Published in final edited form as:

Ophthalmology. 2014 July ; 121(7): 1317–1321. doi:10.1016/j.opththa.2014.01.015.

Use of Statistical Analyses in the Ophthalmic Literature

Renato Lisboa, M.D.^{1,2}, Daniel Meira-Freitas, M.D., Ph.D.^{1,2}, Andrew J. Tatham, FRCOphth¹, Amir H. Marvasti, B.S.^{1,3}, Lucie Sharpsten, Ph.D.¹, and Felipe A. Medeiros, M.D., Ph.D.¹

¹Hamilton Glaucoma Center and Department of Ophthalmology, University of California, San Diego

²Federal University of São Paulo, Department of Ophthalmology, Brazil

³Boston University School of Medicine, Boston, MA, USA

Abstract

Purpose—To identify the most commonly used statistical analyses in the ophthalmic literature and to determine the likely gain in comprehension of the literature that readers could expect if they were to sequentially add knowledge of more advanced techniques to their statistical repertoire.

Design—Cross-sectional study

Methods—All articles published from January 2012 to December 2012 in *Ophthalmology*, *American Journal of Ophthalmology* and *Archives of Ophthalmology* were reviewed. A total of 780 peer-reviewed articles were included. Two reviewers examined each article and assigned categories to each one depending on the type of statistical analyses used. Discrepancies between reviewers were resolved by consensus.

Main Outcome Measures—Total number and percentage of articles containing each category of statistical analysis were obtained. Additionally we estimated the accumulated number and percentage of articles that a reader would be expected to be able to interpret depending on their statistical repertoire.

Results—Readers with little or no statistical knowledge would be expected to be able to interpret the statistical methods presented in only 20.8% of articles. In order to understand more than half (51.4%) of the articles published, readers were expected to be familiar with at least 15 different statistical methods. Knowledge of 21 categories of statistical methods was necessary to comprehend 70.9% of articles, while knowledge of more than 29 categories was necessary to

© 2014 American Academy of Ophthalmology, Inc. Published by Elsevier Inc. All rights reserved.

Corresponding author: Felipe A. Medeiros, M.D., Ph.D., Hamilton Glaucoma Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, fmedeiros@glaucoma.ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Financial Disclosure(s):

The author(s) have made the following disclosure(s):

Renato Lisboa – None; Daniel Meira-Freitas – None; Andrew J. Tatham – None; Amir H. Marvasti – None; Lucie Sharpsten – None; Felipe A. Medeiros – Research support – Carl-Zeiss Meditec, Inc; Heidelberg Engineering, GmbH; Alcon; Allergan; Topcon; Reichert, Inc.

comprehend more than 90% of articles. Articles in retina and glaucoma subspecialties showed a tendency for using more complex analysis when compared to cornea.

Conclusions—Readers of clinical journals in ophthalmology need to have substantial knowledge of statistical methodology to understand the results of published studies in the literature. The frequency of use of complex statistical analyses also indicates that those involved in the editorial peer-review process must have sound statistical knowledge in order to critically appraise articles submitted for publication. The results of this study could provide guidance to direct the statistical learning of clinical ophthalmologists, researchers and educators involved in the design of courses for residents and medical students.

INTRODUCTION

Statistical analyses are an essential component of medical research, which when appropriately applied can further our understanding of risk factors, treatment effects and other aspects of disease. Although it is arguable that one may not have to fully understand all the statistical analyses presented in a research article in order to grasp the article's main message, a basic understanding of statistics is important to allow critical appraisal of the presented material. Clinicians should be able to evaluate the medical literature in order to make informed decisions regarding best practice, a critical component of evidence-based medicine. Knowledge of statistics is also important for those involved in the peer-review process of a research article, as they will be responsible for judging the adequacy and correctness of the statistical methods used. In fact, readers often rely on the “approval stamp” given to articles that have undergone peer-review by a reputable journal. However, this process may have imperfections, such as those originating from the lack of statistical knowledge of the peer-reviewers.

Medical statistics is a growing and increasingly complex discipline, which like all medical specialties is becoming increasingly subspecialized. For many years, basic statistical tests such as the t-test and similar non-parametric tests have been used to test null hypotheses and address the main research question. These tests, which were first described during the early 20th century,^{1,2} are widely cited in the medical literature, and are therefore familiar to most clinicians. Modern research, however, often generates large quantities of data, and hypotheses are posed that may require more sophisticated analyses. For example, analyses often need to factor for multiple potentially confounding variables, repeated measures and inclusion of both eyes of study subjects.^{3–5} By using advanced statistical methods, there is the opportunity to answer questions that might previously have been difficult to address. Recent examples include questions such as: “What are the risk factors for conversion to neovascular age-related macular degeneration?”⁶ and “How to combine different structural and functional tests for assessment of glaucomatous progression?”⁷ However, although advanced statistical analyses have allowed researchers to address previously unsolved issues, its frequent use has made it increasingly difficult to evaluate the published literature.

Although complex statistical methods may be implemented using statistical software packages, the frequency of use of these methods in ophthalmic research is not known. A better knowledge of the frequency of use of different statistical methods in the ophthalmic

literature may provide guidance to the development of training programs aimed at increasing the statistical knowledge of medical students, residents, trainee fellows, clinicians and researchers. Therefore, the purpose of this article was to review the ophthalmic literature to ascertain the most frequently used statistical methods. We also aimed to determine the likely gain in comprehensive understanding of the ophthalmic literature that a reader could expect if he or she were to sequentially add knowledge of more advanced techniques to his or hers statistical repertoire.

METHODS

Three high impact clinically oriented ophthalmology journals were selected for inclusion in this study: Ophthalmology, The American Journal of Ophthalmology and Archives of Ophthalmology (now Journal of the American Medical Association (JAMA) Ophthalmology). All articles published in the print versions of the three journals during the twelve-month period from January 2012 to December 2012 were included. There were a total of 780 peer-reviewed articles published during this period, including 345 articles in Ophthalmology, 264 articles in the American Journal of Ophthalmology and 171 articles in Archives of Ophthalmology. Because this study did not involve human subjects, human derived materials, or human records, Institutional Review Board approval was unnecessary.

Two reviewers with experience in medical statistics examined the full content of each article and assigned categories to each one depending on the type of statistical analyses used. The categorization scheme was similar to that described by Emerson and Colditz,⁸ however, some modifications were required to incorporate more recently introduced statistical methods (Table 1). Additional methods included in our categorization included kappa for agreement, Bland-Altman analysis, univariable logistic regression, multivariable logistic regression, analysis of covariance (ANCOVA), generalized estimated equations, linear mixed models and Bayesian analysis. Any discrepancy between the reviewers was resolved by consensus.

Categories of Statistical Methods

A total of 34 categories of statistical methods were identified and used in the analysis (Table 1). Although most of these categories have been previously described¹ and do not require further explanation, a brief description of selected categories follows.

The first category identified articles that presented only percentages, means, medians, standard deviations and histograms. This was the only category that was exhaustive. In other words, this category applied only to articles that included no other category of statistical analysis. All other statistical categories could be present in an article even when one or more other categories were previously identified. For example, an article that applied both Mann-Whitney test and Bonferroni correction was categorized as including both a nonparametric test and a test of multiple comparisons.

The category “contingency tables” referred to articles that reported odds ratio, chi-squared test, Fisher’s exact test, McNemar test and Mantel-Haenszel test. The category “epidemiologic statistics” identified articles that reported prevalence or incidence. When

these measurements were adjusted for covariates, such as age, the category “adjustment for epidemiologic statistic” was also marked. “Diagnostic proportions” identified articles that reported sensitivity, specificity or diagnostic likelihood ratios. “Multiple comparisons” included any method that adjusted significance levels or P values when several comparisons were made. The “transformation” category identified the examination or analysis of data using a scale other than that of the raw data (e.g., logarithmic). The category “cost-benefit analysis” referred to articles that used the process of combining estimates of cost and health outcomes to compare policy alternatives. Studies categorized as including a “sensitivity analysis” examined sensitivity of outcome to small changes in parameters of a model or in other assumptions. Articles that used data containing multiple measures on the same dependent variable were included in the “repeated measures analysis” category (e.g. time-series data and panel data). “Missing-data methods” referred to articles that performed any type of data imputation due to loss of information during the data collection process. “Non-inferiority trials” intended to show that the effect of a new treatment is not worse than an established treatment by a determined margin. The “resampling” category refers to any method used to estimate the precision of the sample by using subsets of the data (i.e., jackknife), drawing a random sample with replacement from a set of data points (i.e., bootstrapping) or validating a predictive model by fitting this model in a subset of the data that was held out during the construction of the model (i.e., cross-validation). “Generalized estimating equations” is a technique used to estimate the parameters of a linear model when an unknown correlation in the observations is present. This method is frequently applied in ophthalmology to adjust for the inclusion of both eyes of the same patient in the study. The category “linear mixed models” refers to models that combine both fixed effects and random effects. “Bayesian analysis” is a statistical procedure that estimates the parameter of an underlying distribution based on the observed distribution. “Rasch analysis” is usually used to analyze psychometric categorical data, such as answers to questionnaires. It is a function of the trade-off between the respondent’s abilities, attitudes or personal items and the item difficulty. The last category (“other methods”) included methods not adequately described by the other categories.

Study Analysis

The total number and percentage of articles containing each category of statistical method was calculated. Additionally we estimated the accumulated number and percentage of articles that a reader would be expected to interpret depending on their statistical repertoire. To achieve this we ranked the statistical methods in a hierarchy from simplest to most complex. The analysis of accumulated knowledge assumed that if a reader understood the more complex tests, he or she would also have an understanding of the simpler analyses. Although the sequence of tests followed a logical order, it was nevertheless arbitrary.

As articles pertaining to different subspecialties in ophthalmology might differ in the frequency of use of different statistical tests, we also conducted an analysis of accumulated knowledge by subspecialty. The following subspecialties were identified: cornea and external diseases, glaucoma, retina, cataract, strabismus and pediatrics, orbit and oculoplastics, refractive surgery, ocular tumors, uveitis and comprehensive ophthalmology.

Separate analyses were conducted for the three subspecialties with the largest number of articles in the evaluated period, namely cornea and external diseases, glaucoma and retina.

RESULTS

Table 1 summarizes the frequency of statistical tests found in the 780 peer-reviewed articles. One hundred and sixty-two of 780 articles (20.8%) reported no statistical methods or used descriptive statistics only. Therefore readers with little or no statistical knowledge would be expected to be able to understand 20.8% of articles published in *Ophthalmology*, *American Journal of Ophthalmology* and *Archives of Ophthalmology* during 2012.

The next simplest statistical tests were t-tests, contingency tables and non-parametric tests, which were present in 246, 266 and 170 articles respectively. If a reader had sufficient knowledge of statistics to understand non-parametric tests our analysis presumed they would also understand the simpler statistics presented earlier in the table (i.e., no statistical methods, t-tests and contingency tables). Therefore a reader with an understanding of non-parametric tests would be able to comprehend 266 or 34.1% of articles in the 3 journals in 2012. The same reader would have to know the first 15 categories of statistical methods listed on Table 1 to understand more than half (51.4%) of articles published in the same journals during 2012. Knowledge of 21 categories of statistical methods was necessary to comprehend 70.9% of articles, while knowledge of more than 29 categories was necessary to comprehend more than 90% of articles.

The contingency table was the most frequently used category of statistical analysis, present in 266 of 780 articles (34.1%). The least frequent category was Rasch analysis, present in only 3 of 780 articles (0.4%). Only 26 of 780 studies (2.7%) included the “other methods” category of statistical tests. Examples included studies by Rudnicka and colleagues who used I^2 statistics to quantify the extent of heterogeneity in age related macular degeneration prevalence estimates,⁹ and Ramulu and colleagues who used negative binomial regression to analyze accelerometer data in glaucoma.¹⁰

When analyzed by subspecialties articles could be divided as follows: 133 were related to cornea and external diseases, 111 to glaucoma, 228 to retina, 40 to cataract, 44 to strabismus and pediatrics, 14 to orbit and oculoplastics, 12 to refractive surgery, 46 to ocular oncology, 35 to uveitis and 117 to comprehensive ophthalmology. The cumulated ability to understand glaucoma, retina, and cornea and external disease articles is shown in Figure 1. The results indicated a tendency for retina and glaucoma to use more complex statistical techniques compared to cornea, however knowledge of 20 or more statistical techniques would allow one to fully comprehend the analyses of 80% of cornea, retina, and glaucoma-related articles.

DISCUSSION

This study identified the number and categories of statistical tests used in papers published in 3 high-impact ophthalmology journals during 2012. We found that statistical tests were used in most peer-reviewed articles, with 618 of 780 articles (79.2%) reporting some form of statistical analysis beyond descriptive statistics. The most frequently used analyses were

contingency tables (34.1%), t-tests (31.5%), and nonparametric tests (21.8%). Only three other statistical methods were present in greater than 10% of the papers (ANOVA, multiple logistic regression and survival methods).

The proportion of studies that reported using at least one statistical method (almost 80%) was considerably larger than the proportion found in a similar analysis conducted 30 years ago. In 1983 Emerson and Colditz found only 42% of articles published in *The New England Journal of Medicine* included the use of a statistical method.⁸ Although Emerson and Colditz examined articles from a general medical journal it is reasonable to assume that there was a similarly low use of statistics in the ophthalmology literature at the time.

These results indicate that there has been a large increase in the use of statistical analyses during the last 3 decades. These changes are likely to be due, at least in part, to increased volume of medical research data and increased awareness of the importance of appropriate statistical analyses. Moreover, greater access to computers and commercially available statistical software has no doubt facilitated such analyses. The growing emphasis on evidence-based medicine has also heightened the importance placed on appropriate statistics. In 1983, medical knowledge and education depended to a large degree on textbooks. In contrast, today physicians can easily access the latest scientific journals and other resources from personal computers to keep their knowledge up-to-date. The Internet has also substantially impacted evidence-based medicine by disseminating knowledge and enabling the learning of new statistical methods.

The results of this study also demonstrate that knowledge of a large number of statistical methods is necessary in order to fully interpret the results of studies in the ophthalmic literature. From Table 1 it is apparent that knowledge of no single statistical method would substantially increase the ability of the reader to understand a greater number of articles. Rather an accumulated knowledge of many statistical tests is necessary. From Table 1 it is possible to determine the number of statistical methods one must learn in order to acquire a reasonable statistical knowledge. One might assume that knowledge of 5 or 6 statistical methods would be sufficient to comprehend a large proportion of published papers. However, we found that a reader with knowledge of 15 methods would comprehend the analyses of only half (51.4%) of the articles published during the study period. Knowledge of 29 methods was required to understand the analysis in 90% of articles.

The increasing frequency of the use of complex statistical analyses in the clinical ophthalmic research literature has several implications. First, it is clear that statisticians or researchers with sound statistical knowledge should be involved as early as possible in the research, actively participating not only in the data analysis step, but also in study planning and design. In addition, they have important roles in the subsequent steps of reporting results and manuscript preparation, providing input during the translation of research results into clinically relevant messages. Second, the increased frequency of use of complex statistical methods has the potential to make the peer-review process more difficult, by reducing the pool of reviewers qualified to critically appraise a particular article. In fact, the results from Table 1 seem to indicate that the majority of reviewers would actually lack the knowledge to review a large part of the articles submitted for publication. Therefore, the peer-review

process may have to align reviewers that can judge the clinical significance of the research with reviewers possessing significant statistical knowledge. The latter may include biostatisticians who are familiar with ophthalmic research or researchers with sound knowledge of statistics that allows them to make judgments about the adequacy and accuracy of the statistical methods used. Finally, the increased complexity of statistical methods used in ophthalmology research also affects clinicians and eye care practitioners who have to critically appraise these articles to make informed decisions regarding best practice.

The increased complexity of statistical analyses may discourage readers from trying to critically appraise the methodology and analyses of the study and just accept the conclusions as presented in the published article. In this situation, they rely on the belief that an article that has passed through peer-review will have been cleared of potential statistical deficiencies. However, this might actually not be the case due to the potential lack of qualified reviewers for judging the article. This tendency highlights the increased importance of the peer-review process in eliminating the deficient articles. However, as pointed out above, such task will require reviewers with enough knowledge to adequately judge the articles themselves. It is therefore important that editors consider the statistical analyses included in an article when selecting potential reviewers. The results presented in our study may help guide those involved in the peer-review process when deciding whether an article requires a more specialized review by statistical experts.

As different ophthalmic subspecialties may differ in the frequency of use of statistical procedures, we performed a stratified analysis of the data according to different subspecialties. Our results indicated that there was a tendency for articles related to glaucoma and retina to apply more complex statistical procedures than articles related to cornea and external diseases. As Figure 1 indicates, when we plotted the cumulative percentage of articles versus the complexity of statistical methods (from less to more complex), the curve representing articles related to cornea and external diseases tended to achieve a higher cumulative percentage at lower levels of complexity as compared to those curves of articles related to retina and glaucoma. It should be noted, however, that if the reader had cumulative knowledge of the first 20 tests presented in Table 1, he or she would be expected to fully understand the analyses presented in approximately 80% of articles for all these three subspecialties.

The present study has limitations. It does not account for the influence that important concepts such as study design or biases may have on the reader's statistical understanding. Moreover, the sequence of statistical methods used for our analysis was somewhat arbitrary and in practice individuals will have different routes to knowledge. Also, some of our results cannot be extrapolated to other medical fields, as the sequence of methods presented here may not be the most reasonable for other subspecialties. For example, Rasch analysis is likely to be a much more frequent type of analysis in psychology publications than in ophthalmology. Despite these limitations, the current study demonstrates the importance of appropriate statistical knowledge for readers of the ophthalmology literature.

In conclusion, this study has identified the most commonly used statistical methods in the ophthalmic literature and has provided information regarding the likely yield of each method, in terms of improved understanding of the published studies. The frequency of use of complex statistical analyses also indicates that those involved in the editorial peer-review process must have sound statistical knowledge in order to critically appraise articles submitted for publication. In addition, our results could form an important guide to direct the statistical learning of clinical ophthalmologists, researchers and educators involved in the design of courses for fellows, residents and medical students.

Acknowledgments

Supported in part by National Institutes of Health/National Eye Institute grant EY021818 (F.A.M.); Core Grant P30EY022589, Brazilian National Research Council-CNPq grant 200178/2012-1 (D.M.F).

References

1. Student. The probable error of a mean. *Biometrika*. 1908; 6:1–25.
2. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bull*. 1945; 1:80–83.
3. Kleinbaum, DG.; Kupper, LL.; Nizam, A.; Muller, KE. *Applied Regression Analysis and Other Multivariable Methods*. 4th ed. Belmont, CA: Thomson; 2008. p. 114-138.
4. Vittinghoff, E.; Glidden, DV.; Shiboski, SC.; McCulloch, CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models*. 1st ed. New York: Springer; 2010. p. 253-290.
5. Zhou, XH.; Obuchowski, NA.; McClish, DK. *Statistical Methods in Diagnostic Medicine*. New York: Wiley; 2002. p. 274-306.
6. Friberg TR, Bilonick RA, Brennen PM. Risk factors for conversion to neovascular age-related macular degeneration based on longitudinal morphologic and visual acuity data. *Ophthalmology*. 2012; 119:1432–1437. [PubMed: 22551740]
7. Medeiros FA, Zangwill LM, Gitkin CA, et al. Combining structural and functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol*. 2012; 153:1197–1205. [PubMed: 22317914]
8. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med*. 1983; 309:709–713. [PubMed: 6888443]
9. Rudnicka AR, Jarrar Z, Wormald R, et al. Age and gender variations in age-related macular degeneration prevalence in populations of European ancestry: a metaanalysis. *Ophthalmology*. 2012; 119:571–580. [PubMed: 22176800]
10. Ramulu PY, Maul E, Hochberg C, et al. Real-world assessment of physical activity in glaucoma using an accelerometer. *Ophthalmology*. 2012; 119:1159–1166. [PubMed: 22386950]

Knowledge of a variety of statistical methods is necessary to understand the results published in the ophthalmic literature. Our results may serve as a guide for educational programs in statistics among residents, clinicians and researchers.

Figure 1.

Plot of the cumulative percentage of articles that used different statistical methods according to the subspecialties of cornea and external diseases, glaucoma and retina. The complexity of statistical methods is represented in the horizontal axis going from 1 to 34 (from less to more complex). The curve representing articles related to cornea and external disease tended to achieve a higher cumulative percentage at lower levels of complexity as compared to those curves of articles related to retina and glaucoma.

Table 1

Statistical content of selected ophthalmic journals during the year of 2012.

Statistical method		Articles containing the method		Accumulation by article	
number	percent	number	percent		
0	No statistical methods or descriptive statistics only	162	20.8	162	20.8
1	t-Tests	246	31.5	186	23.8
2	Contingency tables	266	34.1	219	28.1
3	Nonparametric tests	170	21.8	266	34.1
4	Epidemiologic statistics	42	5.4	270	34.6
5	Adjustments for epidemiologic statistics	15	1.9	272	34.9
6	Diagnostic proportions	40	5.1	282	36.2
7	Multiple comparisons	49	6.3	289	37.1
8	Pearson's correlation	67	8.6	299	38.3
9	Spearman's correlation	49	6.3	320	41.0
10	Kappa for agreement	25	3.2	326	41.8
11	Bland-Altman	17	2.2	328	42.1
12	Analysis of variance (ANOVA)	99	12.7	357	45.8
13	Analysis of covariance (ANCOVA)	25	3.2	360	46.2
14	Transformation	51	6.5	380	48.7
15	Simple linear regression	64	8.2	401	51.4
16	Multiple linear regression	66	8.5	433	55.5
17	Multiway tables	6	0.8	435	55.8
18	Simple logistic regression	77	9.9	453	58.1
19	Multiple logistic regression	89	11.4	500	64.1
20	Survival methods	85	10.9	553	70.9
21	Power analyses and sample size calculations	58	7.4	586	75.1
22	Cost-benefit analysis	15	1.9	593	76.0
23	Sensitivity analysis	21	2.7	601	77.1
24	Repeated-measures analysis	18	2.3	613	78.6
25	Missing-data methods	23	2.9	626	80.3

Statistical method		Articles containing the method		Accumulation by article	
number	percent	number	percent		
26	Receiver-operating characteristic	27	3.5	643	82.4
27	Resampling	17	2.2	651	83.5
28	Generalized estimating equations	41	5.3	684	87.7
29	Linear mixed models	58	7.4	736	94.4
30	Bayesian analysis	4	0.5	739	94.7
31	Meta-analysis	8	1	742	95.1
32	Rasch analysis and item response theory	3	0.4	744	95.4
33	Generalized linear models (excluding linear and logistic regression)	10	1.3	754	96.7
34	Other methods	25	3.2	780	100.0
	Totals	2038		780	