

Recent trends in utilization of statistical methods in anesthesia research: 2012-2017

Steven J Staffa and David Zurakowski*

Department of Anesthesiology, Critical Care and Pain Medicine, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Abstract

Introduction: Statistics play an integral role in obtaining results and answering research questions. Statistical methods are essential tools for analyzing and interpreting new knowledge in anesthesia research. Using the most appropriate statistical methods allows researchers and anesthesiologists to answer their important questions objectively to advance practice. To best answer questions about potential risk factors of patient outcomes, safety and efficacy of anesthetics and analgesics, hospital costs, and quality of care, applications of statistical methods are invaluable. We examined the current use of study design features and statistical methodologies in the anesthesia literature to illustrate the recent trends and to compare two premier journals: *Anesthesia & Analgesia* and *Anesthesiology*.

Methods: We reviewed every research publication from January 2012 through July 2017 in *Anesthesia & Analgesia* and *Anesthesiology* for over 40 study design features and simple to advanced statistical methods, leading to a sample size of 2,267 articles that included use of inferential statistics.

Results: The most common methods included Student's t-tests (59%), categorical data analysis (34%), nonparametric testing (46%), multiple regression analysis (24%), power analysis (36%), and adjustment for multiple comparisons (47%). The rate of use of propensity score matching, Meta-Analysis, and sample size considerations or power analyses has increased year by year from 2012 to 2017, whereas the frequency of Student t-tests and ANOVA declined (all $P < 0.05$). More basic science articles are published in *Anesthesiology* (42%) than *Anesthesia & Analgesia* (24%) ($P < 0.001$), leading to more repeated measures ANOVAs being found in *Anesthesiology* (29% vs. 14%) ($P < 0.001$). On the other hand, *Anesthesia & Analgesia* features more clinical science articles (76% vs. 58%) ($P < 0.001$), and more categorical data analysis methods (36% vs. 31%) ($P = 0.008$).

Conclusions: Our study demonstrates that a wide range of statistical methods are being utilized in anesthesia research and trends reveal good statistical practice from simple to more advanced techniques. Specialized methods are becoming more common, suggesting a closer collaboration between anesthesiologists and biostatisticians. Our study is the first of its kind, and the results suggest that statistical expertise is found in anesthesia research and will continue to grow, which will improve the rigor and impact of research studies in the specialty and lead to more informed interpretations, understanding, and conclusions from original research.

Introduction

As research methods evolve and modernize over time, equipping the researcher to best address their research questions properly and analyze their data is of paramount importance. As is true with the machinery or technology in any field, new biostatistical methodologies are constantly being developed. Statistician John Tukey once said "The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data." More advanced and modern statistical methodologies are important for the researcher to have in their tool box in order to be able to choose the statistical methodology to achieve the most valid results and extract a reasonable answer from the data. Researchers have a responsibility to ensure that the correct statistical methodology is being used.

Many of the most popular statistical methodologies are the simplest and perhaps have been used for decades. While there may be another inferential statistical method that could be implemented to answer the research question at hand, a researcher may opt to implement a better-known and perhaps less suitable method. Thus it is very important for a research team to be aware of newer statistical methodologies that may improve the quality of their research. It may be the case that many research questions can be adequately answered by analyzing the data with common statistical approaches, and these approaches are valuable and not flawed by any means. Additionally, anesthesia research data may be more suitably analyzed using certain statistical methods as opposed to others.

A recent study by Sato et al. looked at 238 articles published in the *New England Journal of Medicine* in 2015 and summarized the statistical methods used [1]. In our study, we take a "snapshot" of the current state of statistical applications in the anesthesiology literature by reviewing every research article published in the past 5 years, from January 2012 through July 2017 in *Anesthesia & Analgesia* (A&A) and *Anesthesiology*. Additionally, we compare the statistical methods between these two of the most highly respected journals in the field, and analyze the statistical trends from 2012 through 2017 to make a projection through 2020. The emphasis of this study is to find how often various statistical methods are used in published articles, and which methods are used the most, and to determine what this may suggest about the future.

Methods

Eligibility Criteria

We examined the current statistical methods in published Anesthesia research. To do so we chose to look in two of the most

*Correspondence to: David Zurakowski, MS, PhD, Director of Biostatistics, Departments of Anesthesiology and Surgery, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA, Tel: 617-355-7839; Fax: 617-730-0610; E-mail: david.zurakowski@childrens.harvard.edu

Received: May 01, 2018; Accepted: May 14, 2018; Published: May 16, 2018

prestigious journals in the field: *A&A* and *Anesthesiology*. In order to explore the recent statistical trends in anesthesia research, we chose a time window of the most recent 5 years. All published original research articles from January 2012 through July 2017 were reviewed and information was extracted from each primarily on the statistical methods used. Though *A&A* and *Anesthesiology* each have different publication types for journal organization, certain publication types of each journal were excluded as they did not utilize any statistics. These publication types are special articles, narrative reviews, echo rounds, letters to the editor, editorials, images, poems, open-mind articles, medical intelligence articles, and brief reports with no statistics. Included publication types were original research articles (clinical and basic science), systematic reviews and meta-analysis, and brief reports that include statistics. This led to a total of 2,387 articles reviewed, 1,333 from *A&A* and 1,054 from *Anesthesiology*. For analysis, 120 articles that included only descriptive statistics were excluded. Therefore our final sample size was 2,267 articles. Of these 2,267 publications, 1,245 articles were from *A&A* and 1,022 were from *Anesthesiology*.

Data Collection

A variety of information was collected on each publication. Information on year and month of publication, article type (within each journal), and research type (basic science or clinical research) was recorded. The presence of the following study design features was recorded: Randomized Controlled Trial, non-inferiority trial, randomized crossover study, Pharmacokinetics (PK)/Pharmacodynamics (PD) or bioequivalence study, use of surveys or questionnaires, and systematic review or meta-analysis [2].

The use of many statistical methods was recorded. The presence of the following statistical items was determined for each article: Student's t-tests [3], categorical data analysis methods [3] (Chi-square test, Fisher's exact test, or McNemar's test), non-parametric tests or rank statistics [3] (Wilcoxon, Mann-Whitney U, Kruskal-Wallis, Friedman, Kolmogorov-Smirnov, or Shapiro-Wilk), epidemiologic statistics (e.g. incidence, prevalence, relative risk, odds ratio), Pearson correlation [3], Spearman rank correlation [3], intraclass correlation, analysis of the variance (ANOVA), ANOVA for repeated measures, ROC analysis [4], linear regression [5], Poisson regression [5], logistic regression [5], nonlinear regression [5], multiple regression [5], survival methods [6] (Kaplan Meier methods, log-rank test, etc.), Cox proportional hazards regression [6], longitudinal regression [7], mixed effects models [7], multiple comparisons or multiple testing adjustment, power analysis, transformation of data, missing data methods [7], sensitivity analyses (re-analysis after altering/relaxing assumptions or subgroup analysis), reliability of measurements, cost-effectiveness analysis, dose-response analysis, drug combination and synergy, genetics, Bayesian methods, propensity score matching [8], and bootstrap resampling, simulation or validation.

A Microsoft Excel 2010 spreadsheet was used for data recording and organization during the review of the literature. This manuscript adheres to the applicable Equator guidelines.

Statistical Analysis

To analyze our data, we looked at our information overall and the use of statistical methods across the 2012-2017 time period stratified by journal. Since all variables were coded to be binary, we describe our results using frequencies and relative rates. The comparison of the rates of each statistical method and study design feature between the two journals was done using Fisher's exact test. We assessed relative

rates of each statistical technique over the 5 year period first visually by graphing the rates over time, and the Mantel-Haenszel chi-square test for trend was used to test of significant changes over time in the rates. A simple linear projection through the year 2020 was made for select statistical items [9]. The required assumptions were that the trend is truly linear, that our data from 2012 through 2017 accurately represents the direction that the statistical usage is going, and that the years are independent.

Stata version 13.1 (StataCorp, College Station, Texas) was used for statistical analyses. EpiInfo version 7.2.1 was used to perform the chi-square tests for trend. No power calculation was performed, as our sample size was determined by the number of research articles published from January 2012 through July 2017. Our large resulting sample size of 2,267 articles provides high power to detect differences between journals and across the years. The alpha value of 0.05 was used as a threshold to determine statistical significance.

Results

Description of Overall Current Trends

First, we sought to describe the current rates of use of each item in order to determine the current state of statistics in anesthesia research. To do so, we looked at our data across all 5 years from 2012 through 2017 in *A&A* and *Anesthesiology* combined. Table 1 displays the current rates of use of the various study design elements and statistical techniques by journal and total, from 2012 to 2017 combined. Our 5 year window allows us to describe which statistical methods have been used in the modern anesthesia literature.

Looking at study design features among all articles in *A&A* and *Anesthesiology* since the start of 2012, basic science research made 33% of publications and 67% were clinical research articles. Randomized Controlled Trials make up 13% of publications. Since the RCT is the gold standard study design to produce the highest level of evidence and is hard and costly to organize and run, this a respectable rate. Noninferiority trial design (1%) and randomized crossover studies (1%) were rare. Four percent of studies featured a meta-analysis, and 6% featured PK, PD, or bioequivalence. Surveys or questionnaires were used to collect data in 7% of research studies. Economic or cost-effectiveness studies only comprised 1% of publications.

Many classical statistical approaches are very widely used in anesthesia publications from *A&A* and *Anesthesiology*. Student's t-tests (59%), categorical data analysis (34%), and non-parametric methods (46%) were very common. Multiple regression was used in 24% of articles, with 12% using linear regression, 2% using Poisson regression, and 17% using logistic regression. Survival methods appeared in 8% of publications, and longitudinal methods appeared in 29% of publications. Since longitudinal regression also includes mixed effects modelling for hierarchical data, this is why longitudinal regression was found in even more articles (36%). Repeated measures ANOVA (21%) and non-repeated measures ANOVA (30%) were very common.

A formal power analyses was performed *a priori* in 36% of articles from *A&A* and *Anesthesiology* combined, though an additional 9% of articles stated that the sample size was based on previous experience or studies. This was most often found in laboratory research reports. Impressively, a multiple comparison or multiple testing adjustments (e.g. Bonferroni, Tukey, alpha spending functions) was performed in 47% of publications. Transformation of data (10%), missing data methods like multiple imputation or inverse probability weighting (4%), and

Table 1. Summary of Statistical Methods used over 2012–2017. P value from Fisher's exact test comparing A&A and Anesthesiology.

	<i>A & A</i>	<i>Anesthesiology</i>	Total	P value
	n (%)	n (%)	n (%)	
Number of articles	1,245	1,022	2,267	
Basic Science Research	304 (24)	426 (42)	730 (33)	<0.001*
Lab Animal Models	244 (20)	379 (37)	623 (27)	<0.001*
Clinical Research	940 (76)	581 (58)	1521 (67)	<0.001*
RCT	164 (13)	125 (12)	289 (13)	0.527
Noninferiority Trial	15 (1)	8 (1)	23 (1)	0.401
Randomized Crossover	10 (1)	23 (2)	33 (1)	0.005*
Meta-Analysis	58 (5)	31 (3)	89 (4)	0.051
PK/PD/Bioequivalence	65 (5)	70 (7)	135 (6)	0.109
Surveys or Questionnaires	85 (7)	69 (7)	154 (7)	0.999
Student's t-test	676 (54)	661 (65)	1337 (59)	<0.001*
Categorical Data Analysis	454 (36)	318 (31)	772 (34)	0.008*
Nonparametric Tests	589 (47)	446 (44)	1035 (46)	0.090
Pearson Correlation	115 (9)	101 (10)	216 (10)	0.615
Spearman Rank Correlation	82 (7)	62 (6)	144 (6)	0.665
Intraclass Correlation	24 (2)	19 (2)	43 (2)	0.999
Repeated Measures ANOVA	179 (14)	296 (29)	475 (21)	<0.001*
ANOVA - Not Repeated Measures	278 (22)	391 (38)	669 (30)	<0.001*
ROC Analysis	77 (6)	72 (7)	149 (7)	0.444
Linear Regression	142 (11)	129 (13)	271 (12)	0.398
Poisson Regression	29 (2)	16 (2)	45 (2)	0.227
Logistic Regression	213 (17)	178 (17)	391 (17)	0.867
Any Multiple Regression	303 (24)	238 (23)	541 (24)	0.586
Nonlinear/Other Regression	58 (5)	66 (6)	124 (5)	0.064
Survival Methods	78 (6)	100 (10)	178 (8)	0.002*
Cox Regression	41 (3)	58 (6)	99 (4)	0.007*
Longitudinal Analysis	294 (24)	365 (36)	659 (29)	<0.001*
Longitudinal Regression	366 (29)	453 (44)	819 (36)	<0.001*
Mixed Effects Models	168 (13)	140 (14)	308 (14)	0.902
Sample Size Considerations	571 (46)	450 (44)	1021 (45)	0.396
Multiple Comparisons Adjustment	475 (38)	586 (57)	1061 (47)	<0.001*
Transformation of Data	115 (9)	123 (12)	238 (10)	0.033*
Missing Data Methods	34 (3)	54 (5)	88 (4)	0.002*
Sensitivity Analysis	107 (9)	113 (11)	220 (10)	0.054
Bootstrap, Simulation, Validation	129 (10)	113 (11)	242 (11)	0.632
Propensity Score Matching	57 (5)	55 (5)	112 (5)	0.383
Bayesian Methods	7 (1)	12 (1)	19 (1)	0.163
Reliability of Measurements	59 (5)	57 (6)	116 (5)	0.389

*Statistically significant

sensitivity analyses (10%) were not too commonly found. Advanced statistical methods such as Bootstrap validation and simulation (11%), propensity score matching (5%), and Bayesian statistics (1%) were used.

Comparison Between the Journals

We were interested in comparing *A&A* and *Anesthesiology* regarding all study design features and statistical methods. A visual display of the anatomy of the statistical methods in published articles by journal from Table 1 is provided in Figure 1. Of the 2,267 publications that included inferential statistics, 730 (32%) were basic science articles and 1,537 (68%) were clinical research articles. *Anesthesiology* had a higher prevalence of basic science articles (42%) as compared to *A&A* (25%) ($P < 0.001$).

Since 2012, 426 (42%) basic science research articles were published in *Anesthesiology*, as compared to 304 (24%) in *A&A* ($P < 0.001$). There were 164 Randomized Controlled Trials published in *A&A*, while 125 were published in *Anesthesiology*, with rates of 13% and

12%, respectively ($P = 0.527$). Noninferiority trials (1% vs 1%), Meta-analyses (5% vs. 3%), PK/PD/Bioequivalence (5% vs 7%), and surveys or questionnaires (7% vs. 7%) were approximately equal in prevalence between the *A&A* and *Anesthesiology*, respectively (all $P > 0.05$, see Table 1).

Notably for statistical methods, power analysis and sample size and power considerations were done in *A&A* (46%) and *Anesthesiology* (44%) at about the same rate ($P = 0.363$). Student's t-test was used in 65% of *Anesthesiology* publications reviewed, compared to 54% of *A&A* publications ($P < 0.001$). Categorical data analysis methods were more commonly found in *A&A* (36%) than *Anesthesiology* (31%) ($P = 0.008$). ANOVA for repeated measures was found in *Anesthesiology* (29%) more often than in *A&A* (14%) ($P < 0.001$), as this is a commonly used statistical method in basic science research involving animal models. Additionally, adjustment for multiple comparisons or multiple testing was done more often in *Anesthesiology* (57%) as compared to *A&A* (38%) ($P < 0.001$), but this is likely related to the higher prevalence of ANOVA for repeated measured in *Anesthesiology* when adjusting for multiple pairwise post-hoc comparisons is common. There were not statistically significant differences in rates of use of the following statistical techniques between the journals: nonparametric/ rank statistics, Pearson correlation, Spearman rank correlation, Intraclass correlation, ROC analysis, all regression types, mixed-effects modelling, sensitivity analysis, bootstrapping, simulation or validation, propensity score matching, Bayesian methods, and reliability of measurements (all $P > 0.05$, Table 1).

Analysis of Trends from 2012 to 2017

Over these 5 years, many of the statistical methods have kept a relatively constant prevalence rate each year. Table 2 and Table 3 show the trends for all statistical methods recorded and Figure 2 displays the trends for select statistical methods. The following study design features and statistical methods have remained relatively constant in prevalence from 2012 through 2017: categorical data analysis methods (chi-square test, Fisher's exact test, or McNemar's test), Pearson correlation, Spearman rank correlation, multiple comparison adjustment, linear regression, nonlinear/other regression, longitudinal analysis, longitudinal regression, non-parametric tests (Wilcoxon, Mann-Whitney U, Kruskal-Wallis, Friedman, Kolmogorov-Smirnov, or Shapiro-Wilk), transformation of data, missing data methods, bootstrapping and simulation methods, surveys or questionnaire use, reliability of measurements, and ROC analysis (all $P > 0.05$, Table 2 and Table 3).

However, some statistical methods have increased or decreased in usage. Propensity score matching has increased in usage rate, from 2.8% in 2012 to 6.7% in 2017 ($P = 0.005$). This shows that in the anesthesia literature, propensity score matching is becoming an increasingly more popular statistical tool. Multiple regression is increasing in prevalence over time in these journals, from 20% in 2012 to 31% in 2017 ($P < 0.001$), as are survival methods and Cox regression, from 7% and 4% in 2012, respectively, to 12% and 8% in 2017, respectively ($P = 0.011$ and $P = 0.006$, respectively). The same is true for sensitivity analyses being conducted, with rates increasing from 5.7% in 2012 to 13% in 2017 ($P < 0.001$). The rate of publications reporting a sample size and power consideration increased over this 5 year span, from 34% in 2012 to 57% in 2017 ($P < 0.001$). Since the rate of these important and advanced statistical methods is increasing recently, this is an indication that the anesthesia literature is publishing high-quality and statistically thorough research articles. By 2020, for these statistical methodologies we project sample size and power considerations to be found in approximately

Table 2. Trends of Use of Study Design and Basic Statistical Methods. P values based on Mantel-Haenszel Chi-square test for trend.

	2012	2013	2014	2015	2016	2017	Total	P value
Number of articles	386	417	400	390	419	255	2,267	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
Basic Science Research	150 (39)	139 (33)	134 (34)	136 (35)	117 (28)	54 (21)	730 (33)	<0.001*
Lab Animal Models	131 (34)	116 (28)	115 (29)	119 (31)	96 (23)	46 (18)	623 (27)	<0.001*
Clinical Research	235 (61)	277 (66)	255 (64)	253 (65)	300 (72)	201 (79)	1521 (67)	<0.001*
RCT	50 (13)	50 (12)	52 (13)	54 (14)	55 (13)	28 (11)	289 (13)	0.847
Noninferiority Trial	5 (1)	4 (1)	4 (1)	1 (0)	7 (2)	2 (1)	23 (1)	0.832
Randomized Crossover	5 (1)	8 (2)	3 (1)	4 (1)	11 (3)	2 (1)	33 (1)	0.847
Meta-Analysis	10 (3)	14 (3)	16 (4)	14 (4)	23 (5)	12 (5)	89 (4)	0.047*
PK/PD/Bioequivalence	28 (7)	26 (6)	12 (3)	21 (5)	35 (8)	13 (5)	135 (6)	0.984
Surveys or Questionnaires	24 (6)	25 (6)	32 (8)	35 (9)	28 (7)	10 (4)	154 (7)	0.716
Student's t-test	255 (66)	255 (61)	237 (59)	237 (61)	224 (53)	129 (51)	1337 (59)	<0.001*
Categorical Data Analysis	120 (31)	136 (33)	138 (35)	134 (34)	145 (35)	99 (39)	772 (34)	0.055
Nonparametric Tests	166 (43)	202 (48)	181 (45)	178 (46)	199 (47)	109 (43)	1035 (46)	0.682
Pearson Correlation	46 (12)	38 (9)	41 (10)	35 (9)	40 (10)	16 (6)	216 (10)	0.053
Spearman Rank Correlation	25 (6)	24 (6)	25 (6)	25 (6)	34 (8)	11 (4)	144 (6)	0.982
Intraclass Correlation	4 (1)	6 (1)	12 (3)	7 (2)	9 (2)	5 (2)	43 (2)	<0.001*
Repeated Measures ANOVA	85 (22)	100 (24)	89 (22)	90 (23)	74 (18)	37 (15)	475 (21)	0.005*
ANOVA - Not Repeated Measures	147 (38)	129 (31)	122 (31)	108 (28)	117 (28)	46 (18)	669 (30)	<0.001*
ROC Analysis	22 (6)	28 (7)	19 (5)	26 (7)	38 (9)	16 (6)	149 (7)	0.201

*Statistically significant.

Table 3. Trends of Use of Statistical Modeling and Other Statistical Methods. P values based on Mantel-Haenszel Chi-square test for trend.

	2012	2013	2014	2015	2016	2017	Total	P value
Number of articles	386	417	400	390	419	255	2,267	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
Linear Regression	54 (14)	48 (12)	44 (11)	38 (10)	48 (11)	39 (15)	271 (12)	0.949
Poisson Regression	4 (1)	3 (1)	8 (2)	11 (3)	13 (3)	6 (2)	45 (2)	0.011*
Logistic Regression	58 (15)	70 (17)	59 (15)	68 (17)	81 (19)	55 (22)	391 (17)	0.017*
Any Multiple Regression	77 (20)	97 (23)	85 (21)	90 (23)	112 (27)	80 (31)	541 (24)	<0.001*
Nonlinear/Other Regression	17 (4)	23 (6)	19 (5)	28 (7)	24 (6)	13 (5)	124 (5)	0.444
Survival Methods	27 (7)	21 (5)	31 (8)	36 (9)	33 (8)	30 (12)	178 (8)	0.011*
Cox Regression	14 (4)	13 (3)	12 (3)	20 (5)	20 (5)	20 (8)	99 (4)	0.006*
Longitudinal Analysis	105 (27)	142 (34)	113 (28)	123 (32)	114 (27)	62 (24)	659 (29)	0.123
Longitudinal Regression	126 (33)	160 (38)	147 (37)	152 (39)	155 (37)	79 (31)	819 (36)	0.915
Mixed Effects Models	36 (9)	56 (13)	53 (13)	62 (16)	63 (15)	38 (15)	308 (14)	0.016*
Sample Size Considerations	133 (34)	156 (37)	169 (42)	213 (55)	205 (49)	145 (57)	1021 (45)	<0.001*
Multiple Comparisons Adjustment	189 (49)	205 (49)	190 (48)	193 (49)	187 (45)	97 (38)	1061 (47)	0.004*
Transformation of Data	36 (9)	49 (12)	38 (10)	56 (14)	48 (11)	11 (4)	238 (10)	0.380
Missing Data Methods	13 (3)	12 (3)	21 (5)	17 (4)	19 (5)	6 (2)	88 (4)	0.827
Sensitivity Analysis	22 (6)	35 (8)	35 (9)	43 (11)	53 (13)	32 (13)	220 (10)	<0.001*
Bootstrap, Simulation, Validation	41 (11)	47 (11)	41 (10)	40 (10)	46 (11)	27 (11)	242 (11)	0.923
Propensity Score Matching	11 (3)	19 (5)	15 (4)	23 (6)	27 (6)	17 (7)	112 (5)	0.005*
Bayesian Methods	4 (1)	4 (1)	2 (1)	3 (1)	4 (1)	2 (1)	19 (1)	0.752
Reliability of Measurements	13 (3)	22 (5)	19 (5)	29 (7)	29 (7)	4 (2)	116 (5)	0.585

*Statistically significant.

70% of publications, multiple regression to be performed in 35% of publications, survival methods to be used in 13% of articles, propensity score matching to be found in 9% of articles, and Meta-Analysis to be done in 7% of publications. These linear trend projections are depicted in Figure 3, with a linear trend line starting in 2012 and extrapolating beyond 2017 to 2020.

Two statistical methods that have decreased in usage in both *A&A* and *Anesthesiology* from 2012 to 2017 are the Student's t-test and ANOVA. In 2012, 66% of articles used the Student's t-test (or other parametric comparison of means between groups), where in 2017, this reduced to 51% ($P < 0.001$). ANOVA (repeated measures, factorial, or simple) was used in 52% of articles in 2012 but only 26% in 2017 ($P < 0.001$).

Discussion

The data show that anesthesia research publications from two of the premier journals in the field use a wide variety of statistical methods and are of high quality. Statistics such as Student's t-test, chi-square tests, ANOVA, and regression analyses are used often. Additionally, power analysis and multiple comparison adjustments are being done appropriately. Interestingly, only 1% of publications in these journals are economic or cost-effectiveness in nature. However, more advanced statistical methods have been used not only appropriately, but also well-done, including propensity score matching [10-13], Bayesian statistics [14-18], and bootstrapping/simulation/validation [19-22]. Our description of the current state of statistical methodology use shows

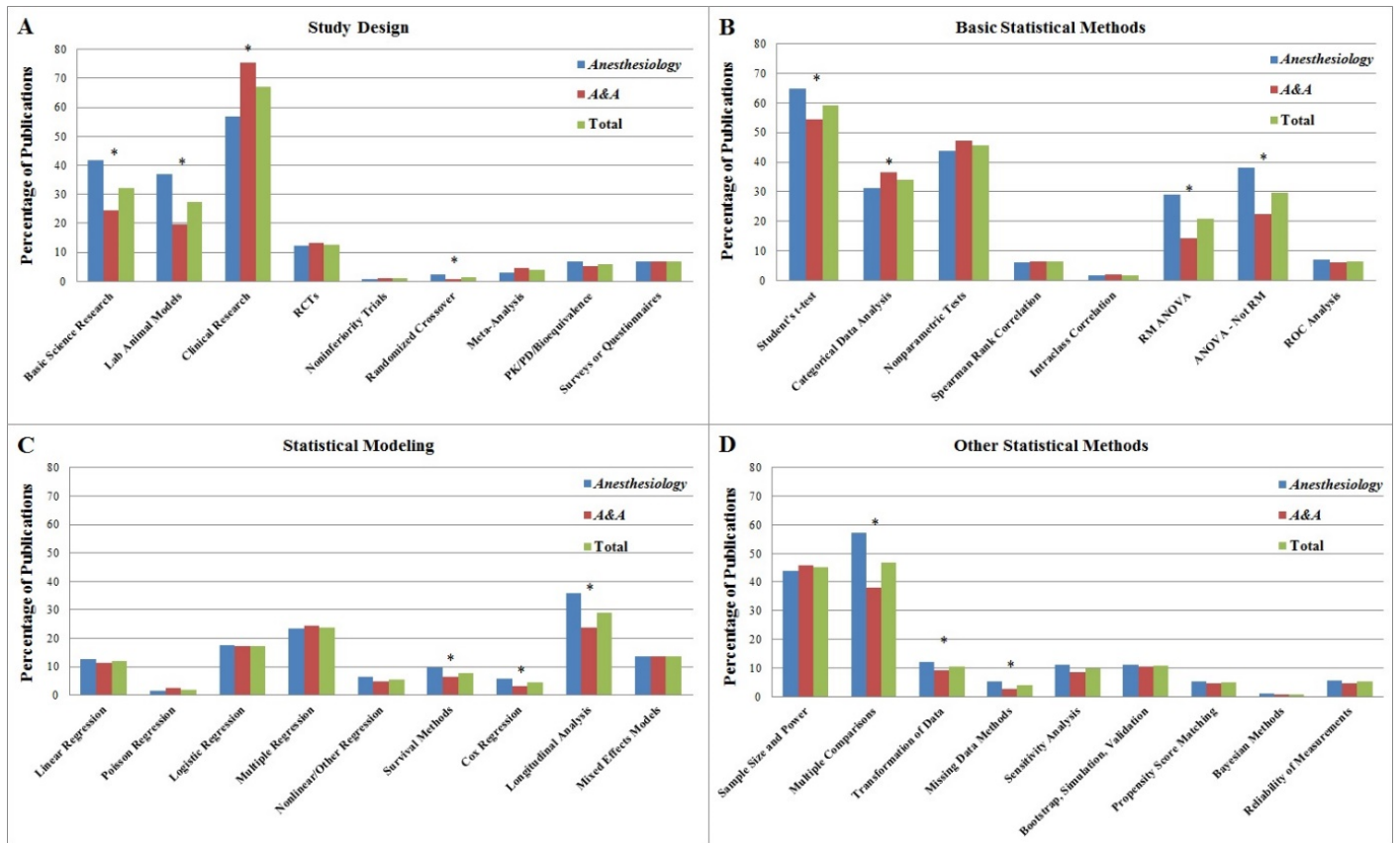


Figure 1. The anatomy of statistical methods in published articles. Percentage of articles from January 2012 through July 2017 by journal and combined overall using various A) study design features, B) basic statistical methods, C) statistical modeling methods, and D) other statistical methods. Statistically significant difference between *A&A* and *Anesthesiology* determined by Fisher's exact test is denoted by an asterisk. Meta-Analysis includes Systematic reviews and Meta-Analyses. RM ANOVA: Repeated Measures ANOVA; ANOVA - Not RM: ANOVA-Not Repeated Measures.

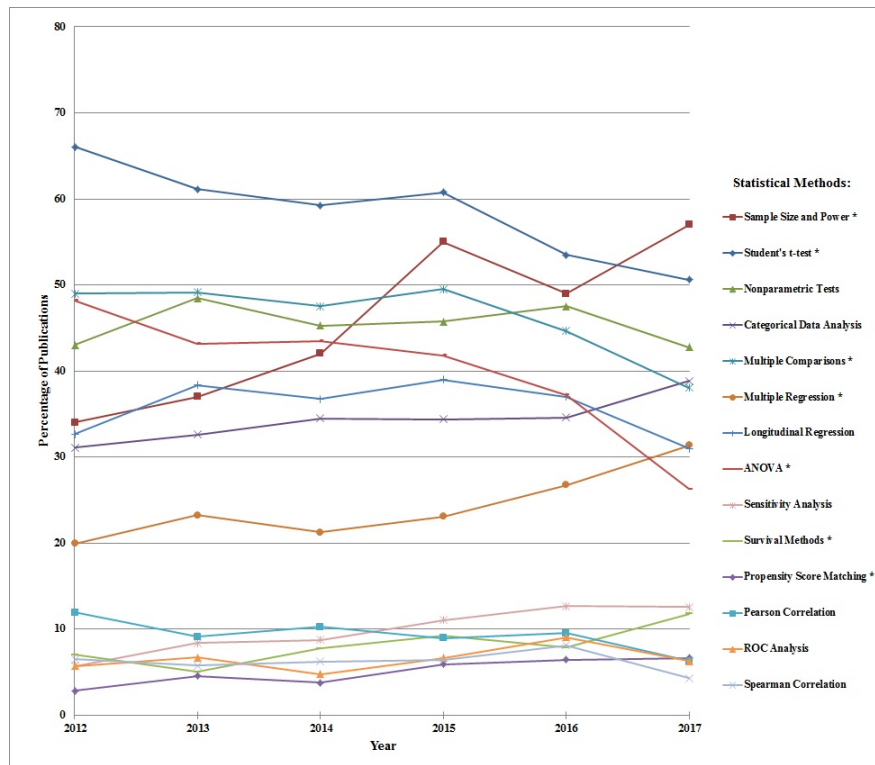


Figure 2. Trends of selected statistical methods from January 2012 through July 2017 for *A&A* and *Anesthesiology* combined. Statistically significant trend as determined by the Mantel-Haenszel Chi-square test is denoted by an asterisk next to the statistical method. ANOVA includes repeated measures, factorial, and simple ANOVA.

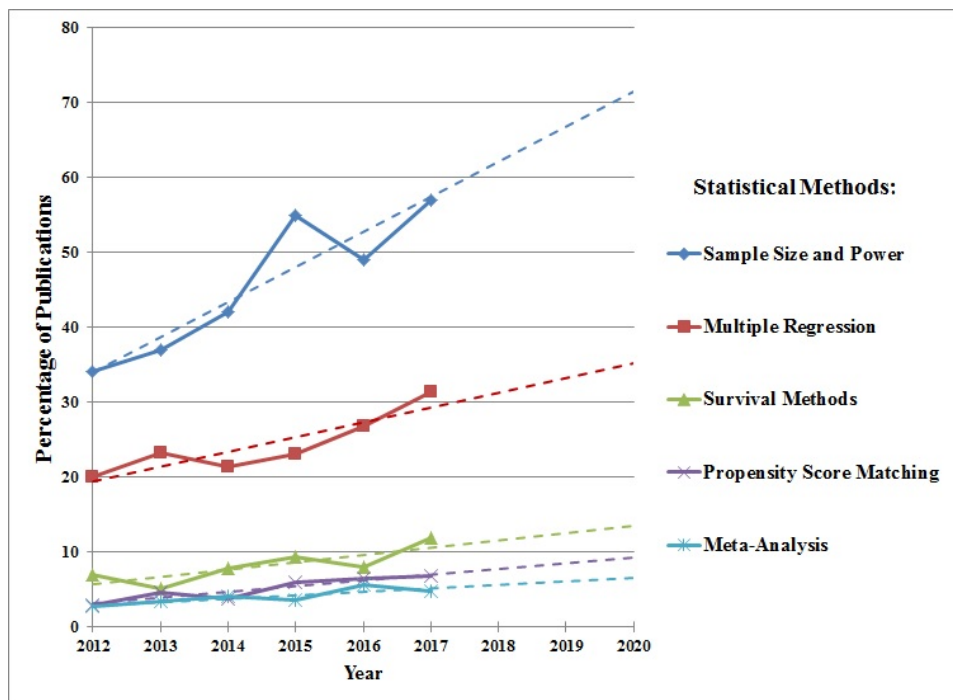


Figure 3. Linear trend forecasting through the year 2020 for select statistical methodologies for *A&A* and *Anesthesiology* combined. The selected methodologies displayed reached statistical significance in the Mantel-Haenszel Chi-square test for trend. Meta-Analysis includes systematic reviews and meta-analyses.

that statistical experience and expertise is present on research teams submitting manuscripts for publication in Anesthesia journals. We see an increasing trend in the rate of use of power analyses or sample size considerations, propensity score matching, multiple regressions, and sensitivity analyses. It is encouraging to see this increasing trend for these important advanced statistical methods. We also found a decreasing trend in the rate of Student's t-tests and ANOVAs. This could be explained by a shifting preference to more advanced statistical methods, or could be simply due to the nature of publications in these given years. Many basic science research articles with animal models used repeated-measures ANOVA, so the rate of use of ANOVA is influenced by the number of such articles.

Our study provides highly relevant information regarding the use of statistical methods that are an integral component of anesthesia research in advancing new knowledge in the specialty. Nevertheless, like all studies, ours has some limitations. There were a wide variety of publications reviewed with a broad range of topics and study designs as well as length of article. Since our data set includes articles which may differ in certain ways, the results we have generated must be interpreted with this in mind. For example, propensity score matching will only be used in the setting of a retrospective data analysis, so depending on how many such manuscripts are submitted to a journal will inherently influence the rate of the use of propensity score matching. Of course there are numerous other journals that publish anesthesia research articles, so while our dataset is very large and we included two premier journals, there are of course many more publications in the anesthesia literature. Since *A&A* and *Anesthesiology* are two of the leading journals in anesthesia, we believe that our results are representative of the gold standard for statistics in anesthesia research, and are generalizable to comparable anesthesia journals. Our projections are limited by the number of years of data we have collected in our database, going back to 2012. We assumed the trend of the rates to be linear over time, and we only projected for 3 years into the future because at a certain point it is unrealistic to look too far out. We chose to only obtain forecasts

for select statistical methods that showed a statistically significant trend and represent more advanced statistical practices.

Though similar to the study in the *New England Journal of Medicine* by Sato et. al, our study is the first to examine the use statistical methods in anesthesia literature, and it looks at 2,267 articles over a 5 year time window from January 2012 through July 2017. It is the first to thoroughly review and record the statistical methods in every research publication in *A&A* and *Anesthesiology*. This puts us in a unique place where we could describe the current trends of statistics in the anesthesia literature and compare utilization of statistical methods between *A&A* and *Anesthesiology*. This study is valuable as it informs the anesthesia research community of its current statistical applications, and may enable the researcher to improve the excellence of their research. It is important that readers and authors of anesthesia literature take a close look at the statistical methods being using in *A&A* and *Anesthesiology* to ensure that high quality research studies are being published with rigorous statistical methods.

Conclusions

Our review of publications from 2012 to 2017 in *Anesthesia & Analgesia* and *Anesthesiology* revealed the current status of the use of various statistical methodologies in anesthesia research. We found that *A&A* and *Anesthesiology* differ on several items, but together show an increasing trend in the rate of more specialized statistical techniques and greater statistical excellence. Our study demonstrates that a wide range of statistical methods are being utilized in anesthesia research, and that close collaboration is occurring between anesthesiologists and biostatisticians. Statistics plays an integral role in clinical and laboratory research, and we are encouraged by the evidence that suggests a spectrum of statistical methods being used, and more advanced techniques being introduced to the field. In the future we expect the statistical rigor in anesthesia journals to continue to grow, leading to increased new knowledge and scientific discoveries.

Financial Disclosures

None

Conflicts of interest

None

Author's individual contribution to the manuscript

Steven J. Staffa, M.S.

Contribution: This author collected the data, helped write the manuscript and conducted the statistical analysis.

David Zurakowski, M.S., Ph.D.

Contribution: This author proposed the topic/concept, helped write the manuscript and conducted the statistical analysis.

References

1. Sato Y, Goshio M, Nagashima K, Takahashi S, Ware JH, et al. (2017) Statistical Methods in the Journal - An Update. *N Engl J Med* 376: 1086-1087. [Crossref]
2. Egger M, Smith GD, Altman DG (2001) Systematic Reviews in Health Care Meta-analysis in Context. BMJ Publishing Group, London.
3. Giuliano KK, Scott SS, Elliot S, Giuliano AJ (1999) Temperature measurement in critically ill orally intubated adults: a comparison of pulmonary artery core, tympanic, and oral methods. *Crit Care Med* 27: 2188-2193. [Crossref]
4. Fletcher RH, Fletcher SW (2005) Clinical Epidemiology The Essentials. (4th Edn) Baltimore, MD: Lippincott Williams & Wilkins.
5. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE (2012) Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. (2nd Edn) Springer, New York, NY.
6. Hosmer DW, Lemeshow S, May S (2008) Applied Survival Analysis: Regression Modeling of Time-to-Event Data. (2nd Edn) Hoboken, NJ.
7. Fitzmaurice GM, Laird NM, Ware JH (2011) Applied Longitudinal Analysis. 2nd ed. Hoboken, NJ.
8. Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 46: 399-424.
9. Fraser C (2016) Business Statistics for Competitive Advantage with Excel 2016: Basics, Modeling Building, Simulation and Cases. New York.
10. O'Leary JD, Janus M, Duku R (2016) A population-based study evaluating the association between surgery in early life and child development at primary school entry. *Anesthesiology* 125: 272-279.
11. Kubota K, Egi M, Mizobuch S (2017) Haptoglobin administration in cardiovascular surgery patients: its association with the risk of postoperative acute kidney injury. *Anesth Analg* 124: 1771-1776.
12. Bateman BT, Tsen LC, Liu J, Butwick AJ, Huybrechts KF (2014) Patterns of second-line uterotonic use in a large sample of hospitalizations for childbirth in the United States: 2007-2011. *Anesth Analg* 119: 1344-1349. [Crossref]
13. Karkouti K, Callum J, Crowther MA, McCluskey SA, Pendergrast J, et al. (2013) The relationship between fibrinogen levels after cardiopulmonary bypass and large volume red cell transfusion in cardiac surgery: an observational study. *Anesth Analg* 117: 14-22. [Crossref]
14. Rothman BS, Shotwell MS, Beebe R (2016) Electronically mediated time-out initiative to reduce the incidence of wrong surgery: an interventional observational study. *Anesthesiology* 125: 484-494.
15. Bayman EO, Parekh KR, Keech J, Selte A, Brennan TJ (2017) A prospective study of chronic pain after thoracic surgery. *Anesthesiology* 126: 938-951.
16. Dexter F, Ledolter J, Davis E, Witkowski TA, Herman JH, et al. (2012) Systematic criteria for type and screen based on procedure's probability of erythrocyte transfusion. *Anesthesiology* 116: 768-778.
17. Stedman JL, Yarmush JM, Joshi MC, Kamath S, Schianodocola J (2017) How long is too long? the prespiked intravenous debate. *Anesth Analg* 124: 1564-1568.
18. Whitlock EL, Torres BA, Lin N (2014) Postoperative delirium in a substudy of cardiothoracic surgical patients in the BAG-RECALL Clinical Trial. *Anesth Analg* 118: 809-817.
19. Brueckmann B, Villa-Urbe JL, Bateman BT (2013) Development and validation of a score for prediction of postoperative respiratory complications. *Anesthesiology* 118: 1276-1285.
20. So-Osman C, Nelissen RG, Koopman-van Gemert AWMM (2014) Patient blood management in elective total hip- and knee-replacement surgery (part 1): a randomized controlled trial on erythropoietin and blood salvage as transfusion alternatives using a restrictive transfusion policy in erythropoietin-eligible patients. *Anesthesiology* 120: 839-851.
21. Maile MD, Engoren MC, Tremper KK, Tremper TT, Jewell ES, et al. (2016) Variability of automated intraoperative st segment values predicts postoperative troponin elevation. *Anesth Analg* 122: 608-615.
22. Olofsen E, Sigtermans M, Noppers I (2012) The dose-dependent effect of s(+)-ketamine on cardiac output in healthy volunteers and complex regional pain syndrome type I chronic pain patients. *Anesth Analg* 115: 536-546.

Copyright: ©2018 Staffa SJ. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.