Understanding the literature: Complexity of statistical methods used in high-impact cardiothoracic surgery research

Michael Gritti, MD,[a,b] Anusha Jegatheeswaran, MD, PhD,[c] Dolev Yissar, MD,[d] M. Anne Harris, PhD,[e,f] and Brian W. McCrindle, MD, MPH[a,b]

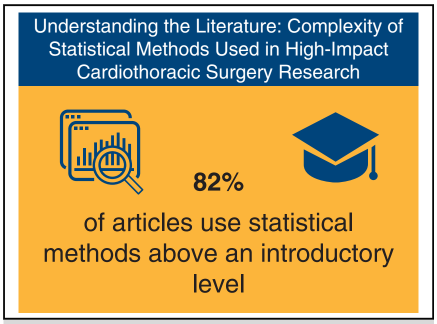
ABSTRACT

Objective: Biostatistics are frequently used in research published in the domain of cardiothoracic surgery. The objective of this study was to describe the scope of statistical techniques reported in the literature and to highlight implications for editorial review and critical appraisal.

Methods: Original research articles published between January and April 2017 in the *Journal of Thoracic and Cardiovascular Surgery, Annals of Thoracic Surgery,* and the *European Journal of Cardio-Thoracic Surgery* were examined. For each article, the statistical method(s) reported were recorded and categorized by complexity.

Results: We reviewed 293 articles that reported 1068 statistical methods. The mean number of different statistical methods reported per article was 3.6 $\pm$ 1.9, with variation by subspecialty and journal. The most common statistical methods were contingency tables (in 59% of articles), *t* tests (49%), and survival methods (49%). Only 4% of articles used descriptive statistics alone. An introductory level of statistical knowledge was deemed sufficient for understanding 16% of articles, whereas for the remainder a higher level of knowledge would be needed.

Conclusions: Contemporary cardiothoracic surgery research frequently requires the use of complex statistical methods. This was evident across articles for all cardiothoracic surgical subspecialties as reported in 3 high-impact journals. Routine review of manuscript submissions by biostatisticians is needed to ensure the appropriate use and reporting of advanced statistical methods in cardiothoracic surgery research. (J Thorac Cardiovasc Surg 2020;■:1-9)



Cardiothoracic surgery research often requires complex statistical methods in this era.

EDU

CENTRAL MESSAGE

Contemporary cardiothoracic surgery research frequently requires the use of complex statistical analyses. Advanced statistical knowledge is often needed for editorial review and critical appraisal.

PERSPECTIVE

Contemporary cardiothoracic surgery research frequently requires the use of complex statistical analyses. Only 4% of cardiothoracic surgery articles report using solely descriptive statistics. A surgeon would benefit from learning *t* tests and survival methods to optimize understanding of the literature. Advanced statistical knowledge is often needed for editorial review and critical appraisal.

See Commentary on page XXX.

There is little information in the literature regarding statistical methods used in medical journals, and none regarding cardiothoracic surgery research specifically. *The New England Journal of Medicine* provided one of the first estimates of the frequency of statistical methods reported in medical research in 1983 and showed that knowledge of

From the [a]Faculty of Medicine and [f]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; Divisions of [b]Pediatric Cardiology and [c]Cardiovascular Surgery, Hospital for Sick Children, Toronto, Ontario, Canada; [d]Medical School for International Health, Ben-Gurion University of the Negev, Be'er Sheva, Israel; and [e]School of Occupational and Public Health, Ryerson University, Toronto, Ontario, Canada.

Received for publication Jan 13, 2020; revisions received Oct 2, 2020; accepted for publication Oct 12, 2020.

Address for reprints: Michael Gritti, MD, The Hospital for Sick Children, 555 University Ave, Toronto, Ontario, Canada M5G 1X8 (E-mail: michael.gritti@sickkids.ca).

0022-5223/$36.00

https://doi.org/10.1016/j.jtcvs.2020.10.144

Scanning this QR code will take you to the article title page to access supplementary information.

descriptive statistics alone is not enough to critically appraise approximately half of articles.[1] A 2017 *The New England Journal of Medicine* update article looked at the frequency of statistical methods again and showed that the number of statistical techniques per article had increased over time and that more complicated statistical methods were being used.[2] A similar assessment was performed for The *Journal of the American Medical Association* and showed an increase in statistical complexity over time as well.[3]

Many medical specialties have been assessing the statistical methods used in their respective fields.[4-7] The lack of literature regarding the statistical methods used, specifically in cardiothoracic surgery research, warrants more investigation. Therefore, we sought to evaluate and systematically categorize the complexity of statistical methods reported in the field of cardiothoracic surgery research in high-impact journals.

MATERIALS AND METHODS

Guided by impact factor, the 3 cardiothoracic surgery journals that were chosen for assessment were the *Journal of Thoracic and Cardiovascular Surgery, Annals of Thoracic Surgery,* and the *European Journal of Cardio-Thoracic Surgery.*[8] All original articles published between January 1, 2017, and April 30, 2017, were identified using a search of Ovid MEDLINE.

The statistical methods of each individual study were reviewed and coded into 28 categories (not including “other” and “descriptive only”) derived from the 2017 *The New England Journal of Medicine* article by Sato and colleagues,[2] with some minor modifications as noted in Table 1. A subgroup analysis dividing the articles based on subspecialty and journal was performed.

The data were collected and reported as follows: (1) the frequency of statistical methods in cardiothoracic surgery original articles; (2) the mean number of statistical methods reported in each article, with the variability being standard deviations; (3) the approximate level of statistical expertise needed to understand the statistical methods; and (4) the statistical analysis program that was used.

Using Ovid MEDLINE, we searched by journals, “Annals of Thoracic Surgery.jn” or “Journal of Thoracic & Cardiovascular Surgery.jn” or “European Journal of Cardio Thoracic Surgery.jn” then limited the year to “2017.” Original articles reporting on original data were examined for the assessment. Case reports, case studies, images in cardiothoracic surgery, expert opinions, letters to the editor, surgical technique papers, and retraction of publications were excluded. A flow diagram can be found in the Figure E1. Articles were independently screened by 2 members of the research team (M.G. and D.Y.) for the number and type of statistical methods used, and discrepancies were resolved on the basis of a consensus process that incorporated the statistician as arbiter. As such, inter-rater agreement was not tracked. No consent or research ethics board approval was required for this study. All of the data were gathered and analyzed using Microsoft Excel (Microsoft Corp, Redmond, Wash).

The articles were assigned into 4 subspecialty groups based on the journal’s secondary categorizations: acquired disease cardiac surgery, congenital cardiac surgery, thoracic surgery, and other (which includes transplantation, mechanical support, education, basic science, evolving technology, and perioperative/quality improvement). We also examined the data by stratifying based on the 3 cardiothoracic surgery journals that were examined: *Journal of Cardiovascular and Thoracic Surgery, Annals of Thoracic Surgery,* and *European Journal of Cardio-Thoracic Surgery.* The dataset is available by request to the corresponding author.

“Understandability” Analysis

Most articles in cardiothoracic surgery will use a combination of statistical methods. Therefore, reviewers would need to understand multiple statistical methods to fully assess the statistics in each article. To quantify this, we first determined how many articles a theoretical reviewer would understand if they knew no statistical methods. Then, we determined how many more articles they would fully understand if this person learned one more type of statistical method. We continued this analysis until the theoretical reviewer knew all the statistical methods used in our dataset in a stepwise fashion. Because there is no standardized order in which these methods are typically learned, we decided on an order that *maximally* increased the percentage of articles fully understood if this theoretical reviewer learned one more method.[1]

Training Analysis

The statistical methods categories were then placed into levels based on training required to understand, critique, and use each method. The levels were derived by a Professor of Biostatistics who is familiar and involved in biostatistical training at our local university (M.A.H.).

The levels we used were Introductory, Intermediate, Advanced, and Specialized. We defined “Introductory” as biostatistical methods taught in introductory undergraduate level university courses. These methods included descriptive statistics, *t* tests, contingency tables, Pearson’s correlation, simple linear regression, analysis of variance, and multiple comparisons. “Intermediate” was defined as statistical methods that would be taught in additional courses that could be taken at an undergraduate level or in an early graduate degree. These included epidemiologic statistics, propensity score, nonparametric tests, multiple regression, adjustment and standardization, multiway tables, power analysis, survival methods, transformation, nonparametric correlation, sensitivity analysis, and meta-analysis. Finally, “Advanced and Specialized” was defined as methods requiring advanced graduate training in biostatistics or epidemiology and specialized techniques in specific individual training. These include repeated-measures analysis, receiver operating characteristic curves, missing data imputation methods, bootstrap resampling, principal component analysis, cluster analysis, Bayesian statistics, genetic analysis, noninferiority trials, and cost-benefit analysis. Many of these statistical methods are summarized for a cardiothoracic surgeon in the textbook *Cardiac Surgery* by Kirklin and Barratt-Boyes in Chapter 6.[9]

We then performed an analysis showing the cumulative percentage of articles in our dataset a reader or reviewer would fully understand with each increasing level of education.

Statistical Analysis Program

The statistical analysis program that was reported in each article was noted, including if no program was cited in the article.

RESULTS

Statistical Methods

A total of 293 articles were included in the study. We ranked the statistical methods (Table 1) in terms of the 10 most commonly used statistical methods (Table 2). The 3

TABLE 1. Statistical method categories

Category	Brief description
No statistical methods or descriptive statistics	No statistical content or descriptive statistics only (eg, percentages, means standard deviations, standard errors, histograms)
t test	1-sample, matched-pair, 2-sample *t* tests, z-test
Contingency tables	Chi-square test, Fisher exact test, McNemar's test, Mantel–Haenszel test
Nonparametric test	Sign test, Wilcoxon signed-rank test, Mann–Whitney test, median/range test, Kruskal–Wallis test, Friedman test, Kolmogorov Smirnov test, Jonckheere–Terpstra test, Begg's test
Epidemiologic statistics	Relative risk, odds ratio, log odds, measures of association, sensitivity, specificity, Bland–Altman plot analysis, Lin's concordance correlation coefficient, Cronbach's alpha
Propensity score	Matching, regression adjustment/ stratification, weighting by using propensity score
Pearson's correlation	Classic product-moment correlation
Simple linear regression	Least-squares regression with 1 predictor and 1 response variable
Analysis of variance	Analysis of variance, analysis of covariance, simple linear contrasts, F-test, Levene test
Transformation/distribution	Use of data transformation, Grubbs test, Shapiro–Wilk test, modified Thompson Tau test
Nonparametric correlation	Spearman's rho, Kendall's tau, monotone regression, trend test
Survival methods	Survival function, Kaplan–Meier plot, proportional hazards model, Other survival model, rate adjustment, log-rank test, Grambsch-Therneau test, Harrell C-statistic, Fine and Gray method
Multiple regression	Includes polynomial regression and stepwise regression, Hosmer-Lemeshow test
Multiple comparisons	Procedures for handling multiple inferences on same data (eg, Bonferroni techniques, Scheffé's contrasts, Holm, Dunnett, Duncan's, Newman-Keuls procedure, false discovery rate)
Adjustment and standardization	Pertains to incidence rates and prevalence rates
Multiway tables	Mantel–Haenszel procedure, log-linear models, logistic regression
Power analysis	Loosely defined, includes use of the size of detectable (or useful) difference in determining sample size
Cost-benefit analysis	The process of combining estimates of cost and health outcomes to compare policy alternatives
Sensitivity analysis	Examines sensitivity of outcome to small changes in parameters of Model or in other assumptions
Repeated-measures analysis	Repeated-measures analysis of variance, generalized estimating equation, mixed-effect models for repeated measures, longitudinal regression
Missing data methods	Listwise deletion, pairwise deletion, mean substitution, simple hot-deck, regression estimation, complete case method, single imputation, multiple imputation
Noninferiority trial	Noninferiority trial is whether the experimental therapy is not inferior to the active control
Receiver operating characteristic	Advanced decision statistics based on analysis of receiver operating characteristic curves
Resampling	Bootstrap, Jackknife, cross-validation, permutation procedures
Principal component analysis	Factor analysis, stepwise discriminant analysis, Varimax rotation
Cluster analysis	Hierarchical, K-Means, 2-step clustering, density-based spatial clustering of applications with noise, multilevel modeling, random-effects models, nested data models
Meta-analysis	Meta-analysis is a statistical technique for combining the findings from independent studies
Genetic analysis or Statistical genetics	Aggregation, heritability and segregation analysis, linkage analysis, genetic association analysis, population substructure, gene-expression Data analysis
Bayesian statistics	Bayesian statistics

Categories of statistical methods used to assess the statistical content of the articles derived from Sato and colleagues[2] with the addition of Bayesian Statistics as a separate category.

EDU

TABLE 2. Top 10 statistical methods used across all articles

Rank	Statistical method	Percentage of total statistical methods (%)	Articles using these methods (%)
1	Contingency tables	16%	59%
2	*t* test	13%	49%
3	Survival methods	13%	49%
4	Nonparametric test	12%	45%
5	Multiway tables	10%	37%
6	Multiple regression	4%	16%
7	Epidemiologic statistics	4%	14%
8	Propensity score	3%	12%
9	Analysis of variance	3%	12%
10	Transformation/distribution	3%	10%

The top 10 statistical methods used, presented as a percentage of total data points collected and a percentage of articles using these methods.

most common statistical methods used were (1) contingency tables (found in 59% of articles, n = 172, 172/293); (2) *t* tests (found in 49% of articles, n = 144, 144/293); and (3) survival methods (found in 44% of articles, n = 144, 144/293). Only 4% (n = 13, 13/293) of the articles reported only descriptive statistics or reported no statistical methods at all. The mean number of statistical methods reported in each article was 3.6 ± 1.9. A few key findings are shown in Figure 1.

Statistical Methods Used by Journal

When stratified by journal, the number of articles analyzed per journal was as follows: 136 (46%) in the *Annals of Thoracic Surgery*, 83 (28%) in *European Journal of Cardio-Thoracic Surgery*, and 74 (25%) in *Journal of Thoracic and Cardiovascular Surgery*. The mean number of statistical methods per article in each journal is 3.9 ± 2.0 for *Journal of Thoracic and Cardiovascular Surgery*, 3.5 ± 1.9 for *Annals of Thoracic Surgery*, and 3.6 ± 1.8 for *European Journal of Cardio-Thoracic Surgery*. The impact factors for these journals were 3.8, 2.3, and 3.1 in 2017, respectively.[8]

Statistical Methods Used by Subspecialty

The most common subspecialty type was acquired disease cardiac surgery (n = 121), followed by thoracic surgery (n = 82) and congenital cardiac surgery (n = 42). The "other" category made up the remaining (n = 48) articles. Analysis by subspecialty showed that the mean of the number of statistical methods reported for articles from each subspecialty included 4.1 ± 1.9 methods per article for acquired disease cardiac surgery, 3.5 ± 1.7 methods for thoracic surgery, and 3.3 ± 1.8 methods per article for congenital cardiac surgery.

The 3 most common statistical method categories are reported for articles associated with each subspecialty. Table 1 specifies the methods in each category. For articles about acquired disease cardiac surgery, these include (1) contingency tables, n = 71, 59% (71/121) of articles; (2) *t* tests, n = 67, 55% (67/121) of articles; and (3) nonparametric tests, n = 62, 51% (62/121) of articles. For congenital cardiac surgery articles, these include (1) contingency tables, n = 25, 60% (25/42) of articles; (2) nonparametric tests, n = 21, 50% (21/42) of articles; and (3) survival methods, n = 18, 43% (18/42) of articles. For thoracic surgery articles these include (1) contingency tables, n = 58, 71% (58/82) of articles; (2) survival methods, n = 50,

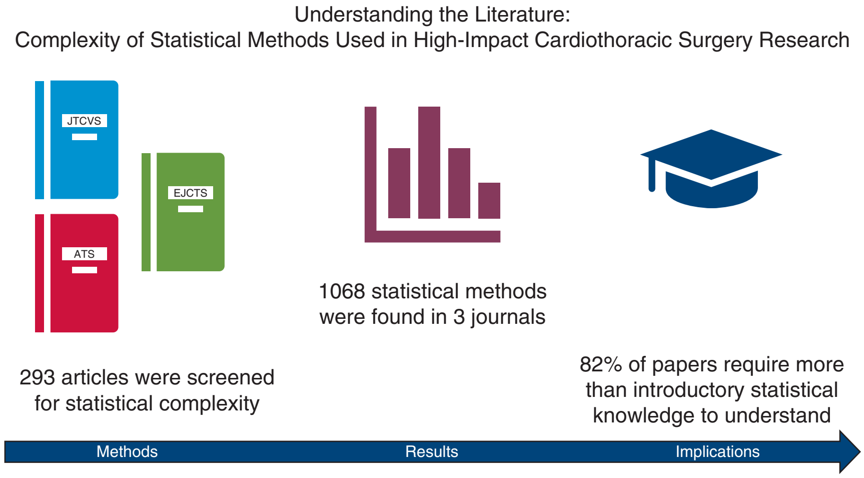


FIGURE 1. Key results. A summary of a few of the key results from the study including the number of articles screened for statistical complexity, the number of statistical methods found, and the percentage of statistical methods in the study that were classified as above an introductory level statistical education. *JTCVS*, *Journal of Thoracic and Cardiovascular Surgery*; *EJCTS*, *European Journal of Cardio-Thoracic Surgery*; *ATS*, *Annals of Thoracic Surgery.*

61% (50/82) of articles; and (3) *t* tests, n = 40, 49% (40/82) of articles.

"Understandability" of Statistical Methods

As Table 3 shows, an understanding of 5 statistical methods would be required for a full statistical understanding of 15% of articles in our data set. An understanding of 10 statistical methods would be required for a full statistical understand of 22% of the literature and 15 statistical methods for 78% of the literature.

Training Required for Understanding

As Table 4 shows, with an introductory level of biostatistical training, only 16% (n = 47) articles in this dataset were fully accessible to the editor, reviewer, or reader. With an intermediate level of biostatistical training, the majority of the articles (83%) were accessible. The data showed that 17% of the articles had statistical methods that required advanced or specialized training to understand.

Statistical Programs Used

When we examined the statistical programs reported, 41 articles (15%) did not mention any specific statistical program. The most common statistical programs reported were the Statistical Package for the Social Sciences (37%, n = 107 articles), Statistical Analysis System (17%, n = 51 articles), and Stata (15%, n = 44 articles). In total, there were 21 different programs reported (Table 5).

TABLE 3. "Understandability" analysis

Summary of statistical methods used across all journals				
Method category	Articles using these methods (n)	Articles using these methods (%)	Accessible articles (n)	Accessible articles (%)
No statistical methods or descriptive statistics	13	4%	13	4% (13/293)
t test	144	49%	20	7% (20/293)
Contingency tables	172	59%	28	10% (28/293)
Nonparametric test	132	45%	39	13% (39/293)
Epidemiologic statistics	41	14%	42	14% (42/293)
Propensity score	36	12%	43	15% (43/293)
Pearson's correlation	13	4%	45	15% (45/293)
Simple linear regression	25	9%	47	16% (47/293)
Analysis of variance	36	12%	57	19% (57/293)
Transformation/distribution	30	10%	61	21% (61/293)
Nonparametric correlation	11	4%	63	22% (63/293)
Survival methods	144	49%	129	44% (129/293)
Multiple regression	48	16%	139	47% (139/293)
Multiple comparisons	22	8%	157	54% (157/293)
Adjustment and standardization	2	1%	158	54% (158/293)
Multiway tables	108	37%	228	78% (228/293)
Power analysis	6	2%	234	80% (234/293)
Cost-benefit analysis	1	0%	234	80% (234/293)
Sensitivity analysis	12	4%	242	83% (242/293)
Repeated-measures analysis	18	6%	253	86% (253/293)
Missing-data methods	17	6%	262	89% (262/293)
Noninferiority trial	0	0%	262	89% (262/293)
Receiver operating characteristic	0	0%	262	89% (262/293)
Resampling	18	6%	276	94% (276/293)
Principal component analysis	0	0%	276	94% (276/293)
Cluster analysis	14	5%	288	98% (288/293)
Meta-analysis	2	1%	290	99% (290/293)
Genetic analysis or statistical genetics	0	0%	290	99% (290/293)
Bayesian statistics	3	1%	293	100% (293/293)

This shows how many articles a theoretical reader or reviewer would understand if he/she initially knew no statistical methods and then how many more articles he/she would fully understand if this person learned one more type of statistical method at a time. We continued this analysis until the theoretical reviewer knew all the statistical methods used in our dataset in a stepwise fashion.

EDU

TABLE 4. Level of statistical training related to "understandability" of the methods used

Training required	Brief description of training	Statistical methods categories	% of articles that are understandable with level of education	Number of articles that require at least this level of education
Introductory	Typical introductory biostatistics course at undergraduate level	No statistical methods or descriptive *t* test Contingency tables Pearson's correlation Simple linear regression Analysis of variance Multiple comparisons	16%	47
Intermediate	Additional courses at undergraduate level (eg, epidemiology, further biostatistics) or training in epidemiology/biostatistics at early graduate degree level	Epidemiologic statistics Propensity score Nonparametric tests Multiple regression Adjustment and standardization Multiway tables Power analysis Survival methods Transformation/distribution Nonparametric correlation Sensitivity analysis Meta-analysis	83%	196
Advanced	Advanced graduate training in biostatistics or epidemiology	Repeated-measures analysis Receiver operating characteristic Missing data methods Resampling Principal component analysis Cluster analysis Bayesian statistics Genetic analysis or Statistical genetics	100%	49
Specialized	Specialized techniques possibly taught in specific individual training	Noninferiority trial Cost-benefit analysis	100%	1

Statistical methods categorized in terms of training required to understand each method. We performed an analysis showing the percentage of articles that readers would fully understand if they were to know all the statistical methods within that level of education and below.

EDU

DISCUSSION

Statistical Methods Used in Other Medical Fields

In addition to *The New England Journal of Medicine* and *Journal of the American Medical Association* studies, many medical specialties have been analyzing the statistical methods in their respective fields, including plastic surgery,[4] ophthalmology,[5] family medicine,[6] and public health.[7] After examining these studies, it becomes clear that *t* tests and contingency tables are consistently the most common statistical methods used. This is consistent with these being the first step to examining unadjusted bivariate relationships. Our study showed similar findings in cardiothoracic surgery research, with contingency tables and *t* tests being the top 2 statistical methods used.

Increasing Complexity of Statistical Methods Over Time in Medical Literature

The medical literature has shown a trend toward a large increase in the number of statistical techniques used over the previous 3 decades.[2,5] Innovation in biostatistical techniques and greater access to computers with statistical software have facilitated the increase in the complexity of statistical methods used and reported in the medical literature.[5] However, as is also evident from the literature, the advances in statistical analysis have not been met with an equivalent advance in the ability of healthcare professionals to understand complex statistical analyses, nor for journal editors and peer reviewers to evaluate them before publication.[10-12] This highlights the importance of the involvement of statistical editors or reviewers more routinely in the review process of manuscript submissions.

Lack of Statistical Expertise Among Healthcare Providers

A study of 11 residency training programs in the United States showed that the majority of residents lacked sufficient statistical knowledge to accurately interpret and evaluate most results reported in the published medical

TABLE 5. Statistical programs used in cardiothoracic surgery research

Packages listed	No. of articles using package	Percentage of articles using package
SPSS	107	37%
SAS	51	17%
Stata	44	15%
No package stated	41	14%
R	38	13%
Prism	19	6%
JMP	10	3%
Microsoft Excel	4	1%
Review Manager	3	1%
MEDCalc	3	1%
Statistica	2	1%
SciPy	1	0%
PS Matching Package	1	0%
StatView	1	0%
MATLAB	1	0%
Meta-analysis	1	0%
SEM	1	0%
FLARE	1	0%
WinBUGS	1	0%
PASW	1	0%
SYSTAT	1	0%
JumpPro	1	0%

Presented are the types of statistical programs used, number of articles using each package, and percentage of total articles that use each package. *SPSS*, Statistical Package for the Social Sciences; *SAS*, statistical analysis system; *SEM*, standard error of the mean.

literature.[10] Remarkably, 75% of the residents indicated that they did not understand all of the statistics that they encountered in journal articles.[10] Moreover, medical fellows and general medicine faculty with research training showed a mean comprehension of only 41% of the statistical methods and results from the medical literature.[10] Lack of statistical knowledge among medical residents is a universal concern in the context of today's research methodology.[11,12] As our data indicate, knowledge of only descriptive statistics (eg, percentages, means, standard deviations, standard errors, histograms) would enable an understanding of only 4% of the articles in our data set. Thus, a limited knowledge of statistics poses a significant problem for the readers of cardiothoracic surgery research, as well as journal editors and peer reviewers.

Statistical Errors Are Common in Medical Literature

Medical learners pursuing academic career pathways are often required to become active researchers in their respective specialties. As medical residents progress to become researchers, their level of statistical knowledge may significantly affect the quality of their research. A study that examined critical and frequent errors in rejected manuscripts from high-impact medical journals found that manuscripts are commonly rejected because of inappropriate applications of statistical methods.[13] Consequently, many researchers resubmit without statistical revision these same articles to subsequent journals of progressively lower-impact factors until publication is achieved.[13] To make matters worse, multiple studies have indicated that most medical journal articles contain multiple statistical errors.[14-18] A recent literature review has indicated that approximately 50% of clinical publications contain at least 1 statistical error, some of which may have deleterious effects on the results and interpretations.[19,20] Even simple and basic procedures, such as the *t* tests or contingency tables, are frequently misused in medical research.[14] For example, *t* tests and chi-square tests are frequently used despite their test assumptions not being evaluated and sufficiently met before application.[14] The misapplication and interpretation of statistical methods may lead to incorrect conclusions and misuse of valuable resources,[14] and may even lead to detrimental clinical practices.[14,21,22] The role for a biostatistician in the review process may be more necessary than in the past.

The Need for Biostatisticians

In response to this problem, many medical journal editors have enhanced their statistical guidelines to filter out manuscripts that contain statistical errors.[14,23-27] However, little improvement has yet been achieved, as evidence continues to show statistical errors.[14,28-31] A tutorial paper meant for junior reviewers in surgical fields stated that many surgical reviewers do not feel adequate in assessing statistical analyses; thus, editors should have a low threshold to request a formal statistical review by a journal's dedicated biostatistical team, if the journal has one.[32] The results of our study confirm that the complexity of modern cardiothoracic surgical research justifies this low threshold for involving a statistical reviewer. The lack of congruence between the increased use of more complex statistics and the lack of statistical knowledge among medical professionals decreases the ability to understand, critically review, and produce medical literature. This may adversely affect the quality of clinical practice, particularly the practice of best evidence-based clinical decision-making. Editorial boards of cardiothoracic surgery journals should ensure original articles are reviewed by individuals with sufficient expertise in statistics to ensure high-quality publications with appropriate use of statistical methods and accurate description and interpretation of results. These individuals would not be the sole statistical reviewers of an

EDU

article but would supplement the peer review process already in place.

Biostatisticians in the Cardiothoracic Surgery Review Process

Although it is encouraged that medical professionals attain knowledge of statistical methods to enhance their capabilities as independent readers, the importance of collaboration with dedicated statisticians should not be overlooked. The progressive complexity of statistical methods used may require reviewer assistance from dedicated statisticians. The higher-impact journals have a higher mean number of statistical methods used per article that further highlight the importance of biostatisticians. The best approach would be for every original article to receive a statistical review, as the *Journal of Thoracic and Cardiovascular Surgery* does, the highest impact factor journal in our study. If this cannot be the case, our study underscores the importance for cardiothoracic surgery editors and reviewers to lower their threshold for requesting a journal's formally designated statisticians to provide additional review. If a journal does not have formally designated statisticians, it may be beneficial to include these individuals in this current era. Collaboration with statisticians, who possess Masters or PhD-level expertise, in the early phases of study conceptualization can ensure that complex statistical methods are used correctly and that the results are valid, reliable, clearly presented, and accurately interpreted.

Educating the Future Cardiothoracic Surgeon

Our study also shows that a basic understanding of introductory statistics is needed at a minimum. If a cardiothoracic residency statistical training program were created, the statistical methods emphasized in the curriculum should potentially be those that would yield the highest degree of understanding of the most number of articles related to the specialty. If implemented, this approach could be further tested with future studies to assess its benefit on biostatistical competence in the trainees. If proven to be beneficial, this focused approach may improve the statistical abilities of the peer reviewers of the future.

Although 14% of articles did not mention a statistical program being used, only 4% of articles did not use a statistical method. The Statistical Package for the Social Sciences and Statistical Analysis System made up the bulk of statistical programs (37% and 17%, respectfully) that were used and reported.

Study Limitations

Although we think the articles chosen as the study sample are a representative valuation of the variety of statistical methods being used in the cardiothoracic surgery literature, it is a snapshot of only 3 journals reviewed over the chosen 4-month period in 2017 and may have missed rarer and more contemporary methods, such as machine learning. A further expansion of this study to cover research in future years would provide valuable information into the trends of statistical methods used over time. No assessment was carried out regarding the type of research of each study, the qualifications or level of statistical expertise of the authors (or acknowledgement of a statistician), or the appropriateness and validity of the statistical methods used in each reviewed article. We also could not assess whether a biostatistician was involved, because this is usually not explicitly stated. Determining the involvement and qualifications of statisticians might be recommended to be included in the manuscript submission and peer-review process. The incorporation of expert-level statisticians and clinicians with advanced statistical expertise routinely in peer review would help in the prevention of publication of statistical errors. Regular statistical audits by journals might also be useful. These are areas for future exploration. Furthermore, the classification of training required for each statistical method used was based on our local institution, and discrepancies among institutions, countries, and education systems would add further valuable information to Table 3. Last, we cannot possibly determine the degree to which our "Understandability" scale reflects the training of all authors in the true order of which they learn statistical methods; however, we think our scale reflects general trends toward increasing statistical competence.

CONCLUSIONS

The field of cardiothoracic surgery research has been moving in the direction of using increasingly complex statistical methods. As with other fields of medicine, a large portion of the statistics used are beyond the scope of introductory statistics. Descriptive statistics alone were used in 4% of studies, and less than 10% of articles in this data set relied on a single method of statistical analysis other than descriptive statistics. Additionally, introductory level statistics will only give one a complete understanding of 16% of the articles in this study. The number of statistical methods used per article is consistent across the various cardiothoracic surgery research subspecialties as well as across the field's top 3 journals by impact factor. Our study lends insights into the type of statistical education readers may want to pursue as it pertains to cardiothoracic surgery research. Statistical training may be more important than ever for a career, because a large portion of adult thoracic surgery and almost all of pediatric thoracic surgery are performed in academic institutions. Finally, the increasing number and complexity of statistical methods, as shown in our study, highlight the importance of journals having statistical experts participate more routinely in the review process. Cardiothoracic surgery journal editorial boards should more routinely involve those with more advanced expertise in biostatistics to ensure that their journal publishes high-quality literature with results that are valid and reliable, and presented and interpreted accurately using appropriate statistical methods.

Conflict of Interest Statement

The authors reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

The authors thank Dean Vaksman for help with the tables in this study.

References

1. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med*. 1983;309:709-13.
2. Sato Y, Gosho M, Nagashima K, Takahashi S, Ware JH, Laird NM. Statistical methods in the journal - an update. *N Engl J Med*. 2017;376:1086-7.
3. Arnold LD, Braganza M, Salih R, Colditz GA. Statistical trends in the Journal of the American Medical Association and implications for training across the continuum of medical education. *PLoS One*. 2013;8:e77301.
4. Al-Benna S, Al-Ajam Y, Way B, Steinstraesser L. Descriptive and inferential statistical methods used in burns research. *Burns*. 2010;36:343-6.
5. Lisboa R, Meira-Freitas D, Tatham AJ, Marvasti AH, Sharpsten L, Medeiros FA. Use of statistical analyses in the ophthalmic literature. *Ophthalmology*. 2014;121:1317-21.
6. Aljoudi AS. Study designs and statistical methods in the Journal of Family and Community Medicine: 1994-2010. *J Family Community Med*. 2013;20:8-11.
7. Karran JC, Moodie EE, Wallace MP. Statistical method use in public health research. *Scand J Public Health*. 2015;43:776-82.
8. Scimago Journal & Country Rank. Scimago .ab. Amsterdam, The Netherlands: Scimago Labs, 2017. Available at: https://www.scimagojr.com/journalrank.php?category=2746&year=2017&type=j. Accessed December 15, 2020.
9. Kirklin JW, Barratt-Boyes BG. Chapter 6: generating knowledge from information, data, and analyses. In: *Cardiac Surgery: Morphology, Diagnostic Criteria, Natural History, Techniques, Results, and Indications*. New York: Wiley; 1986:253-343.
10. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA*. 2007;298:1010-22.
11. Bland JM, Altman DG, Royston JP. Statisticians in medical schools. *J R Coll Physicians Lond*. 1990;24:85-6.
12. Altman DG, Martin Bland J. Improving doctors' understanding of statistics (with discussion). *J R Statist Soc A*. 1991;2:223-67.
13. Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AH. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes*. 2011;4:304.
14. 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.
15. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *Br Med J*. 1977;1:85-7.
16. MacArthur RD, Jackson GG. An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. *J Infect Dis*. 1984;149:349-54.
17. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317:426-32.
18. McKinney WP, Young MJ, Hartz A, Lee MB. The inexact use of Fisher's Exact Test in six major medical journals. *JAMA*. 1989;261:3430-3.
19. Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. *Int Dent J*. 2011;61:163-7.
20. Thiese MS, Ronna B, Robbins RB. Misuse of statistics in surgical literature. *J Thorac Dis*. 2016;8:E726-30.
21. Altman DG. Statistics and ethics in medical research. VIII-Improving the quality of statistics in medical journals. *Br Med J (Clin Res Ed)*. 1981;282:44-7.
22. Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools, and a research agenda. *Account Res*. 2002;9:65-74.
23. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med*. 1998;13:753-6.
24. Altman DG. Statistical reviewing for medical journals. *Stat Med*. 1998;17:2661-74.
25. Gore SM, Jones G, Thompson SG. The Lancet's statistical review process: areas for improvement by authors. *Lancet*. 1992;340:100-2.
26. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)*. 1983;286:1489-93.
27. Murray GD. Statistical guidelines for the British Journal of Surgery. *Br J Surg*. 1991;78:782-4.
28. Cooper RJ, Schriger DL, Close RJ. Graphical literacy: the quality of graphs in a large-circulation journal. *Ann Emerg Med*. 2002;40:317-22.
29. García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. *BMC Med Res Methodol*. 2004;4:13.
30. Olsen CH. Review of the use of statistics in infection and immunity. *Infect Immun*. 2003;71:6689-92.
31. Marshall SW. Testing with confidence: the use (and misuse) of confidence intervals in biomedical research. *J Sci Med Sport*. 2004;7:135-7.
32. Stahel PF, Moore EE. How to review a surgical paper: a guide for junior referees. *BMC Med*. 2016;14:29.

Key Words: cardiothoracic surgery, medical education, statistics

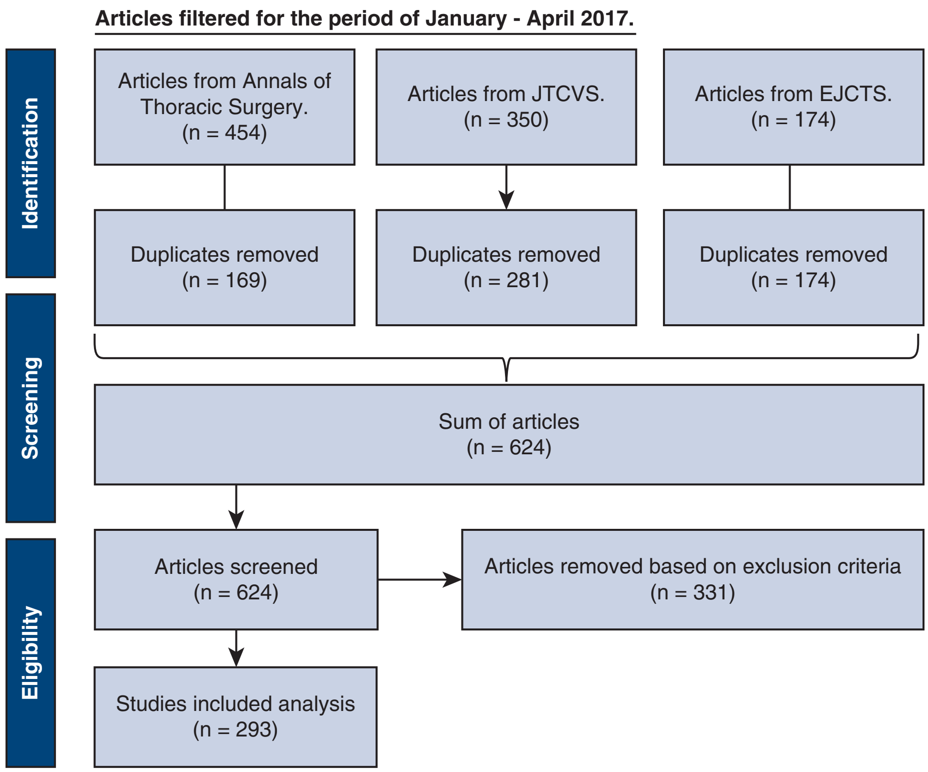


FIGURE E1. Flow diagram showing the number of studies identified from the literature search and the number of articles excluded. Articles were excluded for the following reasons: repeated entries, case reports, case studies, images in cardiothoracic surgery, expert opinions, letters to the editor, surgical technique papers, and retraction of publications. *JTCVS*, *Journal of Thoracic and Cardiovascular Surgery*; *EJCTS*, *European Journal of Cardio-Thoracic Surgery*.