Survey of the Use of Statistical Methods in *The International Journal of Prosthodontics*

Danielle Layton, BDSc, MSc, MDSc[a]

Purpose: This survey aimed to review how scientific articles were reported and to describe the types of statistical tests that had been recently and commonly used in *The International Journal of Prosthodontics*. ***Materials and Methods:*** All 174 articles published in 2012 and 2013 were hand-searched to identify scientific articles (n = 151) and those using at least one statistical test to explore results (n = 111). Editorials, letters, comments, erratum, and award proceedings were excluded. The number and type of statistical tests used within articles were collated, and the 10 commonly used methods were identified and described. ***Results:*** Of the 151 scientific articles, 76% (n = 111) used at least one statistical test and 24% (n = 40) used qualitative methods. Up to 10 tests were used per article, with 237 in total, of which 36 were unique. The 10 most commonly used tests were analysis of variance (ANOVA; n = 34), survival analyses (n = 29), Student *t* test (n = 19), chi-square (n = 19), Mann-Whitney *U* (n = 14), logistic regression (n = 13), Wilcoxon signed rank (n = 12), Fisher exact (n = 11), log-rank (n = 10), and Cox proportional hazards (n = 8), and they accounted for 71% (n = 169) of all tests used. ***Conclusions:*** The vast majority of articles published in recent years in *The International Journal of Prosthodontics* employed statistical analyses. Across 2 years, nearly 250 tests were completed, including 36 unique tests. Statistical test use was common but diverse. *Int J Prosthodont 2015;28:315–322. doi: 10.11607/Int J Prosthodont.2015.3.stat*

Statistics is considered by many to be both complex and tedious, with understanding of this basic science undermined by its relatively sparse incorporation into formal education systems. It is therefore not surprising that clinicians and researchers alike commonly delegate its interpretation to statistical scientists.

Statistics is more than just statistical tests, and its tenets support both qualitative and quantitative analytical methods. Statistics involves four main tasks: collecting, analyzing, presenting, and interpreting data, and its employment commences when the methods of studies are first considered. Without statistical sciences, advances in everyday clinical techniques would not be possible.

There is reluctance for non-statisticians to delve into the intricacies of this science, but not all general concepts in statistics need to be perplexing. In efforts to increase everyday statistical understanding, this paper has been prepared to focus on the second of the statistical tasks: the analysis. These analytical tests help explore data methodically. They help scientists ascertain whether patterns are random or systematic and whether findings are likely to be true.

This survey aimed to review how scientific articles were reported, and to describe the types of statistical tests that had been recently and commonly used in *The International Journal of Prosthodontics*.

Materials and Methods

Articles published in *The International Journal of Prosthodontics* in 2012 and 2013 were screened by hand by one researcher to identify scientific reports. Exclusion criteria were documents that were considered to be editorials, invited commentaries, letters, award proceedings, interviews, and errata.

Once identified, the scientific reports were reviewed to identify articles that used at least one statistical test to explore the results. Exclusion criteria were case reports and articles exploring outcomes qualitatively.

The abstract, methods, and results of articles reporting quantitative results were read, and the details of how authors explored their data were gathered. Data collated included the number of statistical tests used in each article and the frequency of use of individual statistical tests across articles.

The findings of this survey have been explored qualitatively. The descriptions of the 10 most commonly used quantitative methods have been provided. The appropriateness of use of the tests and the quality of reporting was not assessed.

[a]*Private Practice, Brisbane, Australia.*

Correspondence to: *Dr Danielle M. Layton, 217 Wickham Terrace, Brisbane, Queensland, 4000, Australia.*
Email: laytonpros@dlpros.com.au

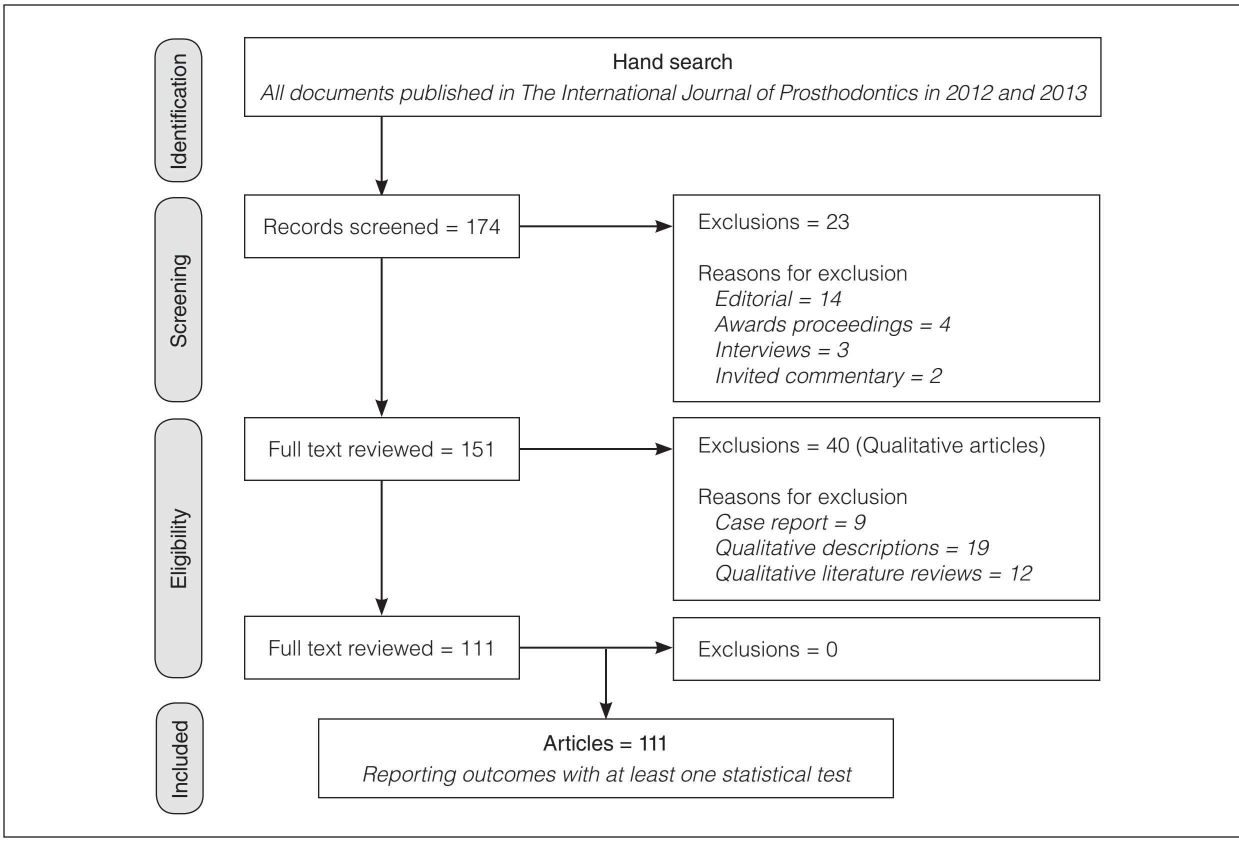

Fig 1 Flowchart of article search.

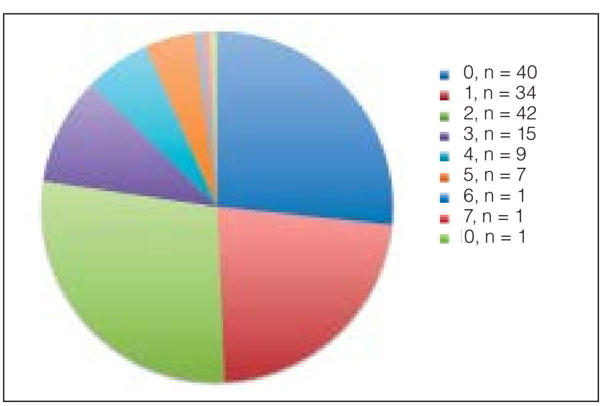

Fig 2 Number of statistical tests used per article across 2012 and 2013.

Results

One hundred seventy-four documents were screened, with 151 retained for further review. Exclusions included 14 editorials, 4 award proceedings, 3 interviews, and 2 invited commentaries. Of the 151 articles, 76% (n = 111) used at least one statistical test to explore the data and 24% (n = 40) used qualitative reporting methods (Fig 1).

Within the 151 articles, between 0 and 10 statistical tests were used, and on average the median was 2 (IQR 0 to 2) (Fig 2).

Among the 40 articles that explored data qualitatively, there were 9 case reports, 12 qualitative literature reviews, and 19 qualitative reports. Many of the qualitative reports used percentages and numbers to describe some outcomes, but no statistical tests were used to explore hypotheses.

Of the remaining 111 articles, 74% used at least one statistical test to explore data. Across these articles, 237 individual tests were used, and these tests comprised 36 different statistical methods (Fig 3). The 10 most commonly used tests were analysis of variance (ANOVA; n = 34), survival analyses (n = 29), Student *t* test (n = 19), chi-square test (n = 19), Mann-Whitney *U* test (n = 14), logistic regression (n = 13), Wilcoxon signed rank test (n = 12), Fisher exact test (n = 11), log-rank (n = 10), and Cox proportional hazards model (n = 8). These accounted for 71% (n = 169) of all tests used, with the remaining 27 tests contributing to the final total (n = 68 out of 237 tests). There is no standard vocabulary to name statistical tests, and several methods were described by different names in different articles.

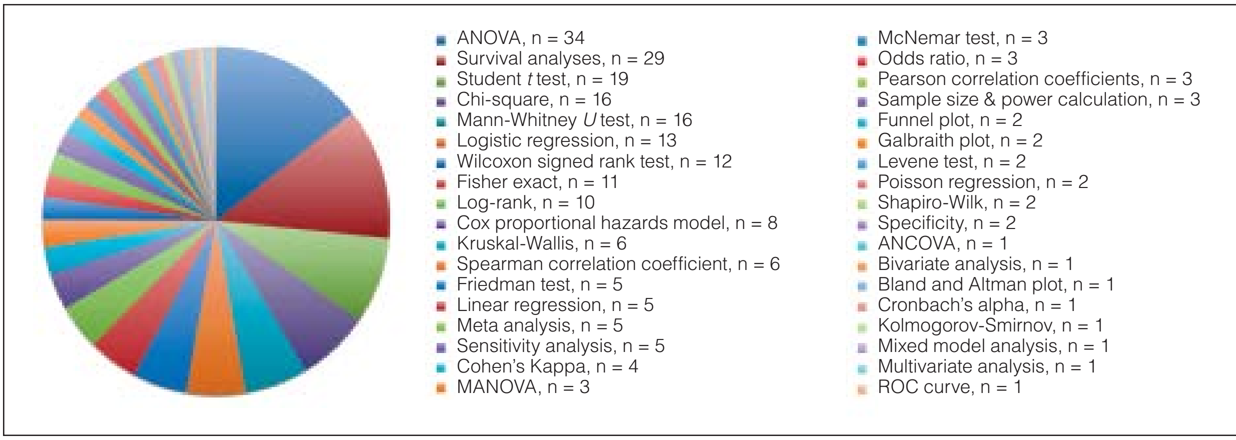

Fig 3 Frequency of use of individual statistical tests across all articles in 2012 and 2013.

Following initial quantitative analysis, various outcomes were then explored with additional post-hoc assessments. This occurred most commonly with the suite of ANOVA tests, with post hoc methods including Tukey (n = 8), Scheffe (n = 2), Bonferroni (n = 2), Student-Newman-Keuls (n = 2), Tamhane (n = 2), and Fisher least significant difference test (n = 1).

Descriptions of the 10 most common methods are provided below and summarized in Table 1. Relationships between these analyses, samples, and data types are outlined in Table 2, and general considerations for reporting are outlined in Table 3.

ANOVA (n = 34 articles)

ANOVA is a range of statistical analyses that can compare the means of more than two groups at the same time. If only two means are compared, the results of ANOVA are the same as the results from the Student *t* test (two independent samples). If the means of multiple groups are tested, two at a time, there is an increased chance that statistical type I error will occur. Therefore, ANOVA is used instead.

ANOVA is used for parametric data. The nonparametric equivalent to the one-way ANOVA is the Kurskal-Wallis test.

Types of ANOVA include one-way (with one independent variable), two-way (with two independent variables), three-way (with three independent variables), and repeated measures. Post-hoc tests to adjust for multiple comparisons are commonly used with ANOVA, because all pairwise comparisons are often tested. Common post-hoc assessments include Tukey, Scheffe, Bonferroni, Student-Newman-Keuls, Tamhane, and Fisher least significant difference test.

For example, Zou and colleagues[1] used the ANOVA method to assess differences in peri-implant parameters between three groups: telescopic crowns, bars, and locator attachments that were supporting overdentures in the maxilla. The outcome measures were continuous, with the mean values for each group compared.

Survival Analyses (n = 29 articles)

Survival analyses are also known as time-to-event analyses. The two most commonly used techniques are Kaplan-Meier and life table analyses. In these analyses, *(1)* events are monitored, *(2)* over time, and *(3)* an estimated cumulative proportion is calculated for these events. In the context of dental research, events such as failures of prostheses may be monitored over time, and a statistic such as the estimated cumulative survival is reported. Events are transitions from one state to another. They can relate to a broad range of medical and nonmedical topics such as time-to disease, time-to recurrence, time-to death, time-to recovery, time-to equipment failure, time-to earthquakes, time-to stock market crash.

For example, Layton and Walton[2] used Kaplan-Meier survival methods to explore the outcome of feldspathic porcelain veneers over 21 years. The variable of interest was binary, survival versus failure, and it was followed over time.

Student t Test (n = 19 articles)

The Student *t* test is a group of parametric tests that assess whether the "test statistic" follows the Student *t* distribution. It is used for small samples, such as those with less than 100 subjects. Large samples use the *Z* test. There are four common tests:

- One-sample *t* test: Assessment of the mean of one group and whether this mean has a value specified by a null hypothesis

Table 1 Summary of the 10 Most Commonly Used Statistical Tests

1.	**ANOVA (n = 34 articles)** Parametric data (normally distributed) More than two groups Can be used with more than one independent variable Multiple comparisons are undertaken and thus post-hoc assessments are often required Data reported as means bounded by a confidence interval Analysis should be reported as F statistic with degrees of freedom (for between and within groups), mean square error, and *P* value; if used, post hoc tests should be acknowledged
2.	**Survival analyses (n = 29 articles)** Binary data (the event happens, or it doesn't) This is not a statistical test, there is no *P* value Data reported as an estimated cumulative proportion (the estimated cumulative survival, ECS) bounded by a confidence interval or standard error; reporting as "survival rate" is incorrect Graphic data should be reported as life tables and survival curves
3.	**Student *t* test (n = 19 articles)** Parametric test Continuous data Small samples Can be used for one-sample, two-sample, or a paired-sample Data reported as means bounded by a confidence interval or standard error Analysis should be reported as the t statistic with degrees of freedom and *P* value
4.	**Chi-square test (n = 16 articles)** Categorical data (for example, female and male) Data reported as proportions bounded by a confidence interval Analysis should be reported as the chi-square statistic with degrees of freedom and *P* value
5.	**Mann-Whitney *U* test (n = 14 articles)** Nonparametric data (not normally distributed) Comparison of two groups of continuous data Numbers are converted to "ranks," and the actual numerical value is discarded Data reported as medians bounded by the interquartile ranges Analysis should be reported as the *U* statistic with degrees of freedom, *Z* value, and *P* value
6.	**Logistic regression (n = 13 articles)** Assesses whether variables can predict an outcome of a categorical variable Data are transformed by a logarithm function and therefore need to be transformed back again Results are best presented in a table; outcomes should be reported as crude and adjusted odds ratios bounded by the confidence interval as well as the fit of the model (goodness of fit) and the significance of the model (*P* value)
7.	**Wilcoxon signed rank test (n = 12 articles)** Nonparametric data Continuous data Matched/paired samples Data reported as median bounded by the interquartile range for each category Analysis should be reported as W statistic with degrees of freedom, *Z* value, and *P* value
8.	**Fisher exact test (n = 11 articles)** Categorical data Chi-square distribution Small sample size (resulting in an expected count of at least one cell being less than 5) Data reported as proportions bounded by a standard error Analyses should be reported as chi-square statistic with degrees of freedom and *P* value
9.	**Log-rank test (n = 10 articles)** Two groups Survival data (time-to-event data) Data reported as an estimated cumulative survival that is bounded by the confidence interval or standard error Analysis should be reported as the chi-square statistic, *P* value
10.	**Cox proportional hazards model (n = 8 articles)** Assesses whether risk factors can predict an outcome of a categorical variable (such as survival) Data reported as an estimated cumulative survival bounded by the confidence interval or standard error Analysis should be reported as crude and adjusted hazard ratios bounded by the confidence interval as well as the fit of the model (goodness of fit) and the significance of the model (*P* value)

- Two-sample (independent) *t* test: Assessment of the means of two groups and whether those groups are from the same population; the groups should be independent and unpaired
- Paired-sample *t* test: Assessment that the means of two responses (such as before and after treatment) from the same subject has a mean value of zero
- Regression: Testing whether the slope of the regression line differs from zero

For example, Hamilton and colleagues[3] compared the fit of titanium CAD/CAM abutments with titanium prefabricated abutments across five implant types with the independent sample *t* test. The variable of interest was continuous, with the marginal gap measured in micrometers and the mean of each group then compared.

Chi-Square Test (n = 16 articles)

The chi-square test is commonly used for cohort studies with two categorical variables. The relationship between each variable can be represented in a 2 × 2 table with columns and rows. For example, one variable may be smoking (yes or no) while the other variable may be periodontal disease (yes or no). Chi-square tests assess whether there is a relationship between the variables, or whether those variables are independent. It tests whether there is a difference between the frequency that an event is observed to occur (the observed frequency) and the frequency that the event was expected to occur if there was no difference between the groups (the expected frequency).

For example, amongst other tests, Gjengedal and colleagues[4] used the chi-square test to assess differences in dietary intake between patients with implant-retained versus conventionally relined dentures. Variables assessed between the two groups were categorical, such as number of patients avoiding at least one food item. When the expected frequency of patients in at least one category is less than 5, then the Fisher exact test rather than the chi-square test should be employed. The authors stated that both the chi-square and the Fisher

Table 2 Relationships Between Analysis, Sample Considerations, and Data Types

	1 sample	2 samples	More than 2 samples	Relationships and predictions (regression)
Categorical	N/A	*Large samples* Chi-square *Small samples* Fisher exact	Chi-square	Logistic regression
Continuous	*Parametric* One-sample *t* test	*Parametric* Two-sample *t* test *Nonparametric* Mann-Whitney *U* test	*Parametric* ANOVA *Nonparametric* Kurskal-Wallis	Linear regression
Paired	N/A	*Parametric* Paired *t* test *Nonparametric* Wilcoxon signed rank test	N/A	N/A
Categorical, over time	Survival analyses (life table, Kaplan-Meier)	Log-rank test	N/A	Cox proportional hazards model

exact tests were used during analysis, but it was unclear from the reporting which test was used to explore which variables.

Mann-Whitney U Test (n = 14 articles)

This is a nonparametric test used when data are not normally distributed. It assesses whether two groups have come from the same population. It is not a simple test of differences in medians, as the spread of the data also influences the results. It is the nonparametric equivalent to the *t* test for two normally distributed samples. When more than two groups are involved, the Kruskal-Wallis test is used. The Mann-Whitney *U* test is also known as the Mann-Whitney-Wilcoxon, the Wilcoxon rank-sum, and the Wilcoxon-Mann-Whitney test. It is not the same as the Wilcoxon signed-rank test. A *U* statistic is calculated.

The numbers (the "results") are converted to ranks (smallest to largest), and the actual numerical value of the information is discarded. This allows comparison of nonparametric data. However, as these data are discarded, the test is not as powerful as a parametric test to assess whether the "groups" come from the same population. Therefore, data are often transformed, such as a log transformation, to see if the transformed data become parametric to allow use of the *t* test (for parametric data) rather than the Mann-Whitney *U* test (the nonparametric equivalent).

For example, Menicucci and colleagues[5] used the Mann-Whitney *U* test to compare the insertion torque and insertion time of straight-walled and tapered implants when placed in patients. The variables of interest were continuous and compared between the two groups.

Table 3 What to report?

Characteristics of the data Measure of central tendency (mean, median, mode) Variance (confidence interval, standard error) Number in each group or subgroup, and if this changes over time Graphics (life tables, survival curves, box plots, histograms)
Causal relationships, with the strength and direction of the relationship Test statistic (chi-square statistic, F statistic, *t* statistic, *U* statistic, W statistic, *Z* statistic, Wald statistic) Degrees of freedom, where appropriate Strength of association (confidence intervals, *P* values)
Relationships between statistical significance and clinical relevance Consider whether a result is clinically significant, regardless of its statistical significance Consider whether a statistically significant result is clinically relevant

Logistic Regression (n = 13 articles)

Logistic regression assesses whether there is a relationship between a categorical variable (dependent variable) and one or more independent variables (continuous or categorical data). The dependent variable can have two categories (binary/binomial logistic regression, such as no and yes responses), multiple categories (multinominal logistic regression), or multiple ordered categories (ordinal data, ordered logistic regression).

The dependent variable is transformed, and the regression analysis works with the natural logarithm of the odds, known as the logit or log-odds. This transforms the binominal data into continuous data, and an analysis similar to linear regression can then be conducted. Linear regression analyzes the least squares to find a best fitting line, while logistic regression estimates the probability of an event occurring (the ratio of the odds of an event occurring to it not occurring). Following assessment, the data is converted back with the inverse log, the exponential function of the log-odds, the exp(B). Multiple tests are conducted on the data, such as the Hosmer-Lemeshow test to assess the goodness-of-fit of the logistic regression model and the

Wald statistic to assess whether individual predictors are statistically significant contributors to the model. Data can be entered into the models in various orders, such as all together (entered), forward, backward, and stepwise.

Data should be reported first as crude odds ratios (the odds of an individual predictor being related to the dependent variable), and adjusted odds ratios (the odds calculated when multiple predictors are included together in the model and are related to the dependent variable).

For example, Listl[6] explored factors associated with denture wearing in older populations in multiple European countries. The outcome (dependent) variable was categorical and dichotomous: denture wearing (yes or no). The independent variables included those such as age, sex, chewing ability, dental insurance coverage, and socioeconomic status.

Wilcoxon Signed Rank Test (n = 12 articles)

The Wilcoxon signed rank test is a nonparametric test to assess paired samples, such as "before" and "after" measurements. It tests whether the median difference between the pairs of observations is zero. This is different from the paired *t* test (mean difference between the pairs is zero) and the sign test (the numbers of differences in each direction are equal). It is the nonparametric equivalent to the matched paired *t* test. The data recorded are continuous but divided into nominal groups (such as subjects and before/after).

The absolute difference between the paired samples is calculated, and then these are ranked (smallest to largest). The ranks of the differences in one direction are added, and the ranks of the differences in the other direction are added, the W test statistic is assessed against the Wilcoxon tables, and a *P* value is reported.

For example, Micarelli and colleagues[7] used the Wilcoxon signed rank test to assess differences in the reverse torque of abutment screws before and after various cleaning methods. The outcome of interest was continuous, but related as before and after measurements.

Fisher Exact Test (n = 11 articles)

The Fisher exact test assesses whether there is a relationship between two categorical samples. The relationship between each variable (such as disease and treatment) can be represented in a 2×2 table with columns and rows. It is similar to the chi-square test, but it is used when the expected count of at least one of the cells in the contingency tables is less than 5. It tests whether there is a difference between the frequency that an event is observed to occur (the observed frequency) and the frequency that the event was expected to occur if there was no difference between the groups (the expected frequency).

For example, Sanita and colleagues[8] explored the management of denture stomatitis when patients were prescribed microwave denture disinfection versus nystatin medication. Between the two groups, the authors used the Fisher exact test to explore whether there was a difference in the categorical outcome: the proportion of patients with no recurring symptoms versus those experiencing recurrence.

Log-Rank Test (n = 10 articles)

The log-rank test assesses the survival distributions of two groups and whether there are differences between their time-to-event outcomes (their estimated cumulative proportions). It is used when censored data (such as loss to follow-up over time) is present. It is incorrect to use general categorical or nonparametric tests to assess groups where not all subjects made it to the end of the study (censored observations).

The log-rank test compares the hazard function (the measure of the tendency to "fail," the instantaneous failure rate) of each group, at each point in time, and then assesses differences in the observed and expected number of events across all time points. The hazard function is also known as the hazard rate.

For example, Walton[9] used the log-rank test to assess differences in the outcome of vital and nonvital high noble ceramometal crowns over a 25-year period. The difference in the binary outcome of (failure: yes, no) was assessed between the two groups over time.

Cox Proportional Hazards Model (n = 8 articles)

The Cox proportional hazards model assesses whether there is a relationship between risk and predictive factors on an event (a categorical dependent variable, such as "failure") over time.

The Cox proportional hazard regression model is similar to logistic regression models, but it is more appropriate when outcomes occur over time and censoring occurs. When the duration of the study is short and the number of events (such as failures) is rare, both models produce similar results.

For example, Beier and colleagues[10] explored the relationship between failure of all ceramic restorations over 20 years and possible risk factors. The dependent variable was categorical and dichotomous: failure (yes, no). The independent variables included potential risk factors of tooth vitality, parafunction, type of cement, type of restoration, and tooth restored.

Discussion

This research has found that over three-quarters of articles published in recent years of *The International Journal of Prosthodontics* used at least one statistical analysis, with 36 specific types of statistical tests employed. This proportion is remarkably similar to the 76% of articles identified in a similar survey of the *Japanese Journal of Clinical Oncology*.[11]

It was found that use of statistical techniques was common but also diverse. It is unlikely that readers or individual reviewers would be familiar with all types of these tests and indeed it is possible that authors were unfamiliar with their own statistics. Over the past few decades, much has been written about reporting quality, but there has been less focus specifically on the reporting of the statistics. A new resource for statistical reporting, rather than study design, became available in 2013 in a book,[12] and on the EQUATOR network.[13] Given the volume and diversity of tests that have been employed, such a resource is much needed.

Across the scientific articles in this survey, 40 explored data qualitatively. Qualitative methods are inductive, rather than deductive, and focus on how findings relate to social and natural environments.[14] These methods can be used to explore why a difference may be present rather than how much of a difference can be observed.

Across the remaining 111 scientific articles, data were explored quantitatively with a total of 237 tests of which 36 were unique. Across the top 10 common analyses, the data types were spread evenly amongst parametric (n = 2), nonparametric (n = 2), categorical (n = 2), categorical over time (n = 2), and predictive (n = 2). These 10 suites of tests accounted for 70% of those conducted, which initially appears surprisingly specific. However, the remaining tests could have been any one of an additional 27, showing great diversity in methods used, and making it difficult for readers and reviewers to have sufficient knowledge to understand all articles.

Of interest, the second most common method, survival analysis, is technically not a statistical test. It arguably straddles quantitative and qualitative reporting, exploring progress of subjects or prostheses over time with ongoing hypothesis testing and using different statistics (eg, log rank or Cox proportional hazards).

When reviewing the tests employed, it was found that different names were commonly used to describe the same analysis (for example, factorial ANOVA and two-way ANOVA). It is not always clear why the names of statistical tests vary. In some cases names relate to the scientists who first published the formula but are then extrapolated to include historical scientists who were originally involved in the mathematics. An example of this includes the Mann-Whitney *U* versus the Mann-Whitney-Wilcoxon test. Alternatively, names of other tests evolved, as different authors suggested more precise descriptors. An example of this includes the independent samples *t* test versus the two-samples *t* test. Although using multiple names is interesting from a historical perspective, it adds to levels of statistical confusion.

Organizations, libraries, societies, and universities, including the International Statistical Institute (ISI), have composed dictionaries of statistical terms. The ISI provides a *Multilingual Glossary of Statistical Terms,*[15] which collates different test names across 31 different languages for over 3,500 terms. This improves our ability to identify overlapping nomenclature, but to date no consensus of standardized nomenclature has been proposed.

When statistical tests were employed, articles used between 1 and 10 different tests. It can be tempting to believe that articles using a larger number of statistical tests are of higher quality than those using a smaller number, or those using no tests at all. Multiple tests can be legitimately useful, improving reporting quality and external applicability. This can involve assessing data usability, testing assumption criteria, or incorporating ongoing sensitivity and subgroup analyses. Conversely, the opposite may well be true, with additional tests being less than useful. For example, the more tests completed, and the more samples tested, the higher the likelihood that statistical significance is found by random chance—data dredging.

This survey did not assess the suitability of the reporting methods used or the quality of such reporting. However, it is clearly incumbent on authors to ensure that outcomes are reported with clarity and accuracy, and equally incumbent on reviewers to assist this process.

Finally, this article cautions: Do not confuse the use of complex statistical tests with a significantly important article; do not confuse statistical significance with clinical significance; and do not confuse a lack of statistical significance with a lack of clinical relevance.

Conclusions

The vast majority of articles published in recent years in *The International Journal of Prosthodontics* employed statistical analyses. Across 2 years, nearly 250 tests were completed, including 36 unique tests. Statistical test use was common but diverse.

References

1. Zou D, Wu Y, Huang W, et al. A 3-year prospective clinical study of telescopic crown, bar, and locator attachments for removable four implant-supported maxillary overdentures. Int J Prosthodont 2013;26:566–573.
2. Layton DM, Walton TR. The up to 21-year clinical outcome and survival of feldspathic porcelain veneers: Accounting for clustering. Int J Prosthodont 2012;25:604–612.
3. Hamilton A, Judge RB, Palamara JE, et al. Evaluation of the fit of CAD/CAM abutments. Int J Prosthodont 2013;26:370–380.
4. Gjengedal H, Dahl L, Lavik A, et al. Randomized clinical trial comparing dietary intake in patients with implant-retained overdentures and conventionally relined denture. Int J Prosthodont 2012;25:340–347.
5. Menicucci G, Pachie E, Lorenzetti M, et al. Comparison of primary stability of straight-walled and tapered implants using an insertion torque device. Int J Prosthodont 2012;25:465–471.
6. Listl S. Denture wearing by individuals among the older segment of European populations. Int J Prosthodont 2012;25:15–20.
7. Micarelli C, Canullo L, Baldissara P, et al. Implant abutment screw reverse torque values before and after plasma cleaning. Int J Prosthodont 2013;26:331–333.
8. Sanita PV, Machado AL, Pavarina AC, et al. Microwave denture disinfection versus nystatin in treating patients with well-controlled type 2 diabetes and denture stomatitis: a randomized clinical trial. Int J Prosthodont 2012;25:232–244.
9. Walton TR. The up to 25-year survival and clinical performance of 2340 high gold-based metal-ceramic single crowns. Int J Prosthodont 2013;26:151–160.
10. Beier US, Kapferer I, Dumfahrt H. Clinical long-term evaluation and failure characteristics of 1,335 all-ceramic restorations. Int J Prosthodont 2012;25:70–78.
11. Fukuda H, Ohashi Y. A guideline for reporting results of statistical analysis in Japanese Journal of Clinical Oncology. Jpn J Clin Oncol 1997;27:121–127.
12. Lang TA, Altman DG. Basic Statistical Reporting for Articles Published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or The SAMPL Guidelines.” In: Smart P, Maisonneuve H, Polderman A, eds. Science Editors’ Handbook. European Association of Science Editors, 2013.
13. Lang TA, Altman DG. SAMPL guidelines for statistical reporting. http://www.equator-network.org/2013/02/11/sampl-guidelines-for-statistical-reporting/. Updated February 2013. Accessed November 2014.
14. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311:42–45.
15. The International Statistical Institute. ISI Multilingual Glossary of Statistical Terms. http://www.isi-web.org/glossary/introduction. Updated June 2011. Accessed November 2014.