e-Health – For Continuity of Care
C. Lovis et al. (Eds.)

doi:10.3233/978-1-61499-432-9-48

Towards a Web-Based Decision Support Tool for Selecting Appropriate Statistical Test in Medical and Biological Sciences

Aslı SUNER[a,1,2], Gökhan KARAKÜLAH[b,2] and Oğuz DİCLE[c,d]
[a] *Ege University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, İzmir, Turkey*
[b] *Dokuz Eylül University, Health Sciences Institute, İzmir, Turkey*
[c] *Dokuz Eylül University, Faculty of Medicine, Department of Radiology, İzmir, Turkey*
[d] *Dokuz Eylül University, Health Sciences Institute, Department of Medical Informatics, İzmir, Turkey*

Abstract. Statistical hypothesis testing is an essential component of biological and medical studies for making inferences and estimations from the collected data in the study; however, the misuse of statistical tests is widely common. In order to prevent possible errors in convenient statistical test selection, it is currently possible to consult available test selection algorithms developed for various purposes. However, the lack of an algorithm presenting the most common statistical tests used in biomedical research in a single flowchart causes several problems such as shifting users among the algorithms, poor decision support in test selection and lack of satisfaction of potential users. Herein, we demonstrated a unified flowchart; covers mostly used statistical tests in biomedical domain, to provide decision aid to non-statistician users while choosing the appropriate statistical test for testing their hypothesis. We also discuss some of the findings while we are integrating the flowcharts into each other to develop a single but more comprehensive decision algorithm.

Keywords. Statistical hypotesis testing, decision making, decision support, biostatistics

Introduction

Statistical hypothesis testing is an essential component of biological and medical studies for making inferences and estimations from the collected data in the study; however, the misuse of statistical tests is widely common [1-6]. Previous studies showed that researchers frequently made errors while deciding the selection of appropriate statistical tests and in sample size calculation, clinical trial planning, reporting and interpreting statistical results, selection of control group, use of figures and tables, and obtained false positive results due to multiple testing [6-9]. Among the statistical errors, inappropriate statistical test selection usually leads to obtaining inaccurate interpretations of results, wrong conclusions and wasting researcher's time

[1] Corresponding author, e-mail: asli.suner@ege.edu.tr
[2] These authors contributed equally to this work.

[8].

In order to prevent possible errors in convenient statistical test selection, it is currently possible to consult available test selection algorithms developed for various purposes [1-5, 9-13]. In these decision algorithms, users are asked a series of questions regarding their data set such as sample size, normality assumption, scale type of variables, and dependent/independent samples and their aims as well. After answering all questions properly, the algorithms provide a suitable statistical hypothesis test, fitting user's demand. Some decision algorithms, for instance, UCLA's flowchart does not only provide decision support in the selection of appropriate statistical method but also introduce detailed documentation regarding how the statistical test suggested by the algorithm can be applied in distinct statistical packages such as SPSS and R [9]. However, none of the currently available algorithms covers and is designed for all parametric and nonparametric statistical tests commonly used in biomedical domain, and this situation causes several problems such as shifting users among the algorithms, poor decision support in test selection and lack of satisfaction of potential users. Therefore, in this study, we aimed at creating single and comprehensive decision algorithm by modifying and integrating distinct algorithms into each other. In addition to this, by creating descriptions for the concepts observed in simple decision questions of our unified algorithm, we created an initial framework of a decision support tool for appropriate test selection in hypothesis testing of biological and medical researches.

1. Methods

In order to seek out available decision algorithms for statistical test selection, we carried out a manual search in the literature and text books as well as on the Internet. After determining decision algorithms, we individually examined algorithms to identify for which statistical tests each algorithm was used to select. Furthermore, each algorithm was studied in terms of decision questions that comprised the algorithms and were answered for the selection of the most appropriate statistical test in the analysis of the data at hand. For this purpose, the number and attributes of decision questions in each algorithm were evaluated. Five statisticians ran the algorithms, expert in biostatistics, to clearly understand whether each of them works properly and provide accurate guidance. By this means, we determined possible potential shortages of each algorithm. For instance, if one wishes to compare non-normally distributed and paired two samples but the algorithm does not offer any non-parametric statistical tests, we noted this as a lack of algorithm. Later, we created a single but comprehensive algorithm for statistical test selection commonly used in biomedical researches by assembling algorithms currently in use. We selected Rosner's flowchart [13] as a base and integrated other algorithms into this flowchart. Additionally, we manually created descriptions for non-statistician users in each decision question of the unified algorithm. A list of fundamental statistical terms, concepts and their meaning was also built to provide user friendly and effective decision support in statistical test selection in biomedical domain.

2. Results

The manual screening of various sources revealed the presence of 10 distinct algorithms and decision tables, which facilitated the selection of the most appropriate statistical test according to data characteristics and researcher's purpose. Table 1 shows the decision algorithms found by manual search and statistical tests related to those algorithms.

Table 1. A list of decision algorithms, suggested statistical tests and number of tests used in these algorithms.

Decision Algorithm	Number of Tests	Suggested Statistical Tests
Rosner (2000)	36	Parametric, nonparametric and multivariate methods
UCLA (2014)	31	Parametric, nonparametric and multivariate methods
Nayak&Hazra (2011)	30	Parametric and nonparametric methods
Jaykaran (2010)	25	Parametric, nonparametric and multivariate methods
Mertler (2002)	17	Multivariate statistical methods
Gunawardena (2011)	14	Parametric and nonparametric methods
McCrum- Gardner (2008)	13	Parametric, nonparametric and multivariate methods
Marusteri&Bacarea (2010)	11	Parametric and nonparametric methods
Twycross&Shields (2004)	11	Parametric and nonparametric methods
Gaddis&Gaddis (1990)	9	Nonparametric methods

Rosner's flowchart is the most detailed one and it provides guidance for 36 different statistical tests used in biomedical research. The statistical tests, including, t-test, ANOVA, chi-square test, Kruskal-Wallis H test, Mann-Whitney U test, and Wilcoxon signed rank test were common in all 10 decision algorithms. The run of each algorithm by the experts showed that all algorithms included in the scope of this study worked properly and presented accurate guidance in test selection; however none of the algorithms covered the whole statistical test range frequently used in biomedical domain in a single decision flowchart. Therefore, the unification of all decision algorithms into a single algorithm resulted in the creation of a complete algorithm bearing 60 separate statistical tests.

Our unified algorithm also includes decision questions for potential users to help for determining the statistical test or tests which is appropriate for the user's aim and might be applied on the data set at user's hand. The questions basically cover the number of variables, number of samples, dependency of samples, normality assumption, and scale type of variables of the data as well as some additional questions for particular cases such as survival analysis and multivariate statistical techniques. All decision questions are in the form of yes/no questions. The longest path of our unified algorithm consists of 10 questions, and that path ends with suggestion of Cox proportional hazards model method. In addition to this, the shortest path ends with simple linear regression, and the path contains four yes/no questions.

Our unified statistical test selection algorithm is mainly built on Rosner's decision algorithm by revising and modifying of available other statistical test selection algorithms. In Figure 1, a sample branch of our decision algorithm for Repeated Measures ANOVA and Friedman tests is given. In this example, our statistical test selection algorithm offers 2 distinct statistical hypothesis tests in a case where the potential user is interested in one variable and would like to compare more than two paired samples. Depending on the normality assumption, one of the two suggested tests can be selected by the user. Additionally, our algorithm also informs the users how normality assumption of the data set can be checked.

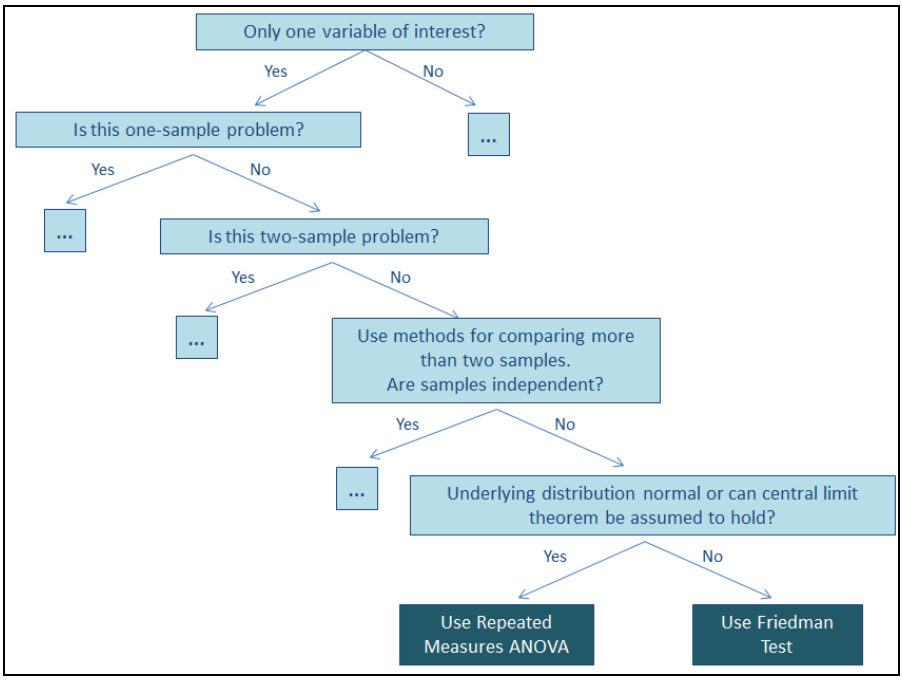

Figure 1. A sample branch of our unified algorithm for appropriate statistical test selection.

3. Discussion

The use of statistical hypothesis testing methods is essential for clarifying the meaning and significance of the collected data in biomedical research. Due to abundance of available statistical tests, the selection of the most appropriate statistical test is generally challenging among the non-statistician researchers. In addition to this, there is a lack of an algorithm presenting the most common statistical tests used in biomedical research in a single flowchart. Thus, this study described a unified algorithm; covers mostly used statistical tests in biomedical domain, to provide decision aid to non-statistician users while choosing the appropriate statistical test for testing their hypothesis.

In this study, Rosner's flowchart, the most comprehensive algorithm with regard to the number of statistical tests, has become the backbone of our unified algorithm, Rosner's flowchart consists of yes/no questions only and guides its users with relatively simple questions at each decision step; therefore we selected Rosner's approach as a base algorithm. On the other hand, Rosner's approach does not include multivariate statistical techniques such as Factor Analysis, Discriminant Analysis, and Principal Analysis. Not limited to these statistical tests however, Rosner's algorithm also does not provide decision support in the selection for multiple comparison methods, including, Mann Whitney U Test, Wilcoxon rank sum test, Dunn Test, Bonferonni's Test, Tukey's HSD Test, Scheffe's Test, and Dunnett's Test as well as nonparametric methods, such as Sign Test, Kendall W Test, Friedman Test and Cochran's Q Test. Thus we modified Rosner's flowchart by integrating other

algorithms and in this way we created a more comprehensive algorithm which could be used for different demands in statistical test selection.

To gain convenient and satisfactory results from the unified algorithm, it is essential to answer questions properly at each decision step. In addition to this, the main goal of this work was to develop a comprehensive decision support algorithm for non-statistician users. Therefore we also manually created descriptions of the statistical terms and concepts observed in the decision questions of our unified algorithm. By doing so, we aimed to simplify the questions asked to non-statistician potential users such as medical practitioners and inexperienced statisticians while determining the most appropriate tests in line with their purposes. Thus, we believe the creation of descriptions manually makes the algorithm as convenient as possible and easy to use.

Another issue regarding our unified algorithm to be discussed in detail is the accuracy of results generated by our unified algorithm. While each algorithm included in this study was manually controlled by experts, obtaining exact results from the flowchart completely depends on correctness of the user's answers at each decision step. Furthermore, we strongly recommend potential users to control the recommendations of the algorithm by referring to an expert statistician to prevent potential misuse of statistical tests.

In conclusion, here we created a framework algorithm of a potential computerized decision support tool, which enables users to select the correct statistical tests in their research. We are currently developing a web based decision support tool to practically apply the unified algorithm in this study. The decision support tool will make use of the structured algorithm proposed in this study and can be utilized while selecting the appropriate method for analyzing the data at hand.

References

[1] B.K. Nayak & A. Hazra, How to choose the right statistical test? *Indian Journal of Ophthalmology*, **59**(2) (2011), 85-86.

[2] Jaykaran, How to select appropriate statistical test? *Journal of Pharmaceutical Negative Results,* **1** (2010), 61-63.

[3] A. Twycross & L. Shields, Statistics made simple: part 4, Choosing the right statistical test. *Paediatric Nursing*, **16**(8) (2004), 24.

[4] N. Gunawardena, Choosing the correct statistical test in research. *Sri Lanka Journal of Child Health,* **40**(4) (2011), 149-153.

[5] E. McCrum-Gardner, Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, **46**(1) (2008), 38-41.

[6] V.B. Nyirongo, M.M. Mukaka & L.V. Kalilani-Phiri, Statistical pitfalls in medical research. *Malawi Medical Journal*, **20**(1) (2008), 15-18.

[7] A.M. Strasak, Q. Zaman, K.P. Pfeiffer, G. Gobel & H. Ulmer, Statistical errors in medical research-a review of common pitfalls. *Swiss Medical Weekly*, **137** (2007), 44-49.

[8] T. Lang, Twenty statistical errors even you can find in biomedical research articles. *Croatian Medical Journal* **45** (2004), 361-370.

[9] What statistical analysis should I use? UCLA: Statistical Consulting Group. from http://www.ats.ucla.edu/stat/mult_pkg/whatstat/ (Last accessed: January 20, 2014).

[10] C.A. Mertler & R. A. Vannatta, *Advanced and Multivariate Statistical Methods*, Pyrczak, Los Angeles, 2002.

[11] M. Marusteri & V. Bacarea, Comparing groups for statistical differences: how to choose the right statistical test? *Biochemia Medica*, **20**(1) (2010), 15-32.

[12] G. M. Gaddis & M. L. Gaddis, Introduction to biostatistics: Part 5, Statistical inference techniques for hypothesis testing with nonparametric data. *Annals of Emergency Medicine*, **19**(9) (1990), 1054-1059.

[13] B.A. Rosner, *Fundamentals of Biostatistics*. (5th Ed), Duxbury Press, USA, 2000.