Editorial

Applications of Medical Informatics and Data Analysis Methods

Pentti Nieminen

Medical Informatics and Data Analysis Research Group, University of Oulu, 90014 Oulu, Finland; pentti.nieminen@oulu.fi

check for updates

1. Introduction

The science of statistics contributes to the development and application of tools for the design, analysis, and interpretation of empirical medical studies. The development of new statistical tools for medical applications depends on the innovative use of statistical inference theory, good understanding of clinical and epidemiological research questions, and an understanding of the importance of statistical software. First, statisticians develop a method in response to a need felt in a particular field of the health sciences, after which the new method is disseminated in the form of presentations, reports, and publications. It is also necessary to develop tools for implementing the method: software and manuals. From this point onwards, the extent to which the procedure is adopted will depend on its usefulness. The broader introduction and acceptance of a new analysis method (as useful as the method might be) into medical and health care publications seems to require the method being incorporated into the standard statistical packages generally used by researchers. In addition, if readers do not understand the mathematics or reporting style, or if the conclusions have been drawn on the basis of advanced mathematics or computationally complex procedures not visible in the data (tables or graphs) presented, then clinicians may not be convinced of the results. The lead time from the description of a new technique to its entering into the practice of medical investigators is long [1].

Unsustainable promises and unfulfillable expectations should be avoided in the context of data mining and machine learning [2]. The broader introduction and expansion of a new analysis method to medical publication seems to require that the method helps to solve a data analysis problem, where basic statistical methods have not been useful or applicable. Simpler classical approaches can often provide elegant and sufficient answers to important questions.

This Special Issue on *Medical Informatics and Data Analysis* was an opportunity for the scientific community to present research on the application and complexity of data analytical methods, and to give insight into new challenges in biostatistics, epidemiology health sciences, dentistry, and clinical medicine. The 13 contributed articles belong to four broad groups: (i) basic statistical methods, (ii) data-oriented practical approaches, (iii) complex machine learning and deep learning predictive algorithms, (iv) medical informatics.

2. Basic Statistical Methods

All basic data analysis methods and multivariable techniques depend on assumptions about the characteristics of the data [3]. If an analysis is performed without satisfying these assumptions, incorrect conclusions may be made on the basis of erroneous results. A normal distribution of main outcome variables is a strong requirement in several statistical techniques and should be verified and reported. In their work, Hanan M. Hammouri and coworkers [4] compare the use of a *t*-test on log-transformed data and the use of a generalized linear model (GLM) on untransformed skewed data. Scientists in biomedical and psychosocial research need to deal with non-normal skewed data all the

time. Hammouri et al. [4] present three examples with real-life data. Their findings show that the *t*-test with log transformation has superior performance over the GLM method for any data that are not normal and follow beta or gamma distributions. Alternatively, for exponentially distributed data, the GLM method has superior performance over the *t*-test with log transformation.

Several findings have demonstrated that too many medical articles do not provide a sufficiently clear, accurate, or complete account of what was done and what was found. In his article, *Ten Points for High-Quality Statistical Reporting and Data Presentation* [5], Pentti Nieminen proposes an applicable checklist for quickly testing the statistical reporting quality of manuscripts and published research papers. The developed instrument is applicable for a wide variety of medical and health care research forums, including both clinical and basic sciences. Editors and reviewers could use the short quality test proposed in this paper for deciding when the presentation in a manuscript is clearly inadequate. If the reviewer cannot find the basic information and description related to the data analysis, the reviewer does not need to read the whole article. After checking tables and figures and reading through the statistical analysis subsection in the methods section, the reviewer can reject the manuscript on good grounds. When the proposed simple quality test shows that the statistical reporting and data presentation are appropriate, the whole article needs to be read and further reviewed [5].

3. Data-Oriented Practical Approaches

Advances in health information technology are enabling a transformation in health research that could facilitate studies that were not feasible in the past, and thus, lead to new insights regarding health and disease. The extent to which new procedures are adopted will depend on their usefulness. It is important that new methods are applied on real data that arise in medical research. Special attention should be given to the practical aspects of analysis and the presentation of the results.

Byung Mook Weon [6] contributes to the Special Issue with applications of modelling life expectancy and population dynamics. The title of this nice piece of work is *Stretched Exponential Survival Analysis for South Korean Females*. The paper focuses on studying current trends of lifespan among South Korean females using modified survival curves. The study shows the quantitative and comparative evidence for a remarkable rapid increase in female lifespan in South Korea during three recent decades, from 1987 to 2016.

Long-term care (LTC) involves a variety of services designed to meet people's health or personal care needs during a short or long period of time. A paper authored by Shun-Hsing Chen, Fan-Yun Pa, and Tsu-Ming Yeh [7] includes an interesting review of different models and methods to examine long-term care service demands and satisfaction improvement. Using data from the older adult population in Taiwan (n = 292), this study demonstrates how two methods can be integrated to serve as a basis for decision makers to adjust LTC service quality design and improve care for older adults. The reproducibility of the proposed integration is easy.

Vocal fatigue may be experienced by any individuals during their lifetime, but it is more frequently encountered by professional voice users in occupational settings. Vocal fatigue increases vocal effort and decreases speaking stamina. Zhengdong Lei and co-authors [8] give in their contribution an extensive examination of the effect of vocal loading on a large number of voice measures and ratings in a small group of vocally normal young women. The novel aspect of the work is the use of vocal dosing as a criterion for performance. Their paper is rich with data, which provides relevant evidence about the acoustic and perceptual manifestations of vocal fatigue.

The paper *Classification Maps in Studies on the Retirement Threshold* by Agnieszka Bielinska and collaborators [9] is an example of a study about retirement age in Poland. The aim of this work is to present new classification maps in health research and to show that they are useful in data analysis. Groups of individuals and their answers to questions of expectations and worries related to the retirement threshold are analyzed. A statistical method, correspondence analysis, is applied for obtaining these maps. With the classification maps, it is possible to find subgroups of these individuals who answer in a similar way to the specific questions. In addition, the authors compare structures of

the maps searching for factors such as gender, marital status, kind of work, and economic situation, which are essential at the retirement threshold.

4. Complex Machine Learning and Deep Learning Predictive Algorithms

During the recent decades, mathematical statisticians have introduced new data analysis methods marked by the rapid expansion of computing efficiency and the advancement in storage capabilities. Examples of these are machine learning and deep learning networks. Many computational methods lie at the nexus of mathematical, statistical, and computational disciplines. Statistical methods often employ approaches that glean predictive capability from diverse and enormous databases of information. Emerging complex computational methods can provide impressive prediction models. However, it is unclear how widely these methods are applied in different medical domains [10,11]. This Special Issue includes four articles that focus on these predictive methods.

It is difficult to predict a patient's outcome with serial data that is collected irregularly, including medications, treatments, and laboratory tests. Typical deep learning methods can be used to analyze serial data. However, they must be improved to handle irregularly sampled serial datasets. In their study, Park and colleagues [12] investigate the accuracy of the phased long-term short-term memory (phased-LSTM) deep learning method in the prediction of patients with prostate cancer who might have castration-resistant prostate cancer (CRPC). The authors found that the phased-LSTM model was able to predict the CRPC outcome with 91.6% and 96.9% using 120 and 360 days of data, respectively.

The paper *A Comparison of Deep Learning Methods for ICD Coding of Clinical Records* authored by Moons and colleagues [13] presents a survey of various deep learning methods for text classification in a hierarchical framework for the domain of medical documents. Methods based on exploiting the taxonomy structure and also flat methods are discussed. These methods are evaluated on publicly available datasets corresponding to ICD-9 and ICD-10 coding, respectively.

In their contribution, de la Torre and co-authors [14] demonstrate the particularities of applying machine learning techniques in the field of healthcare. They focus on cervical assessment, where the goal is to predict the potential presence of cervical pain in patients affected with whiplash diseases. Using a sample of 302 patients, they compared several predictive models, including logistic regression, support vector machines, k-nearest neighbors, gradient boosting, decision trees, random forest, and neural network algorithms.

Afnan M. Alhassan and Wan Mohd Nazmee Wan Zainon [15] present in their article *Taylor Bird Swarm Algorithm Based on Deep Belief Network for Heart Disease Diagnosis* an approach to classify medical data for medical decision making. The method uses a feature selection step, where a sparse Fuzzy-c-mean (FCM) approach is used to select the significant features. Then, the selected features are passed into a deep belief network, which is trained using the Taylor-based bird swarm algorithm. The result of the analysis shows that the method is a promising approach.

5. Medical Informatics

Medical informatics focuses on the information technology that enables the effective collection of data using technology tools to develop medical knowledge and to facilitate the delivery of patient medical care [16]. The goal of medical informatics is to ensure access to critical patient medical information at the precise time and place it is needed to make medical decisions. Medical informatics also focuses on the management of medical data for research and education. Three papers in this Special Issue present applications for clinical decision making.

Daniel Clavel and his co-authors [17] present a decision support system to organize and order possible surgeries. Their study has the potential to reduce the workload of the healthcare system in scheduling—which is very labor-intensive work. A heuristic algorithm is proposed and included in the decision support system. Different features are implemented in a software tool with a friendly user interface. A simulation comparison of the scheduling obtained using the approach presented in this

paper and other similar approaches is shown and analyzed. In addition, the impact of the software tool on the efficiency and quality of surgical services is studied in one hospital setting.

In their paper, *A Prostate MRI Segmentation Tool Based on Active Contour Models Using a Gradient Vector Flow* [18], Joaquín Rodríguez, Gilberto Ochoa-Ruiz, and Christian Mata describe in a fully and detailed way a new GUI tool based on a semi-automated prostate segmentation. The purpose is to facilitate the time-consuming segmentation process used for annotating images in clinical practice. To support the efficiency of their method, the authors describe an experimental case.

The paper entitled *Medical Assistant Mobile Application for Diabetes Control by Simulating a Compartmental Model* authored by Hernandez-Ordonez and his coworkers [19] is very interesting and innovative. The authors present an application for mobile phones to assistant patients with type 1 diabetes. The proposed application is based on four mathematical models that describe glucose–insulin–glucagon dynamics using a compartmental model, with additional equations to reproduce aerobic exercise, gastric glucose absorption by the gut, and subcutaneous insulin absorption. Such developments are always welcome since diabetes became a civilization disease that affects a number of people every year.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Nieminen, P.; Miettunen, J.; Koponen, H.; Isohanni, M. Statistical methodologies in psychopharmacology: A review. *Hum. Psychopharmacol. Exp.* **2006**, *21*, 195–203. [CrossRef]
2. Caliebe, A.; Leverkus, F.; Antes, G.; Krawczak, M. Does big data require a methodological change in medical research? *BMC Med. Res. Methodol.* **2019**, *19*, 125. [CrossRef]
3. Indrayan, A. Reporting of Basic Statistical Methods in Biomedical Journals: Improved SAMPL Guidelines. *Indian Pediatr.* **2020**, *57*, 43–48. [CrossRef]
4. Hammouri, H.M.; Sabo, R.T.; Alsaadawi, R.; Kheirallah, K.A. Handling Skewed Data: A Comparison of Two Popular Methods. *Appl. Sci.* **2020**, *10*, 6247. [CrossRef]
5. Nieminen, P. Ten points for high-quality statistical reporting and data presentation. *Appl. Sci.* **2020**, *10*, 3885. [CrossRef]
6. Weon, B.M. Stretched exponential survival analysis for South Korean females. *Appl. Sci.* **2019**, *9*, 4230. [CrossRef]
7. Chen, S.H.; Pai, F.Y.; Yeh, T.M. Using the importance-satisfaction model and service quality performance matrix to improve long-term care service quality in Taiwan. *Appl. Sci.* **2020**, *10*, 85. [CrossRef]
8. Lei, Z.; Fasanella, L.; Martignetti, L.; Li-Jessen, N.Y.K.; Mongeau, L. Investigation of vocal fatigue using a dose-based vocal loading task. *Appl. Sci.* **2020**, *10*, 1192. [CrossRef] [PubMed]
9. Bielińska, A.; Bielińska-Waz, D.; Waz, P. Classification maps in studies on the retirement threshold. *Appl. Sci.* **2020**, *10*, 1282. [CrossRef]
10. Nieminen, P.; Kaur, J. Reporting of data analysis methods in psychiatric journals: Trends from 1996 to 2018. *Int. J. Methods Psychiatr. Res.* **2019**, *28*, e1784. [CrossRef] [PubMed]
11. Nieminen, P.; Vähänikkilä, H. Use of data analysis methods in dental publications: Is there evidence of a methodological change? *Publications* **2020**, *8*, 9. [CrossRef]
12. Park, J.; Rho, M.J.; Moon, H.W.; Lee, J.Y. Castration-resistant prostate cancer outcome prediction using phased long short-term memory with irregularly sampled serial data. *Appl. Sci.* **2020**, *10*, 2000. [CrossRef]
13. Moons, E.; Khanna, A.; Akkasi, A.; Moens, M.F. Article a comparison of deep learning methods for ICD coding of clinical records. *Appl. Sci.* **2020**, *10*, 5262. [CrossRef]

14. De la Torre, J.; Marin, J.; Ilarri, S.; Marin, J.J. Applying machine learning for healthcare: A case study on cervical pain assessment with motion capture. *Appl. Sci.* **2020**, *10*, 5942. [CrossRef]

15. Alhassan, A.M.; Wan Zainon, W.M.N. Taylor Bird Swarm Algorithm Based on Deep Belief Network for Heart Disease Diagnosis. *Appl. Sci.* **2020**, *10*, 6626. [CrossRef]

16. Melton, B.L. Systematic Review of Medical Informatics–Supported Medication Decision Making. *Biomed. Inform. Insights* **2017**, *9*, 117822261769797. [CrossRef] [PubMed]

17. Clavel, D.; Mahulea, C.; Albareda, J.; Silva, M. A decision support system for elective surgery scheduling under uncertain durations. *Appl. Sci.* **2020**, *10*, 1937. [CrossRef]

18. Rodríguez, J.; Ochoa-Ruiz, G.; Mata, C. A Prostate MRI Segmentation Tool Based on Active Contour Models Using a Gradient Vector Flow. *Appl. Sci.* **2020**, *10*, 6163. [CrossRef]

19. Hernández-Ordoñez, M.; Nuño-Maganda, M.A.; Calles-Arriaga, C.A.; Rodríguez-Leon, A.; Ovando-Chacon, G.E.; Salazar-Hernández, R.; Montaño-Rivas, O.; Canseco-Cortinas, J.M. Medical Assistant Mobile Application for Diabetes Control by Simulating a Compartmental Model. *Appl. Sci.* **2020**, *10*, 6846. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.