European Journal of Therapeutics DOI: 10.5152/EurJTher.2018.1012

Are Lung Cancer Publications Up-to-Date in terms of Advances in Statistics and Bioinformatics?

Seval Kul[1], İlkay Doğan[2]
[1]Department of Biostatistics, Gaziantep University School of Medicine, Turkey
[2]Department of Biostatistics, Faculty of Veterinary, Afyon Kocatepe University, Turkey

ABSTRACT
This study was performed to evaluate whether literature of lung cancer follow advances in statistics and bioinformatics. Four medical journals with high impact factors were reviewed between January 2013 and December 2017. Among 1649 published manuscript, 514 of them were about lung cancer. Also, Medline was searched with key words combinations of e-learning AND education AND cancer AND patient for last 5 years. New statistical methods weren't applied in the cancer researches performed by clinicians. Furthermore, unlike increasing number of successful studies using internet and computer technologies, number of the study is limited. Working with professional statisticians or collaboration to Biostatisticians will increase the quality of lung cancer papers.
Keywords: Regression analysis, statistical methods, e-learning

INTRODUCTION

Statistics is an essential component of medical research from design to reporting, data collection to analysis and interpretation of data (1). Editors of medical journals want to ensure the quality and accuracy of the statistical methods of the papers. Standards of a manuscript were determined by international committee of medical journal editors and several checklist such as strobe for STROBE checklist for observational cohort, case control and cross sectional studies and CONSORT checklist for randomized controlled trials are available to identify basic requirements of a report (2, 3). All of the checklists have a special section for standards of the statistical methods used in the report. Because of the trend of improving quality of papers, researchers pay more attention to statistical analysis part. Some studies were performed to investigate how accurate statistical analysis are (4-6). But there is no study to show how up-to-date statistical analysis used in medical literature. Parallel to medical research, medical statistics is also improving (7). This study was performed to evaluate whether literature of lung cancer follow advances in statistics and bioinformatics.

CLINICAL AND RESEARCH CONSEQUENCES

First, Medical journals with high impact factors namely; The Lancet Oncology, The Annals of Thoracic Surgery, The Journal of Thoracic and Cardiovascular Surgery and European Journal of Cardio-Thoracic Surgery were reviewed between January 2013 and December 2017. Among 1649 published manuscript, 514 of them were about lung cancer. All of the published lung cancer manuscripts were classified in terms of statistical method used. Second, lung cancer word was searched in one of the most popular biostatistics journal, Statistics in Medicine, to review recent statistical methods were introduced for lung cancer research questions and applied to real lung cancer data. The last Medline was searched with key words combinations of e-learning AND education AND cancer AND patient for last 5 years.

Frequency of the statistical method used in the same year and overall for 5 years were given in Table 1 and Figure 1. Kaplan-Meier method was the most commonly used method to estimate survival analysis with 28.71%. Frequencies of using the method were relatively similar across the years. Chi-square test was the second most frequently used method to show relationship between categorical variables. Student t/ Mann Whitney u test and one way ANOVA/Kruskal Wallis tests relatively lost their popularity in 2017.

Regression methods including hazard, logistic and linear regression were still not frequently used methods. Area under the roc curve and ROC curve were rarely used statistical methods. Furthermore power analysis was only reported by 5% of the published study.

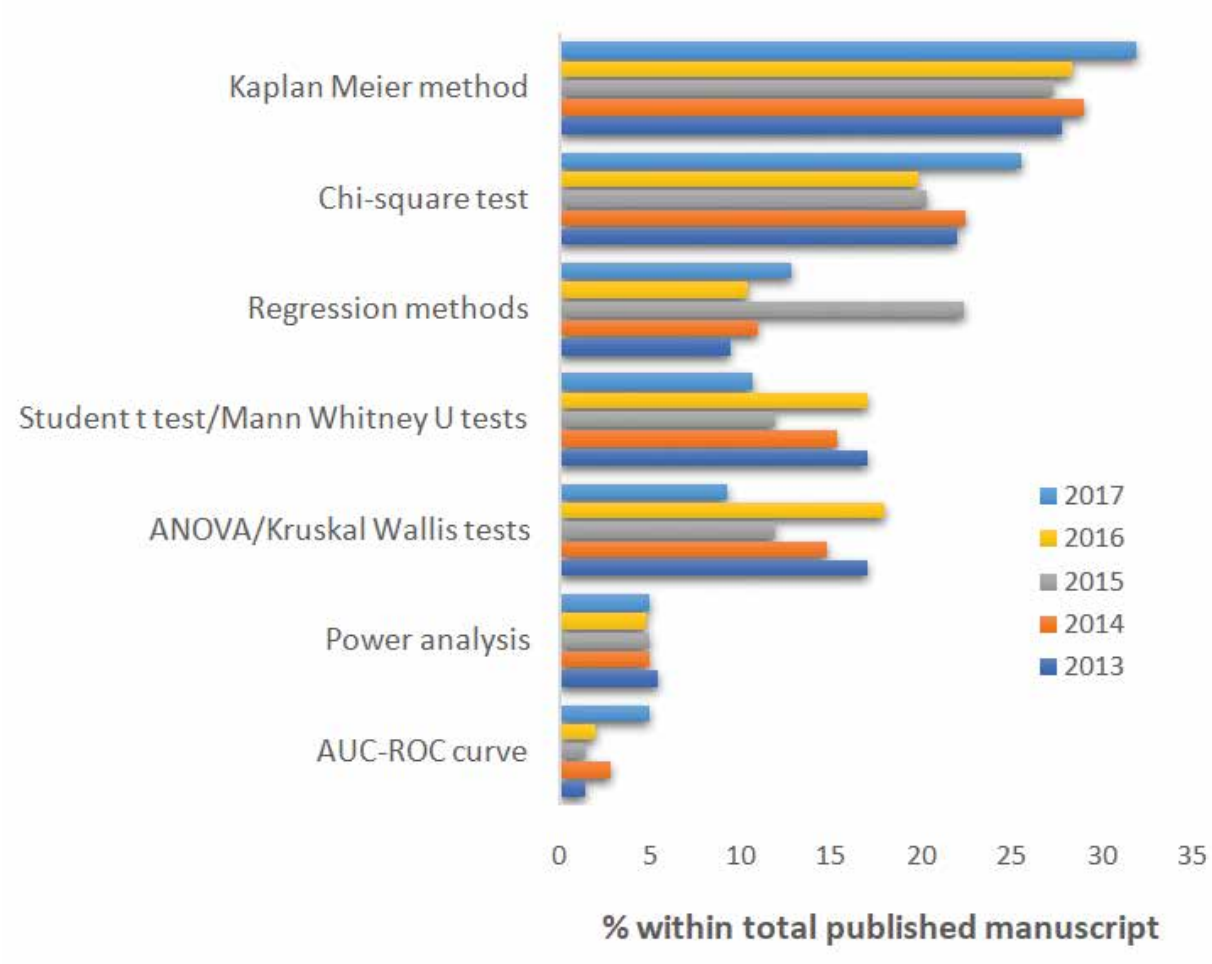
Starting from the design issue, determining the minimum sample size for a study convinces an adequate power to detect statistical significance and consequently, it is a critical step in the design of a lung cancer research (8). Among the published studies, only approximately 5% of the studies reported their power analysis which is very low. Additionally, majority of statistical method applied in the published studies was univariate analysis (87.47%). Considering applied statistical methods, it can be concluded that in lung cancer studies complex rela-

Corresponding Author: Seval Kul **E-mail:** sevalkul@gantep.edu.tr
Received: 11.03.2018 • **Accepted:** 26.04.2018

Table 1. Frequency (%) of the statistical method used in the same year and overall for 5 years

	Year of the publication					
Statistical Methods	**2013**	**2014**	**2015**	**2016**	**2017**	**Overall (%)**
Kaplan Meier method	27.80	28.96	27.27	28.30	31.91	28.71
Chi–square test	21.97	22.40	20.28	19.81	25.53	21.84
Student t test/Mann Whitney U tests	17.04	15.30	11.89	16.98	10.64	14.86
One way ANOVA/Kruskal Wallis tests	17.04	14.75	11.89	17.92	9.22	14.75
Regression methods	9.42	10.93	22.38	10.38	12.77	12.53
AUC–ROC curve	1.35	2.73	1.40	1.89	4.96	2.32
Power analysis	5.38	4.92	4.90	4.72	4.96	4.99

Figure 1. Frequency (%) of the statistical method used in the same year



tionships were not investigated enough. Main target of most of the publication was to estimate mean or median overall survival and risk factors affecting survival time. Kaplan Meier method is the most popular method for estimating mean or median time from censored and non-censored data (9). A competing risk is an event whose occurrence stops the occurrence of the primary event of interest (10). But Kaplan Meier method doesn't take into consideration of confounders and competing risks which is a very common situation for cancer studies. Several practical methods for competing risks analysis were mentioned in the study of Bakoyannis and Touloumi (11). It is known that especially in the observational and retrospective studies, confounding factors should be eliminated from studies (12). In other words, results of causal relationships should

be adjusted by possible confounding variables to eliminate the bias (13). Besides, e-learning practices or use of mobile technologies become very popular in medical research to support health professionals and patients (14-17). We believe, another problem about the lung cancer studies is rarely using this recent bioinformatics technologies in lung cancer research. In the following part we will review recent improvements in biostatistics and bioinformatics.

Recent Advances in Biostatistics

There are several statistical research journals publish papers about new statistical methods with real data applications. For this study we reviewed Statistics in medicine, which is one of well-known biostatistics journal, with key word of lung cancer for last 5 years. Several papers introduced novel and advance statistical methods with application of real data. To identify genetic markers associated with the prognosis of lung cancer Wu et al. (18) advised a penalized robust semiparametric approach for gene-environment interactions. Furthermore, Wu et al. (19) also showed effectiveness of penalized robust approach to estimate the association between lung cancer prognosis with gene expression measurements and clinical covariates). Schipper et al. (20) used a dataset of lung cancer patients treated with radiation therapy and applied a special statistical model for toxicity and efficacy with dose and biomarkers as covariates. Receiver operating characteristic (ROC) curve analysis is used to determine the optimal cut off values for a numerical variable and to investigate diagnostic value of a continuous medical test (21). But this method is usually used for classification of two categories. Wang et al. (22) proposed methods for classification of 3 or more categories and in the applications a microarray data set for lung cancer was used. Branscum et al. (23) developed flexible regression model for evaluating the accuracy of a continuous medical test or biomarker with or without a gold standard. Gasparini et al. (24) modelled the relationship between occupational exposure to radon with distributed lag non-linear models and lung cancer mortality by using the data from the Colorado Plateau miner's cohort.

Recent Advances in Bioinformatics

Milne et al. (25) conducted a cross sectional study to determinate level of eHealth Literacy in primary lung cancer survivors. They showed 78% of the survivor had access to eResources via computer, Internet, or smartphone. Because of the increasing number of smartphone users and internet users, E-learning and mobile technologies have become recent issue to inform and support patients, update doctors and health professions knowledge, (15, 16, 26-28). In Medline, several bioinformatic studies with successful results were available. For example; in some studies web based support and decision-making systems were used for clinical decision. Masood et al. (29) proposed a Computer-Assisted Decision Support System in Pulmonary Cancer detection by using the learning based. Murgu et al. (30) designed interactive a program (GAIN 3.0) to enhance interdisciplinary collaboration for effective Non-small Cell Lung Cancer diagnosis, assessment, and treatment. And the program improved participants' knowledge, competence, and likely the clinical care provided to patients. In the study performed by Basch et al. (31) tablet computers in clinic waiting areas were given to patients and reporting of adverse events at 6 time points was asked. DuBenske et al. (32) introduced a Web-based lung cancer information, communication, and coaching system for caregivers (family members of patients). Lower burden and negative mood were observed among caregivers who joined the eHealth intervention.

CONCLUSION

New statistical methods weren't applied in the cancer researches performed by clinicians. Working with professional statisticians or collaboration to Biostatisticians will increase the quality of papers. Furthermore, unlike increasing number of successful studies using internet and computer technologies, number of the study is limited.

Author Contributions: Concept - S.K, İ.D.; Design - S.K, İ.D.; Supervision - S.K, İ.D.; Resource - S.K, İ.D.; Materials - S.K, İ.D.; Data Collection and/or Processing - S.K, İ.D.; Analysis and/or Interpretation - S.K, İ.D.; Literature Search - S.K, İ.D.; Writing - S.K, İ.D.; Critical Reviews - S.K, İ.D.

Peer-review: Internally reviewed.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The author declared that this study has received no financial support.

REFERENCES

1. Swinscow TD. Statistics at Square One. IV. Variation between samples. Br Med J 1976; 1: 1585. [CrossRef]
2. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007; 4: e297. [CrossRef]
3. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Consort Group. Consort 2010 statement: extension to cluster randomized trials. BMJ 2012; 345: e5661. [CrossRef]
4. Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research--a review of common pitfalls. Swiss Med Wkly 2007; 137: 44-9.
5. Parsons NR, Price CL, Hiskens R, Achten J, Costa ML. An evaluation of the quality of statistical design and analysis of published medical research: results from a systematic survey of general orthopedic journals. BMC Med Res Methodol 2012; 12: 60. [CrossRef]
6. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. Biochem Med (Zagreb) 2015; 25: 5-11. [CrossRef]
7. Mahapatra D, Agarwal K, Khosrowabadi R, Prasad DK. Recent Advances in Statistical Data and Signal Analysis: Application to Real World Diagnostics from Medical and Biological Signals. Comput Math Methods Med 2016; 2016: 1643687. [CrossRef]
8. Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci 2012; 5: 7-13. [CrossRef]
9. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ 1998; 317: 1572-80. [CrossRef]
10. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation 2016; 133: 601-9. [CrossRef]
11. Bakoyannis G, Touloumi G. Practical methods for competing risks data: a review. Stat Medhods Med Res 2012; 21: 257-72. [CrossRef]

12. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013; 158: 280-6. [CrossRef]
13. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. Evid Based Spine Care J 2012; 3: 9-12. [CrossRef]
14. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med 2006; 81: 207-12. [CrossRef]
15. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008; 300: 1181-96. [CrossRef]
16. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007; 9: e2. [CrossRef]
17. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015; 17: e52. [CrossRef]
18. Wu C, Shi X, Cui Y, Ma S. A penalized robust semiparametric approach for gene–environment interactions. Statistics in medicine. 2015; 34: 4016-30. [CrossRef]
19. Wu C, Jiang Y, Ren J, Cui Y, Ma S. Dissecting gene-environment interactions: A penalized robust approach accounting for hierarchical structures. Stat Med 2018; 37: 437-56. [CrossRef]
20. Schipper MJ, Taylor JM, TenHaken R, Matuzak MM, Kong FM, Lawrence TS. Personalized dose selection in radiation therapy using statistical models for toxicity and efficacy with dose and biomarkers as covariates. Stat Med 2014; 33: 5330-9. [CrossRef]
21. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013; 4: 627-35.
22. Wang D, Attwood K, Tian L. Receiver operating characteristic analysis under tree orderings of disease classes. Stat Med 2016; 35: 1907-26. [CrossRef]
23. Branscum AJ, Johnson WO, Hanson TE, Baron AT. Flexible regression models for ROC and risk analysis, with or without a gold standard. Stat Med 2015; 34: 3997-4015. [CrossRef]
24. Gasparrini A. Modeling exposure–lag–response associations with distributed lag non-linear models. Stat Med 2014; 33: 881-99. [CrossRef]
25. Milne RA, Puts MT, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of High eHealth Literacy in Primary Lung Cancer Survivors. J Cancer Educ 2015; 30: 685-92. [CrossRef]
26. Corbeil JR, Corbeil ME. Are we ready for mobile learning now? 2007 Mobile learning predictions revisited. Issues Inform Syst 2011; 12: 142-52.
27. Schilling K, Wiecha J, Polineni D, Khalil S. An interactive web-based curriculum on evidence-based medicine: design and effectiveness. Fam Med 2006; 38: 126-32.
28. Lai CY, Wu CC. Promoting Nursing Students' Clinical Learning Through a Mobile e-Portfolio. Comput Informa Nurs 2016; 34: 535-43. [CrossRef]
29. Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, et al. Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. J Biomed Inform 2018, 117-28. [CrossRef]
30. Murgu S, Rabito R, Lasko G, Jackson C, Mino-Kenudson M, Ettinger DS, et al. Impact of a Non-small Cell Lung Cancer Educational Program for Interdisciplinary Teams. Chest 2018; 153: 876-87. [CrossRef]
31. Basch E, Pugh SL, Dueck AC, Mitchell SA, Berk L, Fogh S, et al. Feasibility of Patient Reporting of Symptomatic Adverse Events via the Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) in a Chemoradiotherapy Cooperative Group Multicenter Clinical Trial. Int J Radiat Oncol Biol Phys 2017; 98: 409-18. [CrossRef]
32. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010; 30: 732-44. [CrossRef]

How to cite:

Kul S, Doğan İ. Are Lung Cancer Publications Up–To–Date in terms of Advances in Statistics and Bioinformatics? Eur J Ther 2018; 24(Suppl 1); S57–S60.