

A Multivariate Ranking Procedure to Assess Treatment Effects

Alan Y Chiang, Eli Lilly and Company, Indianapolis, IN
 Grace Li, Eli Lilly and Company, Indianapolis, IN
 Ying Ding, University of Michigan, Ann Arbor, MI
 Ming-Dauh Wang, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

In early phase clinical studies, it is often difficult to assess the effects of a set of biomarker variables and/or clinical endpoints when the individual variables do not appear to have statistically significant effects. To address this situation, we propose a method of U-scores applied to subsets of multivariate data. A SAS[®] macro is developed and we illustrate the usefulness of this approach through simulations, considering various combinations of correlations and underlying distributions, and compare its statistical power with the existing tests: Hotelling's T^2 and nonparametric rank-sum tests. Finally we apply this approach in a Phase I clinical study to help assess the treatment effects of an investigational drug on rheumatoid arthritis.

INTRODUCTION

In early phase clinical studies, treatment comparisons in clinical studies can typically involve several biomarkers and/or distinct endpoints. Parametric approaches to the analysis of multiple variables have been proposed, notably by O'Brien (1984) and Pocock, Geller, and Tsiatis (1987). See also Wassmer et al. (1999) for further discussion on the procedures for testing multiple endpoints. In contrast, as had been noted by Lehmacher et al. (1991) and Koziol (1991), multivariate rank statistics can be of particular value in the analysis of multiple endpoint data, particularly when the variables are heterogeneous or distributional assumptions are suspect.

The majority of these methods are based on the linear model, either explicitly, as in regression, factor and cluster analysis, or implicitly, as in neural networks (Wittkowski et al., 2004). One scores each variable individually on a comparable scale, either present/absent, low/intermediate/high, 1 to 10, or z-transformation, and then defines a global score as a weighted average of these scores. In other words, data are interpreted as points in a Euclidian space of (independent) dimensions. The number of dimensions is reduced by assuming the dimensions to be related by a specific function of known type, allowing one to determine for each point the Euclidian distance from a hyperspace. While mathematically elegant and computationally efficient, this approach has shortcomings when applied to real world data. Since the relative importance of the variables, the correlations among them, and the functional relationship of each variable with the immeasurable latent factor (e.g. efficacy, safety, risk, or overall usefulness) are typically unknown, the linear model validity may not be established.

As an alternative to the commonly used Hotelling's T^2 and multivariate rank statistics approaches, we have investigated a method of U-scores approach (Wittkowski et al., 2004) to the multiple biomarker/endpoint problems. We demonstrate how this method leads to a simple non-parametric statistical test for comparing treatments with respect to several ordinal outcomes, some of which may be graded. In addition, we apply U-scores to help assess the treatment effects and dose-response of an investigative drug on rheumatoid arthritis. The scoring mechanism is introduced in the next section.

U-SCORES FOR MULTIPLE VARIABLES

As described in Wittkowski et al. (2004), the scoring mechanism is based on the principle that each subject i

$$\vec{y}_i = (y_{i1}, \dots, y_{ip})', \quad i = 1, \dots, n$$

is compared with every other subject in a pairwise manner. When the observed outcomes can be assumed to be correlated with an unobservable latent factor, a partial ordering among the subjects can easily be defined. If the second of two subjects has values at least as high among all variables $k = 1, \dots, p$, but higher in at least one variable, it will be called 'superior':

$$\vec{y}_i > \vec{y}_j \Leftrightarrow \{ \forall_{k=1, \dots, p}, y_{ik} \geq y_{jk} \wedge \exists_{k=1, \dots, p}, y_{ik} > y_{jk} \}.$$

We also define:

$$\bar{y}_i \leq \bar{y}_j \Leftrightarrow \{\forall_{k=1,\dots,p}, y_{ik} \leq y_{jk}\}.$$

If identical observations (ties) are present, two cases need to be considered. Ties may be due to the underlying phenomena, and are often caused by discretization or by observing a discrete surrogate variable for a continuous phenomenon (e.g. interval censoring). Even though a partial ordering does not guarantee that all patients can be ordered on a pairwise basis, all subjects can be scored. The vectors \bar{y}_i and \bar{y}_j are compared through a kernel function t :

$$t(\bar{y}_i > \bar{y}_j) = \begin{cases} 1, & \text{if } \bar{y}_i > \bar{y}_j \\ 0, & \text{if } \bar{y}_i \leq \bar{y}_j \end{cases}$$

Otherwise, we define $t(\bar{y}_i > \bar{y}_j) = 0$ if \bar{y}_i and \bar{y}_j can not be ordered. One can then assign a score $u(\bar{y}_i)$ to each subject by simply counting the number of subjects being inferior and subtracting the number of subjects being superior:

$$u(\bar{y}_i) = \sum_{j=1, j \neq i} \{t(\bar{y}_i > \bar{y}_j) - t(\bar{y}_j > \bar{y}_i)\}.$$

To illustrate, Table 1 shows a subset of Phase 1 clinical trial data from an investigational drug on rheumatoid arthritis. C-reactive protein (CRP), erythrocyte sedimentation rate (ESR) and matrix metalloproteinase (MMP3) are three inflammatory biomarkers of interest and the measurements were changes from baseline after a single dose of treatment. Ten subjects are listed. Comparing S1 and S2, we find S2 has a greater value of CRP than S1, while S1 has a greater ESR value than S2. Therefore S1 and S2 can not be ordered and each receives a score of 0. Comparing S1 and S4, we find S1 has greater values in all three biomarkers than S4. Therefore S1 has a score of 1 in comparison of S4, and S4 has a score of -1 in comparison with S1. By repeating this procedure, we can summand the total score for each subject and this is recorded in last column of the table. The SAS macro for calculating $u(\bar{y}_i)$ for each subject is included in the Appendix.

Table 1: Subset of a Phase 1 study

ID	CRP	ESR	MMP3	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	u(y)
S1	15	57	12.9	-	0	0	1	1	1	1	1	1	1	7
S2	48.1	30	22.9	0	-	0	0	1	1	1	1	1	1	6
S3	24.2	38	12.2	0	0	-	0	1	1	1	1	1	1	6
S4	11.91	40	-1.3	-1	0	0	-	0	0	1	0	1	1	2
S5	7.09	18	8	-1	-1	-1	0	-	0	0	1	1	1	0
S6	13.13	1	-6.8	-1	-1	-1	0	0	-	0	0	0	1	-2
S7	-2.04	25	-5.2	-1	-1	-1	-1	0	0	-	0	0	1	-3
S8	-3.98	-10	2.2	-1	-1	-1	0	-1	0	0	-	0	0	-4
S9	-0.1	14	-1.9	-1	-1	-1	-1	-1	0	0	0	-	1	-4
S10	-15.5	0	-13.8	-1	-1	-1	-1	-1	-1	-1	0	-1	-	-8

TEST STATISTICS AND POWER

In this section, we describe the multivariate U -statistics for comparison of two groups, which we call it multivariate U -test. Given the observations from two groups $\{\bar{y}_{1i}\}_{i=1,\dots,n_1}$ and $\{\bar{y}_{2j}\}_{j=1,\dots,n_2}$, define U -statistics as follows:

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} t(\bar{y}_{1i} > \bar{y}_{2j}).$$

It can be seen that U is the total number of pairwise comparisons that is in favor of Group 1. Asymptotic properties of the U -statistics can be found in Hoeffding (1948) and Wittkowski (2004). For small sample size trials, we can permute the data to obtain the empirical distribution of U under the null hypothesis that there is no difference between two groups.

We examine the performance of multivariate U -test against two commonly used multivariate tests, namely, Hotelling's T^2 and multivariate rank-sum tests under three scenarios:

SCENARIO 1:

We test four possible correlation structures between variables. We assume there are 4 variables and $n_1 = n_2 = 10$. The underlying distribution is multivariate log-normal: $\log(\vec{y}_i) \sim MVN(\vec{\mu}_i, \Sigma)$, where $\vec{\mu}_1 = (2, 0.5, 3, 5)'$ and $\vec{\mu}_2 = (1, 0, 2, 3)'$. The results are summarized in Figure 1. It is observed that multivariate U -test outperforms the other two tests.

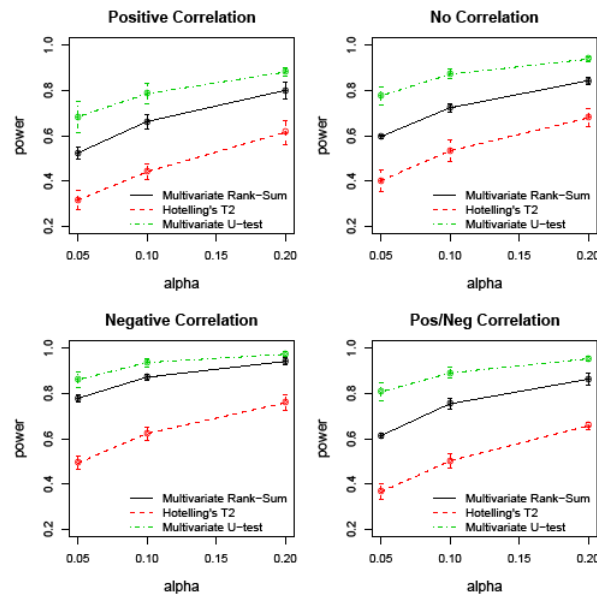


Figure 1. Statistical power comparison for Scenario 1.

SCENARIO 2:

We test three different underlying distributions. We assume there are 4 variables, $n_1 = n_2 = 10$, and $\vec{\mu}_1 = (2, 0.5, 3, 5)'$ and $\vec{\mu}_2 = (1, 0, 2, 3)'$. The results are summarized in Figure 2. It is observed that multivariate U -test outperforms the other two tests even when the data follow a multivariate normal distribution.

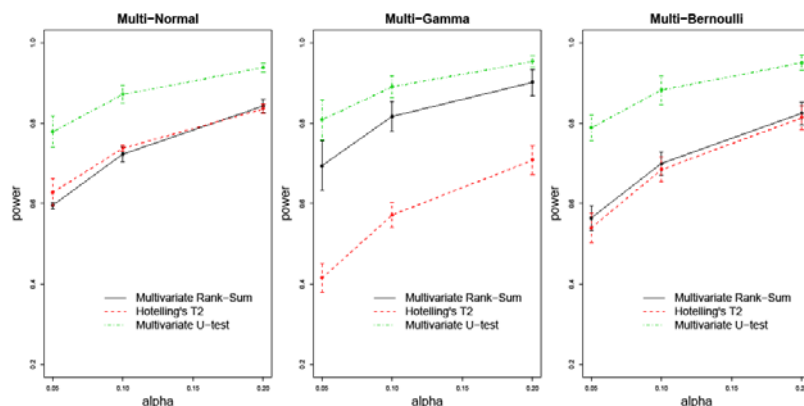


Figure 2. Statistical power comparison for Scenario 2.

SCENARIO 3:

We test three different numbers of variables, i.e. the dimension of the vector is either 4, 6 or 8. We assume $n_1 = n_2 = 10$, and the underlying distribution is multivariate log-normal: $\log(\vec{y}_i) \sim MVN(\vec{\mu}_i, \Sigma)$. The results are summarized in Figure 3. It is observed that multivariate U -test outperforms the other two tests when the number of variables is less than 8. When the number of variables increases, the number of incomparable pairs also

increases, resulting a 'loss' of information. In this situation, the multivariate rank-sum test can be a more suitable choice.

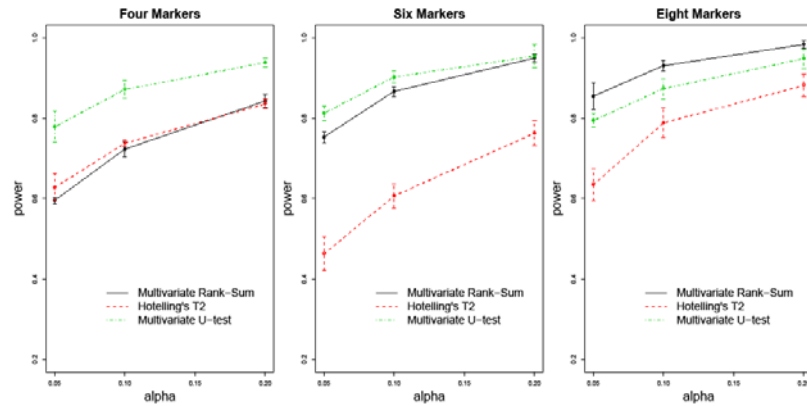


Figure 3. Statistical power comparison for Scenario 3.

DOSE FINDING

We explore the dose response relationships of an investigational drug for rheumatoid arthritis based on the pharmacodynamic effects in three inflammatory biomarkers (CRP, ESR and MMP3) and two clinical endpoints (ACR20 and DAS28). ACR20 is defined as the proportion of subjects who have $\geq 20\%$ improvement in both tender and swollen joint counts. DAS28 consists of a composite score of the following variables: tender joint count, swollen joint count, CRP, and the Patient's Global Assessment of Disease Activity. This is a measurement of disease progress activity over time. The individual dose response plots are shown in Figure 4.

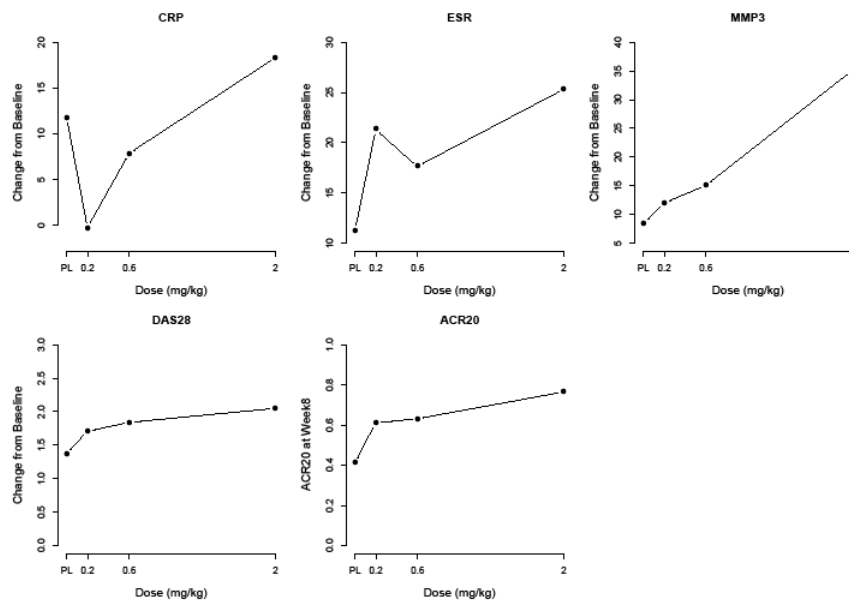


Figure 4. Dose response relationship for biomarker or clinical endpoint individually.

It can be seen that individual dose response is not clear. Important response characteristics for the clinical endpoints such as ED50 or ED80 can not be easily assessed. We apply the scoring procedure as discussed previously to various combinations of biomarkers and clinical endpoints. By plotting the mean score for each group, the dose response relationships appear to be clear and the results are shown in Figure 5.

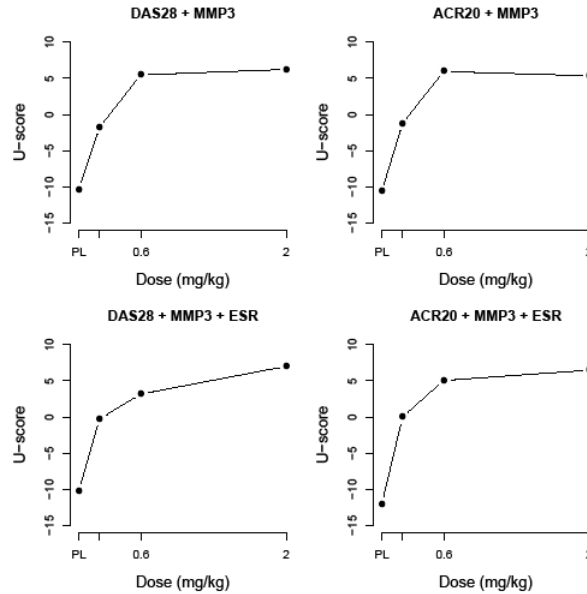


Figure 5. Dose response relationships for combination of biomarker(s) and clinical endpoint(s).

CONCLUSION

We illustrate the usefulness of U-scores through simulations, considering various combinations of correlations and underlying distributions, and compare its statistical power with the existing tests. %U-score macro attached in the Appendix is a user-friendly SAS[®] macro which can calculate the U-score from multiple biomarkers/endpoints for each subject. This can be used for formal statistical comparison or help in dose-finding. Anyone who has minimum SAS[®] programming skills and a basic understanding of the non-parametric statistic can use this macro with ease.

REFERENCES

- O'Brien PC (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, 40, 1079-1087.
- Pocock PC, Geller NL, and Tsiatis AA (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, 43, 487-498.
- Wassmer G, Reitmeir P, Kieser M, and Lehmacher W (1999). Procedures for testing multiple endpoints in clinical trials: An overview. *Journal of Statistical Planning and Inference*, 82, 69-81.
- Lehmacher W, Wassmer G, and Reitmeir P (1991). Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics*, 47, 511-521.
- Kozioł JA (1991). Nonparametric approaches to the analysis of multiple endpoints, *Biom. Inform. Med. Biol.*, 2, 44-47.
- Wittkowski KM, Lee E, Nussbaum R, Chamian FN, and Krueger JG (2004). Combining several ordinal measures in clinical studies. *Statistics in Medicine*, 23, 1579-1592.
- Hoeffding W (1948). A class of statistics with asymptotically Normal distribution. *Annals of Mathematical Statistics*, 19, 293-325.
- Wittkowski KM (2004). Novel methods for multivariate ordinal data applied to Olympic Medals, risk profiles, genomic pathways, genetic diplotypes, pattern similarity, and array normalization. *Computing Science and Statistics*, 35, 626-646.

ACKNOWLEDGMENTS

We are grateful to Dr. Patrick Haslett and Dr. Jin Feng for their valuable comments, and Dr. Soomin Park for her careful reading of this paper.

APPENDIX

```
%macro Uscore(data=, var=, id=, trt=);

proc sql noprint;
    select compress(put(count(unique &id),9.)) into: Nsub from &data;
```

```

quit;

data &data;
    set &data(rename=(&ID=old&ID));
    &id='S' || left(put(_n_,9.));
run;

proc transpose data=&data out=one;
    id &id;
    var &var;
run;

data one1(drop=i j);
    set one;
    array S(&Nsub);
    array bbb(&Nsub,&Nsub);
    do i=1 to &Nsub;
        do j=1 to &Nsub;
            if i^=j then bbb(i,j)=(s(i)>s(j))+(s(i)<s(j))*(-1);
        end;
    end;
run;

ods listing close;

%macro x(stat);
    ods listing close;
    proc means data=one1;
        output out=&stat(drop=_type_ _freq_) &stat.(bbb:)=;
    run;

    proc transpose data=&stat out=&stat(rename=(col1=&stat));
    proc sort data=&stat; by _name_;
    run;
%mend;

%x(min);
%x(max)

data all(keep=_name_ sum);
    merge min max; by _name_;
    if min=. and max=. then sum=.;
    else if min>=0 and max=1 then sum=1;          *bigger than any other subject;
    else if min=-1 and max<=0 then sum=-1;      *smaller than any other subject;
    else sum=0;                                  *could not make a call;
run;

proc transpose data=all out=all(drop=_name_);
    id _name_;
    var sum;
run;

data all(keep=s:);

```

```

set all;
array bbb(&Nsub,&nsub);
array S(&nsub);
do i=1 to &nsub;
    do j=1 to &nsub;
        s(i)+bbb(i,j);
    end;
end;
run;

proc transpose data=all out=all(rename=(col1=U));
run;

ods listing;
proc sql;
    select b.*, U
    from all a, &data b where left(_name_)=left(uppercase(&id)) order by &ID;
    select &trt, mean(U) as sU from (select b.*, U
    from all a, &data b where left(_name_)=left(uppercase(&id))) group by &trt;
drop table all, one1, min, max;quit;
quit;
run;

%mend;

```

CONTACT INFORMATION

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

Alan Y Chiang
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
E-mail: chiangay@lilly.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.