İstanbul Üniversitesi İşletme Fakültesi Dergisi
Istanbul University Journal of the School of Business Administration
Cilt/Vol:40, Sayı/No:1, 2011, 22-34
ISSN: 1303-1732 – www.ifdergisi.org © 2011

Dedicated to the 40th Anniversary of the Journal

A new multiple test approach for nursing care administration of deep vein thrombosis patients

John Chen[1]
Department of Mathematics and Statistics, Faculty of Arts and Sciences
Bowling Green State University, Bowling Green, OH 43403, USA

Mary Eileen Walsh
Department of Vascular Surgery
Jobst Vascular Institute,
The Toledo Hospital, Toledo, OH 43606, USA

Gail Davis
Department of Vascular Surgery
Jobst Vascular Institute,
The Toledo Hospital, Toledo, OH 43606, USA

Anthony Comerota
Department of Vascular Surgery
Jobst Vascular Institute,
The Toledo Hospital, Toledo, OH 43606, USA

Marilyn Gravett
Department of Vascular Surgery
Jobst Vascular Institute,
The Toledo Hospital, Toledo, OH 43606, USA

Dennis Wojnarowski
Department of Vascular Surgery
Jobst Vascular Institute,
The Toledo Hospital, Toledo, OH 43606, USA

Abstract

In nursing care administration, it is critical to evaluate the risk assessment ability of the nurses for patients at different risk status. For an investigation carried out in the Jobst Vascular Institute at The Toledo Hospital, the nurse risk evaluation necessitates the analysis of risk assessment data for prophylaxis of deep venous thrombosis when comparing nurse risk assessment scores with master scores simultaneously at different risk categories. While conventional statistical methods fail to make any conclusion from the data, we construct a new stepwise confidence procedure that strongly controls the family-wise error rate and successfully detects the difference between the nurse score and the Master score. Compared with existing statistical methods, the new bivariate method is more powerful than the Bonferroni procedure and the Holm's step-down algorithm for this data set. It is also more robust than the Hochberg's step-up approach (which relies on an un-checkable assumption of positive dependence among test statistics to strongly control the family-wise error rate). In the data analysis of patients with deep vein thrombosis, the new method successfully detects the difference between the master risk assessment score and the nurse score, while the conventional statistical methods are unable to make any conclusive statement. The new statistical method is applicable to other fields of administration research simultaneously comparing management performances of two different groups under different scenarios.

Keywords: *Simultaneous Confidence Intervals, Multiple Testing, Partition Approach, Nursing-care Administration, Deep Vein Thrombosis, Risk Assessment, Disease Prophylaxis*

[1] *jchen@bgsu.edu (J. Chen), corresponding author*

Derin ven trombozu olan hastalara verilen hastabakıcılık hizmetlerinin yönetimine ilişkin bir çoklu test yaklaşımı

Özet

Hastabakıcılık yönetiminde, hemşirelerin farklı risk düzeylerindeki hastaların taşıdıkları riski değerleme becerileri önem arz etmektedir. Toledo Hastanesi Jobst Damar Enstitüsü'nde yürütülen bir araştırma, farklı risk kategorileri için hemşirelerin ve uzmanların risk değerleme puanlarını eş zamanlı olarak karşılaştırabilmek üzere derin ven trombozunun önlenmesi ile ilgili risk değerleme verilerinin analizini gerektirmiştir. Geleneksel istatistik teknikler verilerden herhangi bir sonuç çıkarmada başarısız olurken, yapılandırdığımız yeni aşamalı güven yöntemi alfa hatasını güçlü bir biçimde kontrol altında tutmakta ve hemşireler ile uzmanların puanları arasındaki farkı başarı ile saptamaktadır. Mevcut istatistik teknikler ile kıyaslandığında, bu yeni iki değişkenli yöntemin, söz konusu veri seti için Bonferroni ve Holm'un basamaklı işleme algoritmasından daha güçlü olduğunu belirtebiliriz. Aynı zamanda Hochberg'in artış yaklaşımından (ki bu yaklaşım aşamalı hata oranını sıkı kontrol altında tutma amacıyla test istatistikleri arasında sınanamaz bir pozitif bağımlılık varsayımına dayanmaktadır.) da daha sağlamdır. Derin ven trombozu olan hastaların verilerinin analizinde, geleneksel istatistik teknikler nihai bir yorum sağlayamazken, bu yeni teknik hemşire ve uzman risk değerleme puanları arasındaki farkı başarıyla tayin etmektedir. Önerilen yeni teknik, eş zamanlı olarak farklı koşullar altında iki farklı grubun yönetim performanslarını kıyaslamaya yönelik yönetim araştırmalarının diğer alanlarına da uyarlanabilir niteliktedir.

Anahtar Sözcükler: *Eş Zamanlı Güven Aralıkları, Çoklu Sınama, Üleştirme Yaklaşımı, Hastabakıcılık Yönetimi, Derin Ven Trombozu, Risk Değerleme, Hastalıktan Korunma*

1. Introduction

Nursing care is an important part in medical and hospital management and administration. Effective evaluation of the ability, knowledge, and experience of the nurses critically affects the function of nurse training programs and the quality of patient care in a medical unit. For patients with deep vein thrombosis (DVT), the accuracy of nurses performing risk assessment directly affects the quality of nursing care and indirectly influences treatment outcomes.

As pointed out by Silverstein et al. [1] and White [2] in population studies, the overall age- and gender-adjusted annual incidence of venous thromboembolism (VTE) is 1 to 2 per 1000 people in the United States (most of the cases are recurrent). Almost one-third of the symptomatic VTE patients have pulmonary embolism (PE), whereas the rest have isolated DVT. In terms of VTE mortality, according to an estimate in 2006 (see, for example, Heit [3]), more than 1,000,000 cases of VTE are diagnosed annually in the United States, with at least 300,000 VTE related deaths annually.

Normally, venous thrombosis occurs when red blood cells and others (such as fibrin, platelets and leukocytes) form a mass within an intact vein. Main manifestations of VTE are PE and DVT. Pulmonary embolism occurs when a piece of thrombus detaches from a vein wall, travels to the lungs, and lodges within the pulmonary arteries. In fact, PE and DVT are highly associated with each other. According to Browse and Thomas [4], the majority of pulmonary emboli originate in the pelvic and deep veins of the lower extremities. Since a fatal PE normally occurs within a very short time without any specific symptoms, prophylaxis of PE naturally includes preventive measures for DVT.

Besides acute PE, complications of DVT include chronic postthrombotic syndrome. Comerota et al. [5] and Kearon et al. [6] investigated the effect of postthrombotic syndrome on the quality of patients' life. As suggested by Galson [7], it is desirable to develop diagnostic instruments and risk assessment methods for symptomatic DVT

patients. This paper is partly motivated by the investigation of risk assessment methods for DVT patients at Jobst Vascular Institute (JVI).

In the literature, there are substantial publications on DVT risks and prophylaxis. For example, Geerts et al. [8] indicate that hospitalization of critically ill patients increases the risk of DVT; Goodacre et al. [9] show that the incidence rate of DVT increases with the age of patients under study, to list just a few. Although many papers have aimed at the evaluation of DVT prophylaxis, few have investigated the reliability of DVT risk assessment methods that are usually performed by nurses.

In the DVT risk assessment, risks are normally sorted into the following three risk categories according to the extent of risk factors involved.

The first category consists of risk factors such as age, obesity, pulmonary disease, etc. All the risk factors in category 1 are listed in Table 1 in the appendix. The second category of DVT risks includes risk factors such as stroke, hip fracture, family history of DVT, etc. All the risk factors in this category are listed in Table 2 in the appendix. The third category of DVT risks contains the history of DVT and acute spinal cord injury.

Preventive measurements and forms of DVT prophylaxis are highly dependent on the risk category that the DVT patient is in. Thus, it is critical to correctly access the risk category for medical and surgical patients. To search for reliable and efficient risk assessment approaches for DVT patients, a research team at JVI collected a data set specific to these patients. The data set consists of DVT risk assessment scores evaluated by nurses (the nurse score) and by a master (an experienced nursing professional), independently.

2. Methods and Setting

The preliminary data set on DVT risk assessment scores includes 108 general medical and surgical patients. The patient DVT risk assessment score is the sum of the indicators of the risk factors (1 for the presence and 0 for the absence of the risk factor) specified in each risk category.

The patient population of the JVI data set consists of 66 females and 42 males with age ranging from 21 to 96. For each patient, the DVT risk status was first evaluated by a nurse who interviewed the patient and filled in the risk assessment report to generate a nurse risk assessment score. The patient information is then transferred to an experienced nurse (the master) who independently reviewed the patient information and generated another risk score (the master score) for the patient. One of the goals of the investigation is to seek whether the nurses underestimate the risk of the patients, namely whether the nurse risk assessment score is lower than the master score on average.

Without any specific model assumption on the data, the Wilcoxon signed rank statistic is used to compare the median score difference for matched observations between the nurse and the master scores, the p-value is 0.036 for the median score difference in risk category 1, and 0.042 for the median score difference in risk category 2.

For the inference of risk score difference separately within each risk category, the p-value of 0.036 indicates that within the risk category 1, the median risk scale of the master score is higher from that of the nurse score. And the p-value of 0.042 shows that for risk category 2, the median master score is higher from that of the nurse score. However, in this study, the primary interest is to see the overall reliability of the nurse scores, and both the nurse and the master scores are strongly influenced by the risk category that the patient is in. Under this scenario, simultaneous inference across strata is called.

The story of making simultaneous inference (for the two risk categories at the same time) is different from making individual inference. This is because the source of errors for simultaneous inference involves more than one random factor. Making simultaneous inference on for median score difference between risk categories necessitates statistical techniques of multiple comparisons, which is discussed in the next section.

3. Bivariate Comparisons

In the statistical analysis of the DVT risk score data, type I error occurs when the procedure incorrectly claims the difference between the two risk assessment scores (but actually there is no difference). At the same time, type II error occurs when the procedure incorrectly claims no difference (but there is actually a difference). Thus, to make a correct conclusion, traditional hypothesis methods keep the rate of type I error at a fixed level (say, 0.05) while minimizing the type II error (which is equivalent to finding a more powerful test).

Notice that for individual inference, controlling type I error within category 1 implies the control of inference error without caring about DVT in the risk category 2. This type of comparison overlooks the relation between the two risk categories. In fact, for this study, the inference method should control all possible false rejections for any risk category in which the master score agrees with the nurse score. Namely, we need to control the error rate of incorrect rejection for DVT merely in risk category 1, merely in risk category 2, or in both risk categories. In multiple comparisons, such requirements are formulated with the terminology of the strong control of familywise error rate (see, for example Hsu [10], or Hochberg and Tamhane [11]).

Although procedures with other testing criteria (such as the control of false discovery rate, or the control of generalized familywise error rate) are available in the literature, these procedures do not always strongly control the familywise error rate. For the DVT risk assessment problem, the special condition on the strong control of familywise error rate limits the candidate inference procedures to the Bonferroni method and the Holm's step-down procedure.

In terms of stepwise multiple comparison procedures, another well-known method is the Hochberg's step-up procedure (which coincides with the celebrated Simes' procedure for bivariate comparisons [12]). However, it should be noted that the Hochberg's step-up procedure is invalid for the JVI inference problem because it does not always strongly control the familywise error rate. The Hochberg's procedure works only for positively dependent populations. It is risky and careless to assume positive dependence (or independence) in this data analysis since the plots of the data indicates that the nurse risk scores may be negatively dependent with the master risk assessment scores for DVT patients. Publications on the invalidity of the Hochberg's step-up procedure include Block et. al [13], Chen [14], Huang and Hsu [15], and Sakar [16], among others. These papers clearly emphasize that the Hochberg's step-up procedure is valid only when the test statistics are positively dependent, including independent, or jointly follow a positively correlated distribution such as the MTP_2 (multivariate totally positive dependence with degree two) distribution. The assumption on the unknown correlation (such as independence or MTP_2) is critical for the validity of the Hochberg's procedure. Ignoring such model assumption may result in an over- or under-estimation in the inference of the risk score difference, and consequently enlarge the error rate. Due to this reason, the next section will focus on the Bonferroni method and the Holm's procedure for methodology comparison of data analysis for the DVT risk assessments scores.

Consider the problem of testing H_1, ..., H_k based on corresponding p-values $\hat{p}_1$ ..., $\hat{p}_k$. Denote the ordered p-values $\hat{p}_{(1)} \le \hat{p}_2 ... \le \hat{p}_k$, and the hypotheses associated with $\hat{p}_{(1)}$, $\hat{p}_{(2)}$, ... $\hat{p}_{(k)}$, as $H_{(1)}$ ..., $H_{(k)}$, respectively.

One-step Bonferroni's procedure: *If* $\hat{p}_{(k)} \le \frac{\alpha}{k}$, *reject all* $H_{(i)}$ *for i=1, 2, ..., k; Otherwise, no conclusive statement for the overall hypotheses, namely at least one of the hypotheses can not be rejected.*

Holm's step-down procedure: *If m is the smallest integer j for which* $\hat{p}_{(j)} \ge \frac{\alpha}{k-j+1}$, *do not reject* $H_{(i)}$ *for i= m,...,k (equivalently, reject* $H_{(i)}$ *for i=1, 2, ..., m-1 when m > 1).*

The above two procedures are applied to the data of DVT risk assessment score as follows. Computing the p-values of individual comparisons and ordering them gets $\hat{p}_{(1)}$=0.036 and $\hat{p}_{(2)}$=0.042. For the Bonferroni method, since $\hat{p}_{(1)}$>0.025 and $\hat{p}_{(2)}$>0.025, no hypothesis is rejected at 0.05 significance level. For the Holm's algorithm, since $\hat{p}_{(1)} \ge \frac{\alpha}{2}$=0.025, Holm's procedure does not reject any hypothesis, (no conclusive statement) either.

Thus, both of the two currently available multiple testing procedures (for the strong control of familywise error rate) fail to detect any difference when the difference exists. The inference conclusion for the two test procedures is that the p-values are not small enough to detect any difference between the nurse DVT risk assessment score and that of the master score *simultaneously* for the two risk categories at 0.05 significance level.

Although such an inconclusive statement is interpretable statistically, it is awkward for practical interpretations, especially when both p-values are small enough to claim statistical difference for each pairwise comparison separately. Recall that the primary interest in the project is to seek whether the median of the nurse risk assessment score is higher than that of the master score in term of DVT risks at the two different risk levels.

In the next section, we develop a new confidence procedure method to make inference for bivariate comparisons. The new procedure detects the median score difference for the JVI data and make inferential conclusions coherent with the conclusions of individual inference on pairwise comparisons.

4. A New Bivariate Confidence Procedure

Consider the one-sided testing problem: $H_1: \eta_1 \le \eta_{10}$ versus $K_1: \eta_1 > \eta_{10}$ for the median risk assessment score comparison between the nurse and the master for patients in risk category 1, and $H_2: \eta_2 \le \eta_{20}$ versus $K_2: \eta_2 > \eta_{20}$ for the median risk assessment score comparison between the nurse and the master score for patients in risk category 2.

Denote $\xi_1 = \eta_1 - \eta_{10}$, and $\xi_2 = \eta_2 - \eta_{20}$, $P(\xi_1 > L_{1j}(Y)) = P(\hat{P}_1 > \frac{\alpha}{j}) = \frac{\alpha}{j}$ and similarly $P(\xi_2 > L_{2j}(Y)) = P(\hat{P}_2 > \frac{\alpha}{j}) = \frac{\alpha}{j}$ for j=1, 2, where $L_{1j}(Y)$ and $L_{2j}(Y)$ are $1-\alpha$ confidence lower bounds (based on the data) of the median score differences $\xi_1 = \eta_1 - \eta_{10}$ and $\xi_2 = \eta_2 - \eta_{20}$ at two risk strata, respectively.

As usual, let W be the Wilcoxon signed rank statistic for the matched observations, and w_α be the value so that $P(W > w_\alpha) = \alpha$ for any $\alpha \in (0, \ 1)$, and $(W_{(1)}, \ W_{(2)})$ be the order statistic of $(W_1, \ W_2)$ corresponding to the ordered p-values $\hat{P}_1 \le \hat{P}_2$, respectively. Also, denote $H_{(1)}$ and $H_{(2)}$ the hypotheses corresponding to the ordered p-values $\hat{P}_{(1)}$ and $\hat{P}_{(2)}$, respectively. With above notations, we have

4.1. Procedure A: Bivariate Confidence Procedure

Step 1: If $\hat{P}_{(2)} < \alpha$, reject $H_{(2)}$, go to Step 2;

else claim $\xi_{(2)} > L_{(2)1}(Y)$, stop.

Step 2: If $\hat{P}_{(1)} < \alpha/2$, reject $H_{(1)}$, stop;

else claim $\xi_{(1)} > L_{(1)2}(Y)$, stop.

In the first step, the procedure examines the largest p-value to see whether it is greater than α. If so, reject the corresponding hypothesis and move on to examine the smallest p-value; if not, stop the procedure and make a confidence estimation on the associated median. In the second step, the procedure examines whether the smallest p-value is greater than $\alpha/2$. If so, it rejects the hypothesis associated with the smallest p-value; otherwise, it produces a confidence estimate for the median score difference that is associated with the smallest p-value.

The following theorem guarantees that Procedure A strongly controls the familywise error rate in the setting of bivariate testing.

Theorem 1: The confidence set obtained from Procedure A has confidence level at least $1-\alpha$, and it strongly controls the familywise error rate at level α in the setting of bivariate testing.

The proof of Theorem 1 is given in the appendix.

In terms of the data set given in the investigation of DVT risk assessment scores between the nurse and the master at two risk categories, we have $\hat{P}_{(1)}$=0.036 and $\hat{P}_{(2)}$=0.042, $L_{11}(Y) = 4$, $L_{21}(Y) = 2$, $L_{12}(Y) = -1$, $L_{22}(Y) = -2$. Thus, the proposed procedure works as follows.

4.2. Data Analysis of the Risk Score Difference in Procedure A

Step 1: Since $\hat{P}_{(2)} = 0.042 < 0.05$, reject $H_{(2)}$, go to Step 2;

Step 2: Since $\hat{P}_{(1)} = 0.036 > 0.025$, claim $\xi_{(1)} > -2$, stop.

Thus, with the strong control of familywise error rate at 0.05, the result of data analysis is that the median nurse score is significantly lower than the median master score for risk category 2, and that the median nurse risk assessment score is lower than two plus the master score for risk category 1.

Note that the above conclusion is coherent with the individual inference that the median nurse risk score is lower than that of the master score in risk category 1, and the same conclusion for the risk category 2.

5. Comparisons with Existing Procedures

Compared with the Bonferroni approach, the new method is more powerful because it rejects any hypothesis that is rejected by the one-step Bonferroni approach while strongly controlling the familywise error rate in the setting of bivariate testing. This is because for a hypothesis to be rejected by the one-step Bonferroni approach, the two p-values should be less than $\alpha/2$. In this case, the new procedure will reject both hypotheses because it only needs the larger p-value less than α and the smallest p-value less than $\alpha/2$ to reject the two hypotheses.

Compared with the Holm's approach, the new method is more powerful when the largest p-value is less than α. In this case, if the smallest p-value is less than $\alpha/2$, both the Holm's method and the new method reject the two hypotheses. However, if the smallest p-value exceeds $\alpha/2$, the Holm's method is unable to make any conclusion while the new method is able to reject the hypothesis that is associated with the largest p-value. This is the case for the DVT risk score analysis between the nurse assessment and the master assessment at JVI.

It should be noted that in the case when the largest p-value is larger than α, the new procedure is unable to reject any hypothesis, but the Holm's method may reject the hypothesis associated with the smallest p-value when it is below $\alpha/2$. In the case when the largest p-value exceeds α, the individual inference is unable to detect any median risk score difference. Under this scenario, the new procedure provides a confidence lower bound for the information on the extent of the median score difference.

As discussed in the previous section, the Hochberg's step-up testing procedure is unable to strongly control the familywise error rate for the JVI data. However, the new method takes the form of step-up (in the sense that it screens the p-value from larger to smaller) and keeps the strong control of familywise error rate without making any distribution assumption on the positive dependence for the test statistics.

6. Conclusion

The new confidence procedure proposed in this paper is versatile in the sense that it is applicable to different scenarios according to the need of data analysis. For example, we use the Wilcoxon signed rank statistic in this paper; however, the procedure is readily used when the test statistics are t-test statistics, without assuming the condition of positive dependence. The new method may also be extended to the comparisons of survival curves stratified at different risk levels.

Besides data analyses on vascular nursing care administration, the new method can be applied to statistical problems posted for other fields of data analysis for administration and management when the positive dependence assumption is implausible. For example, to compare two treatments with a placebo in lung cancer investigations [17], to compare the effect of oral cladribine on multiple sclerosis [18] at two different risk status, or to compare for safety issues of elderly persons [19] to list just a few. In terms of

methodological development, the new method can be applied to binary data to enhance the power of data analysis for the efficacy of a drug [20] or [21] at two treatment levels. In conjunction with the simultaneous confidence segments [22] or [23], the new method can be applied to improve the analysis of shelf-lives for two chemical compounds. The new bivariate testing procedure can also be applied for parametric models with p-values computed from a specific model, in which a good resource of model selection criterion contains [24,25].

Although the new method advances the methodology of multiple comparisons, the limit of the new procedure is that it is only for simultaneous comparison of two populations. Further investigation is focused on the extension of the current procedure to the comparisons of any k populations.

Acknowledgements

The authors would like to thank the Editor-in-Chief, Professor Eyüp Çetin, for his kind invitation to publish their article in this important volume of the Istanbul University Journal of the School of Business Administration.

References

[1] M. Silverstein, et al., Trends in the Incidence of Deep Vein Thrombosis and Pulmonary Embolism: A 25-year Population-based Study. *Archives of Internal Medicine*, 158, 585-593 (1998).

[2] R.H. White, The Epidemiology of Venous Thromboembolism. *Circulation*, 107; I4-I8 (2003).

[3] J. Heit, *American Society of Hypertension Meeting*, 2006.

[4] N. Browse and M. Thomas, Source of Non-lethal Pulmonary Embolism. *Lancet, 1,* 258-259 (1974).

[5] A. Comerota, et al., Endovenectomy of the Common Femoral Vein and Intraoperative Iliac Vein Recanalization for Chronic Iliofemoral Venous Occlusion. *Journal of Vascular Surgery*, North American Chapter, 52, 1, 243-247 (2010).

[6] C. Kearon, et al., Antithrombotic Therapy for Venous Thromboembolic Disease. *American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition) CHEST*, 133, 6, 454S-545S (2008).

[7] S. Galson, Surgeon General's Perspectives, Prevention of Deep Vein Thrombosis and Pulmonary Embolism*, Public Health Reports*, July-August, 123 (2008).

[8] W. Geerts, et al., American College of Chest Physicians. Prevention of venous thromboembolism: *American College of Chest Physicians Evidence-Based Clinical Practice Guidelines* (8th Edition). *Chest* 2008; 133(6 Suppl):381S-453S.

[9] S. Goodacre, et al., Meta-analysis: the value of clinical assessment in the diagnosis of deep venous thrombosis. *Ann Intern Med*.; 143(2): 129-39, (2005).

[10] J.C. Hsu, Multiple Comparisons: Theory and Methods, London, Chapman and Hall, 1996.

[11] Y. Hochberg, and A. C. Tamhane, *Multiple Comparison Procedures*, New York, Wiley, (1987).

[12] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800-802 (1988).

[13] H. Block, T. Savits, and J. Wang, Negative Dependence and the Simes Inequality. *Journal of Statistical Planning and Inference*, 138, 12, 4107-4110 (2008).

[14] J.T. Chen, A Two-stage Estimation Procedure*. Biometrics*, 64, 406-412 (2008).

[15] Y. Huang and J. Hsu, Hochberg's Step-up Method: Cutting Corners off Holm's Step-down Method. *Biometrika*, 94, 965-975 (2007).

[16] S.K. Sarkar, Some Probability Inequalities for Ordered MTP_2 Random Variables: A Proof of the Simes Conjecture. *Annals of Statistics*. 26, 494-504 (1998).

[17] F. Shepherd, et al., Erlotinib in Previously Treated Non-mall-cell Lung Cancer. *The New England Journal of Medicine*, 353, 2, 123-132 (2005).

[18] G. Giovannoni, et al., A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. *The New England Journal of Medicine*, Vol 362, 5, 416-426 (2010).

[19] T. Gill, et al., A program to prevent functional decline in physically frail elderly persons who live at home. *The New England Journal of Medicine*, Vol 347, 14, 1068-1074 (2002).

[20] J.T. Chen, Inference on the Minimum Effective Dose Using Binary Data. *Communications in Statistics Theory and Methods*, 37, 2124-2135, (2008).

[21] J.C. Hsu, and R. Berger, Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies, *Journal of the American Statistical Association*, 94, 468-482 (1999).

[22] X. Lu and J.T. Chen, Exact Simultaneous Confidence Segments for all Contrast Comparisons. *Journal of Statistical Planning and Inference*, 139, 2816-2822 (2009).

[23] J.T. Chen and F. Hoppe, Simultaneous Confidence Intervals. *The Encyclopedia of Biostatistics*, 5, 4114-4116, Editted by Peter Armitage, John Wiley & Sons, New York, 1998.

[24] H. Bozdogan, A New Class of Information Complexity (ICOMP) Criteria with an Application to Customer Profiling and Segmentation. *Istanbul University Journal of the School of Business Administration*, 39, 2, 370-398 (2010).

[25] H. Bozdogan and P. Bearse, Information Complexity Criteria for Detecting Influential Observations in Dynamic Multivariate Linear Models Using the Genetic Algorithm*. Journal of Statistical Planning and Inference,* 114, 31-44 (2003).

Appendix

Proof of Theorem 1: We need the following two lemmas for the proof of Theorem 1. For any two sets A and B, we denote $A \cap B$ as AB in the sequel for notational convenience. Let $Q_i = \{\eta_i \le 0\}$, $C_i = (L_{i1}(Y), \quad \infty)$, and $D_i = (L_{i2}(Y), \quad \infty)$ for i=1, 2.

- ***Lemma 1***

$$P(\eta_2 \in C_{(2)} \mid \hat{P}_{(2)} > \alpha) \ge P((\eta_{1,}\eta_2) \in [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}D_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}] \mid \hat{P}_{(2)} > \alpha) \quad (1)$$

Proof: Since $\hat{P}_{(2)} > \alpha$, the corresponding $1-\alpha$ confidence lower bound is lower than zero. Thus $C_{(2)} = (L_{(2)1}(Y), \quad \infty) \supseteq Q^c_{(2)}$ and $Q^c_{(2)} = Q^c_{(2)} \cap C_{(2)}$, now when $\hat{P}_{(2)} > \alpha$,

$$\begin{aligned}
&[Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}D_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}] \\
&\subseteq [Q_{(2)}C_{(2)}] \cup Q^c_{(2)} \\
&= [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}C_{(2)}] \\
&= C_{(2).}
\end{aligned}$$

This completes the proof of Lemma 1.

- ***Lemma 2***

$$\begin{aligned}
&P((\eta_{1,}\eta_2) \in Q^c_{(2)}D_{(1)} \mid \hat{P}_{(2)} < \alpha, \hat{P}_{(1)} > \frac{\alpha}{2}) \\
&\ge P((\eta_{1,}\eta_2) \in [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}C_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}] \mid \hat{P}_{(2)} < \alpha, \hat{P}_1 > \frac{\alpha}{2})
\end{aligned}$$

Proof: When $\hat{P}_{(2)} < \alpha$, the corresponding $1-\alpha$ lower confidence bound is larger than zero, thus $C_{(2)} \subseteq Q^c_{(2)}$, and $Q_{(2)} \cap C_{(2)} = \phi$.

When $\hat{P}_{(1)} > \alpha/2$, the corresponding $1-\alpha$ confidence lower bound is less than zero, thus $D_{(1)} \supseteq Q^c_{(\backslash 1)}$, Therefore, when $\hat{P}_{(2)} < \alpha$, $\hat{P}_{(1)} > \alpha/2$, we have

$$\begin{aligned}
&[Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}D_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}] \\
&= [Q^c_{(2)}Q_{(1)}D_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}] \\
&= [Q^c_{(2)}D_{(1)}] \cap [Q_{(1)} \cup Q^c_{(1)}D_{(2)}] \\
&\subseteq Q^c_{(2)}D_{(1).}
\end{aligned}$$

This completes the proof of Lemma 2.

- ***Lemma 3***

$$P((\eta_1,\eta_2) \in Q_{(2)}^c Q_{(1)}^c \mid \hat{P}_{(2)} < \alpha, \hat{P}_{(1)} < \frac{\alpha}{2})$$

$$\geq P((\eta_1,\eta_2) \in [Q_{(2)} C_{(2)}] \cup [Q_{(2)}^c Q_{(1)} C_{(1)}] \cup [Q_{(2)}^c Q_{(1)}^c D_{(1)} D_{(2)}] \mid \hat{P}_{(2)} < \alpha, \hat{P}_1 < \frac{\alpha}{2})$$

Proof: When $\hat{P}_{(2)} < \alpha$, the corresponding $1-\alpha$ lower confidence bound is larger than zero, thus $C_{(2)} \subseteq Q_{(2)}^c$, and $Q_{(2)} \cap C_{(2)} = \phi$.

When $\hat{P}_{(1)} < \alpha/2$, the corresponding $1-\alpha$ confidence lower bound is larger than zero, thus $D_{(1)} \subseteq Q_{(\backslash 1)}^c$, and $Q_{(1)} \cap C_{(1)} = \phi$. Therefore, when $\hat{P}_{(2)} < \alpha$, $\hat{P}_{(1)} < \alpha/2$, we have

$$[Q_{(2)} C_{(2)}] \cup [Q_{(2)}^c Q_{(1)} D_{(1)}] \cup [Q_{(2)}^c Q_{(1)}^c D_{(1)} D_{(2)}]$$
$$= [Q_{(2)}^c Q_{(1)}^c D_{(1)} D_{(2)}]$$
$$\subseteq Q_{(2)}^c Q_{(1)}^C$$

This completes the proof of Lemma 3.

- ***Lemma 4***

$$P(\eta_{(2)} \in C_{(2)}) \geq 1-\alpha.$$

Proof: If $\hat{P}_1 > \hat{P}_2$, by the relationship between p-values and the confidence bounds,

$$P(\eta_{(2)} \in C_{(2)})$$
$$= P(Y : \hat{P}_{(2)} > \alpha)$$
$$= P(Y : \{\hat{P}_1 > \alpha\} \cup \{\hat{P}_2 > \alpha\})$$
$$\geq P((Y : \hat{P}_1 > \alpha)$$
$$= 1-\alpha.$$

This completes the proof of Lemma 4.

With Lemmas 1-4, we are ready to prove Theorem 1 as follows.

Proof of Theorem 1: To prove the validity of Procedure A in terms of the strong control of familywise error rate, consider any subset $A \subset I = \{1,2\}$, we have

P(Procedure A rejects $H_i, i \in A \mid H_i, i \in A$ true)

$\leq$ P(Procedure A makes at least one error on the inference for $(\eta_1 \quad \eta_2)$)

=1- P(Confidence set of Procedure A correctly covers the true parameters $(\eta_1 \quad \eta_2)$). (2)

Now for the partition of the sample space $\{\hat{P}_{(2)} > \alpha\}$, $\{\hat{P}_{(2)} < \alpha, \hat{P}_{(1)} > \alpha/2\}$, and $\{\hat{P}_{(2)} < \alpha, \hat{P}_{(1)} < \alpha/2\}$, we have

P(Confidence set of Procedure A correctly covers the true parameter vector $(\eta_1 \quad \eta_2)$)

$$= a_1 P(\hat{P}_{(2)} > \alpha) + a_2 P(\hat{P}_{(2)} < \alpha, \hat{P}_{(1)} > \alpha/2) + a_3 P(\hat{P}_{(2)} < \alpha, \hat{P}_{(1)} < \alpha/2), \quad (3)$$

where the constants a_1, a_2, and a_3 are as follows,

$a_1 =$ P(Procedure A claims $\eta_{(2)} \in C_{(2)} \mid \hat{P}_2 > \alpha)$

$a_2 =$ P(Procedure A claims $\eta_{(2)} \in Q^c_{(2)}, \eta_{(1)} \in D_{(1)} \mid \hat{P}_2 < \alpha, \hat{P}_{(1)} > \alpha/2)$

$a_3 =$ P(Procedure A claims $\eta_{(2)} \in Q^c_{(2)}, \eta_{(1)} \in Q^c_{(1)} \mid \hat{P}_2 < \alpha, \hat{P}_{(1)} < \alpha/2)$.

By Lemma 1, we know that in (3),

$a_1 =$ P(Procedure A claims $\eta_{(2)} \in C_{(2)} \mid \hat{P}_2 > \alpha)$

$$\geq P((\eta_1, \eta_2) \in [Q_{(2)} C_{(2)}] \cup [Q^c_{(2)} Q_{(1)} D_{(1)}] \cup [Q^c_{(2)} Q^c_{(1)} D_{(1)} D_{(2)}] \mid \hat{P}_{(2)} > \alpha) \quad (4)$$

By Lemma 2, we know that in (3)

$a_2 =$ P(Procedure A claims $\eta_{(2)} \in Q^c_{(2)}, \eta_{(1)} \in D_{(1)} \mid \hat{P}_2 < \alpha, \hat{P}_{(1)} > \alpha/2)$

$$\geq P((\eta_1, \eta_2) \in [Q_{(2)} C_{(2)}] \cup [Q^c_{(2)} Q_{(1)} C_{(1)}] \cup [Q^c_{(2)} Q^c_{(1)} D_{(1)} D_{(2)}] \mid \hat{P}_{(2)} < \alpha, \hat{P}_1 > \frac{\alpha}{2}). \quad (5)$$

When $\hat{P}_{(2)} < \alpha, \hat{P}_{(1)} < \frac{\alpha}{2}$, by Lemma 3, we have in (3),

$a_3 =$ P(Procedure A claims $\eta_{(2)} \in Q^c_{(2)}, \eta_{(1)} \in Q^c_{(1)} \mid \hat{P}_2 < \alpha, \hat{P}_{(1)} < \alpha/2)$

$$\geq P((\eta_1, \eta_2) \in [Q_{(2)} C_{(2)}] \cup [Q^c_{(2)} Q_{(1)} C_{(1)}] \cup [Q^c_{(2)} Q^c_{(1)} D_{(1)} D_{(2)}] \mid \hat{P}_{(2)} < \alpha, \hat{P}_1 < \frac{\alpha}{2}). \quad (6)$$

Considering Equation (3) in conjunction with (4), (5), and (6) yields

P(Confidence sets of Procedure A correctly cover the median differences (η_1, η_2))

$$\geq P((\eta_1, \eta_2) \in [Q_{(2)} C_{(2)}] \cup [Q^c_{(2)} Q_{(1)} C_{(1)}] \cup [Q^c_{(2)} Q^c_{(1)} D_{(1)} D_{(2)}]). \quad (7)$$

Since $L_{ij}(Y)$ is the lower confidence bound corresponding to $\hat{P}_i > \alpha / j$ for i=1, 2 and j=1, 2, we have the following conclusions. If $\hat{P}_1 < \hat{P}_2$, then $L_{11}(Y) > L_{21}(Y)$, $Q_{(i)} = Q_i$, $C_{(i)} = C_i$, and $D_{(i)} = D_i$, for i=1, 2. If $\hat{P}_1 > \hat{P}_2$, then $L_{11}(Y) < L_{21}(Y)$, $Q_{(1)} = Q_2$, $Q_{(2)} = Q_1$ $C_{(1)} = C_2, C_{(2)} = C_1$ and $D_{(1)} = D_2$, $D_{(2)} = D_1$.

Now notice that the parameter space can be partitioned into $Q_{(2)} \cup [Q^c_{(2)} Q_{(1)}] \cup [Q^c_{(2)} Q^c_{(1)}]$, by the method of simultaneous confidence intervals (see, for example, [17]), we have

$$P((\eta_1,\eta_2) \in [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}C_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}])$$

$$= P(\eta_{(2)} \in C_{(2)}), \text{ when } (\eta_1,\eta_2) \in Q_{(2)}$$

$$\geq 1-\alpha;$$

$$P((\eta_1,\eta_2) \in [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}C_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}])$$

$$= P(\eta_{(1)} \in D_{(1)}), \text{ when } (\eta_1,\eta_2) \in Q^c_{(2)}Q_{(1)}$$

$$\geq 1-\alpha;$$

$$P((\eta_1,\eta_2) \in [Q_{(2)}C_{(2)}] \cup [Q^c_{(2)}Q_{(1)}C_{(1)}] \cup [Q^c_{(2)}Q^c_{(1)}D_{(1)}D_{(2)}])$$

$$= P(\eta_{(2)} \in D_{(2)}, \eta_{(1)} \in D_{(1)}), \text{ when } (\eta_1,\eta_2) \in Q^c_{(2)}Q^c_{(1)}$$

$$\geq 1-\alpha;$$

Because the vector of parameters (η_1,η_2) can only belong to one of the subsets of the partition of the parameter space, the above three cases exhausts all the possible cases. This completes the proof of Theorem 1.

Table 1 Risk Factors for Patients in Risk Category 1

1	Age from 40 to 59
2	Bed confinement greater than 48 hours
3	Presence of varicose veins
4	Presence of leg edema, ulcer, or stasis
5	Obesity
6	Myocardial infarction
7	Congestive heart failure
8	Serve chronic obstructive pulmonary disease
9	Crystalloids greater than 5L/24 hrs
10	Confining travel for more than 4 hours in the past 2-4 weeks
11	Postpartum in the past one month
12	Inflammatory bowel disease
13	Severe infection/sepsis
14	Estrogen replacement
15	Operation greater than 2 hours

Table 2: Risk Factors for Patients in Risk Category 2

1	Age greater than or equal to 60
2	Stroke
3	Trauma
4	Pelvic operation
5	Joint replacement
6	Hip fracture
7	Malignancy
8	Pelvic/long bone fracture
9	Hyper-coagulable state (thrombophylia, factor V Leiden, protein S deficiency)
10	Family history of DVT