-VALUE AND THE P-VALUE ---*/
z=abs(pw-0.5)/( SQRT( (pw*(1-pw) )/(Nw+Nl) ) );
p=(1-PROBNORM(z))*2;
```

Please refer to the Figure 3 below for visual illustration. From left to right are shown: the characteristics for the formation of the pairs, the endpoints death and VF with examples, as well as the evaluation into the categories a) to e).



The first two columns on the left describe the risk factors for forming the pairs. In the centre, the endpoints are displayed, with the days from randomization to death for the first endpoint and the number of days without mechanical ventilation for the second endpoint. On the right, the assignment of the pairs is given in categories a) to e).

CONCLUSION

The method is simple and understandable. In contrast to the classic evaluation of composite endpoints, subject risk profiles are considered. The profile, based on the examined indication, can take into account the medical history of the subject as well as medically relevant test results. In addition, multiple endpoints can be defined and subdivided into priority lists according to their medical relevance.

As illustrated by the example, the method is not limited to cardiovascular studies. It also shows that the methodology is accepted and used in the statistical world showed by a publication published in March 2015 to calculate confidence intervals [2].

Imbalance between treatment groups can be an issue, if the total number of subjects is small, e.g. if the characteristic for determining the risk profile has not been adequately selected and more subjects must be excluded in order to achieve balance between the treatment groups. Certainly, the question of an additional sensitivity analysis is justified, in which it must be shown that the exclusion of other randomly selected subjects leads to the same results. Overall, the methodology becomes difficult if there are no risk characteristics for pairing. This variant was not covered in this article, but in such a case every possible combination of subjects must be considered. Please refer to the publication [1].

The following is a note from the author to encourage further discussion on using the method:

- By exclusion of subjects from the analysis (in case of imbalance), the method conflicts with the intention-to-treat principle, which expects the evaluation of all randomized subjects. Sensitivity analyzes are required to demonstrate that the same results are observed when excluding other subjects.
- Stratification factors should be used in the selection of the risk profile in order to minimize imbalance in the treatment groups. Risk profiles which are observed uncontrolled can lead to a greater imbalance and thus to an increased exclusion of subjects.
- Another way to avoid the exclusion of subjects could be a 1:n "matching" for subjects for whom no counterpart subject exists due to the different group size. In this case, subject i from the smaller group would be used several times for pairing, so that each subject from the larger group can be used for pairing. Attention should be made when calculating statistical inferences as number of winner and losers will increase but with the same number of subjects. Correction of the outcome will be required.
- It is recommended to include the method as "supportive analysis" in the analysis plan. Thoughts on using the method as the primary method of analysis should be well discussed and tested with existing data or analyzed based on simulations.

REFERENCES

[1] Stuart J. Pocock, Cono A. Ariti, Timothy J. Collier, and DuolaoWang; “The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities”; European Heart Journal 33 (2012), 176–182; doi:10.1093/eurheartj/ehr352.

[2] Xiaodong Luo, Hong Tian, Surya Mohanty, and Wie Yann Tsai; „An Alternative Approach to Confidence Interval Estimation for the Win Ratio Statistic“; Biometrics (March 2015) 71, 139-145; doi:10.1111/biom.12225

ACKNOWLEDGMENTS

I would like to thank my colleague and friend Rui Pedro Raminho Simoes for his review and suggested edits.

CONTACT INFORMATION

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

Murat Ipek
Parexel International
Spandauer Damm 130
Klinikum Westend, Haus 18
Berlin, Germany
murat.ipek@parexel.com
www.parexel.com