



Reporting data analysis methods in high-impact respiratory journals

Pentti Nieminen¹, Tuula Toljamo² and Hannu Vähänikkilä¹

Affiliations: ¹Medical Informatics and Data Analysis Research Group, University of Oulu, Oulu, Finland. ²Dept of Pulmonary Diseases, Lapland Central Hospital, Rovaniemi, Finland.

Correspondence: Pentti Nieminen, Medical Informatics and Data Analysis Research Group, University of Oulu, Aapistie 5A, 90240 Oulu, Finland. E-mail: pentti.nieminen@oulu.fi

ABSTRACT Data analysis methods play an important role in respiratory research. We evaluated the application and complexity of data analytical methods in high-impact respiratory journals and compared the statistical reporting in these respiratory articles with reports published in other eminent medical journals.

This study involved a total of 160 papers published in 2015 in the *European Respiratory Journal*, *American Journal of Respiratory and Critical Care Medicine*, *Chest* and *Thorax*, and 680 papers published between 2007–2015 in other medical journals including the *Lancet* and *New England Journal of Medicine*. We manually reviewed the articles to determine the way in which they reported the methods applied in data analysis.

The statistical intensity in the respiratory journals was equal to that in eminent medical journals. Traditional ways of testing statistical significance were widely used in respiratory articles. Statistical procedures were not always described in sufficient detail, and the prominent respiratory journals did not display different profiles with respect to their statistical content.

Readers of the prominent respiratory journals need to possess a substantial level of statistical expertise if they wish to critically evaluate the design, methodology, data analysis and interpretation of the findings published in these journals.



@ERSpublications

Statistical data analysis methods play an important role in prominent respiratory journals

<http://ow.ly/DMQz30k7PPS>

Cite this article as: Nieminen P, Toljamo T, Vähänikkilä H Reporting data analysis methods in high-impact respiratory journals. *ERJ Open Res* 2018; 4: 00140-2017 [<https://doi.org/10.1183/23120541.00140-2017>].



Introduction

The medical journal is one of the primary channels for communicating medical information to fellow physicians. The amount of information being published has been exponentially increasing [1, 2]. Healthcare professionals who are anxious to maintain their professional skills are forced to decide which items they will focus on and those to be ignored. To be an effective consumer, the journal reader needs to be familiar with methodological issues, especially when elaborate techniques, such as statistical procedures, are invoked to clarify findings or to draw conclusions from raw data. Statistical methods also play an important role in medical publications [3]. This is reflected in the high proportion of articles that are essentially statistical in character. In fact, most papers published in medical journals contain some elements of statistical methods, analysis and interpretation [4, 5].

Respiratory healthcare professionals clearly recognise the importance of biostatistical and data-analytical concepts [6, 7]. However, there has been no evaluation of which statistical methods are adopted in respiratory research papers. Earlier reviews of respiratory research have indicated the need for improved methodologies [8, 9] and this should be based on a systematic review of the present status of respiratory research.

Because of the importance of critical reading of the medical literature, it is essential to include an awareness of statistics in respiratory (undergraduate and postgraduate) training. In addition, clinicians and graduate readers of medical journals should appreciate the frequency with which various statistical concepts are reported in journals that are important in their sub-fields. This can help readers to identify the major statistical skills needed to critically evaluate the published literature [10].

Table 1 lists the commonly used methods to evaluate statistical significances in medical research as presented in textbooks of medical statistics [11–14]. However, this table does not encompass by any means all of the statistical techniques applied in medical reports. Respiratory studies often involve longitudinal data where individuals are measured repeatedly over time and the relationship between an outcome variable and several explanatory variables is analysed. New statistical techniques have been developed to analyse longitudinal relationships allowing use of all of the available longitudinal data [15, 16]. In addition, it is often necessary to evaluate the agreement of the raters involved in classifying subjects into predefined categories, or to measure the reliability of developed questionnaires or test instruments. During recent decades, mathematical statisticians have introduced new data analytical methods compatible with the rapid expansion in computing capability [17]. Bayesian methods, artificial neural networks and machine learning are some examples of these more sophisticated approaches [18]. However, it is unclear how widely these methods are being applied in medical domains [10].

The *New England Journal of Medicine (NEJM)* and the *Lancet* are two of the most eminent medical interdisciplinary journals. These journals have a wide range of readers, and they publish high-quality

TABLE 1 Basic statistical methods used in medical research by research goal and type of outcome variable

Research goal	Type of outcome variable			
	Measurement from symmetric distribution	Measurement from very skewed distribution	Categorical variable	Time to event
Describing one variable	Mean±SD	Median, interquartile range	Proportion	Kaplan–Meier curve
Comparing two independent groups	Comparing means with independent samples t-test	Comparing medians with Mann–Whitney test	Cross-tabulation with Chi-squared test	Kaplan–Meier curves and log-rank test
Comparing three or more independent groups	Comparing means with one-way ANOVA	Comparing medians with Kruskal–Wallis test	Cross-tabulation with Chi-squared test	Kaplan–Meier curves and log-rank test
Comparing two repeated measurements	t-test for repeated measurements	Wilcoxon test	McNemar’s test	
Comparing three or more repeated measurements	Repeated-measures ANOVA	Friedman test		
Quantifying association between two variables	Pearson correlation	Spearman correlation	Cross-tabulation with Chi-squared test, risk ratio or odds ratio statistics	
Explaining variation with several explanatory variables	Multivariable linear regression	Negative binomial regression	Logistic regression	Cox proportional hazard regression

Reproduced and modified from [10].

research findings in an understandable and clinically useful format. Hence, comparisons between the journals publishing research articles in respiratory medicine and these two eminent general medical journals would help respiratory researchers to improve their use of data analysis methods. As far as we are aware, no studies have assessed recent practices in the use of statistical methods in articles published in respiratory journals in comparison with how they are utilised in these two eminent general medical journals.

Our aim in this study was to evaluate the application and complexity of data analytical methods in high-impact respiratory journals, and to compare the statistical methods and reporting in these respiratory articles with reports published in other eminent medical journals. In addition, we aimed to examine differences in study designs and sample sizes between the journals.

Material and methods

Set of articles

This study collated the statistical techniques used and statistical reporting in 840 articles appearing in the 14 journals listed in table 2. The collection represents a total of 160 articles published in the 2015 issues of four respiratory journals (*European Respiratory Journal (ERJ)*, *American Journal of Respiratory and Critical Care Medicine (AJRCCM)*, *Chest* and *Thorax*); these journals are cornerstones of respiratory medicine and are read by physicians working in this speciality. These respiratory journals were selected because they have consistently been among the most visible and widely cited respiratory journals.

For comparison purposes, we used original research articles published between 2007 and 2015 in 10 journals. For the first group, we selected two eminent medical journals (*Lancet* and *NEJM*). For the second group (selected medical journals), we chose eight journals from different sub-fields (*Journal of Dental Research*, *Journal of Dentistry*, *Cell*, *International Journal of Epidemiology*, *European Journal of Public Health*, *American Journal of Psychiatry*, *Scandinavian Journal of Public Health* and *Research in Nursing and Health*) to be included in this evaluation. We chose these journals to validate inferences about the wide range of statistical reporting in medicine and related fields. We analysed a total of 320 papers published between 2007 and 2017 in the *Lancet* and *NEJM*, and 360 papers published in the other journals from 2009 to 2016.

Several articles have reviewed medical papers and reported that the proportion of articles adequately using statistical methods lies in the range of 50–70% [19–22]. Assuming a 60% rate in respiratory papers, a sample size of 100 respiratory articles was calculated to be the minimum number of articles required for the present purpose, allowing a maximum difference of 10% units between the sample rate and the true population rate at a 95% significance level. However, we anticipated that 25 articles per journal would be insufficient to make comparisons between the journals and therefore we decided to increase the number to 40 articles per journal. The starting articles were chosen randomly, the only criteria being that there would

TABLE 2 Respiratory and other medical journals surveyed for use of statistical techniques

	Impact factor 2016	Publication year(s) surveyed	Number of articles
Respiratory journals			
<i>American Journal of Respiratory and Critical Care Medicine</i>	13.204	2015	40
<i>Chest</i>	6.044	2015	40
<i>European Respiratory Journal</i>	10.569	2015	40
<i>Thorax</i>	8.272	2015	40
Eminent general medical journals			
<i>New England Journal of Medicine</i>	72.406	2007, 2008, 2010, 2017	160
<i>Lancet</i>	47.831	2008, 2009, 2011, 2017	160
Selected medical journals			
<i>American Journal of Psychiatry</i>	14.176	2011, 2016	80
<i>Cell</i>	30.410	2010	40
<i>European Journal of Public Health</i>	2.431	2009	40
<i>International Journal of Epidemiology</i>	7.738	2011	40
<i>Journal of Dental Research</i>	4.755	2010	40
<i>Journal of Dentistry</i>	3.456	2010	40
<i>Research in Nursing and Health</i>	1.693	2013	40
<i>Scandinavian Journal of Public Health</i>	1.339	2015	40

be at least 39 eligible subsequent articles published that year in the journal in question. Letters, brief reports, narrative reviews and editorials were excluded from this sample.

Classification of the study design and sample size

We classified the articles into three groups: experimental studies (including randomised clinical trials and animal studies), non-experimental studies (including cross-sectional surveys, cohort studies and case-control studies) and others (including reliability, methodology and basic science studies). We then divided the studies into four groups based on sample size: under 30, 30–99, 100–300 and over 300.

Categories of statistical procedures

We manually reviewed all 840 papers for their statistical content. We used the Statistical Intensity of Medical Articles (SIMA) tool [10] to collect information about the statistical methods and reporting. The SIMA instrument has been developed to assess the statistical intensity of an article or manuscript. The current version of the instrument includes 67 items pertaining to the description of statistical and data management procedures, applied statistical methods, and reporting of methods and results, but the instrument does not include items listing statistical errors. In the instrument, the items are categorised into 16 groups. These sub-groups are denoted with capital letters (from A to P) on the evaluation form. Each group includes items measuring the usage of specific statistical analysis methods or reporting styles. We have calculated the sums of sub-group items and a total score by summing all 67 items. We have designated the total score as the statistical intensity score and it ranged from 0 to 74. However, in practice, values over 30 are very rare. A paper with several outcomes and explanatory variables, application of multivariable methods, overuse of p-values and confidence intervals, and a very high number of tables and figures can be awarded a high SIMA score, but readers with a medical background might find it difficult to read. An article with a low value of the statistical intensity score means that it has hardly used any statistical methods (e.g. laboratory studies or narrative studies). The reliability of the statistical intensity score has previously been shown to be high, the interobserver agreement and test-retest reliability as measured by the intraclass correlation coefficient are >0.80 [10].

Statistical analysis

The Chi-squared test was applied to evaluate the statistical significant differences in the use of statistical procedures between the journal groups. Analysis of variance was used to reveal the statistically significant differences for the statistical intensity score between the journal groups. Tukey's *post hoc* test was used for pairwise comparisons. IBM SPSS Statistics version 24 (IBM Corp., Armonk, NY, USA) was used in the data analysis.

Results

Table 3 shows the study design of the evaluated articles. Most of the respiratory articles (67.5%) had applied an observational approach. An experimental design was used in 16.8% of the respiratory articles and in 16.7% of the selected medical articles. Experimental clinical trials were more common (54.1%) in the *Lancet* and *NEJM* articles. Sample sizes were lower both in respiratory studies and other medical studies when compared with the *Lancet* and *NEJM* articles: the sample size exceeded 300 in 64.4% of the articles published in the eminent general medical journals, but in only 41.8% of the respiratory articles.

The reported statistical significance of the primary outcome of articles in the four respiratory journals is summarised in table 3. All the eligible articles were classified according to whether or not the authors had reported or decided that the primary outcome was statistically significant or nonsignificant. If the results of formal statistical significance testing or confidence interval estimation were not included in the reporting of the primary finding, the article was classified as not having evaluated statistical significance. We found that 128 (80.0%) of respiratory articles reported a statistical test that had been conducted on the primary outcome. Of these, 106 (82.8%) reported a statistically significant result. This proportion was similar in the selected medical subject-specific journals (218 (84.2%) out of 259 articles). However, this proportion differed from the eminent general journals (table 3), with 162 (69.2%) out of 234 articles published in the *Lancet* and *NEJM* reporting statistically significant results. The proportion of studies that report significant or positive results seems to be higher in the respiratory journals than in the eminent general journals (*Lancet* and *NEJM*).

The statistical methods and reporting used in the journals are shown in table 4. About one third of the respiratory papers provided an incomplete description of their statistical procedures. This serious reporting problem seems to be prevalent even in eminent medical journals. *Lancet* and *NEJM* gave references to the statistical literature most often, whereas the respiratory journals more frequently detailed which statistical software had been used.

TABLE 3 Distribution of study design, sample size and main outcome by the evaluated journals

	AJRCCM	ERJ	Chest	Thorax	NEJM+Lancet	Selected medical journals	All articles
Total articles n	40	40	40	40	320	360	840
Study design							
Cross-sectional survey	14 (35.0)	10 (25.0)	10 (25.0)	12 (30.0)	43 (13.4)	107 (29.7)	196 (23.3)
Longitudinal cohort study	14 (35.0)	10 (25.0)	11 (27.5)	10 (25.0)	49 (15.3)	62 (17.2)	156 (18.6)
Case-control	4 (10.0)	7 (17.5)	3 (7.5)	3 (7.5)	7 (2.2)	18 (5.0)	42 (5.0)
Interventional study (clinical trial)	3 (7.5)	7 (17.5)	6 (15.0)	11 (27.5)	173 (54.1)	60 (16.7)	260 (30.9)
Laboratory work	4 (10.0)	5 (12.5)	1 (2.5)	2 (5.0)	7 (2.2)	69 (19.2)	88 (10.5)
Meta-analysis	0	1 (2.5)	5 (12.5)	1 (2.5)	15 (4.7)	18 (5.0)	40 (4.8)
Other	1 (2.5)	0	4 (10.0)	1 (2.5)	26 (8.1)	26 (7.2)	58 (6.9)
Sample size							
<30	4 (10.0)	8 (20.0)	7 (17.5)	5 (12.5)	34 (10.6)	65 (18.1)	120 (14.3)
30-99	9 (22.5)	9 (22.5)	7 (17.5)	9 (22.5)	27 (8.4)	62 (17.2)	123 (14.6)
100-300	12 (30.0)	7 (17.5)	9 (22.5)	7 (17.5)	39 (12.2)	45 (12.5)	117 (13.9)
>300	15 (37.5)	16 (40.0)	17 (42.5)	19 (47.5)	206 (64.4)	145 (40.3)	423 (50.4)
Missing	0	0	0	0	14 (4.4)	43 (11.9)	57 (6.8)
Statistical significance of primary outcome							
Not significant	6 (15.0)	4 (10.0)	7 (17.5)	5 (12.5)	72 (22.5)	41 (11.4)	135 (16.1)
Significant	30 (75.0)	28 (70.0)	21 (52.5)	27 (67.5)	162 (50.6)	218 (60.6)	486 (57.8)
Not evaluated	4 (10.0)	8 (20.0)	12 (30.0)	8 (20.0)	86 (26.9)	101 (28.0)	219 (26.1)

Data are presented as number of articles (%), unless stated otherwise. AJRCCM: American Journal of Respiratory and Critical Care Medicine; ERJ: European Respiratory Journal; NEJM: New England Journal of Medicine.

The percentage frequencies of the statistical procedures used in the three samples of published medical research articles highlights the broad use of statistical methods (table 4). The procedures most commonly reported in the respiratory papers were a comparison of independent groups in its various forms and regression modelling. Almost 70% of the articles compared frequencies, mean values, median values or time-to-event curves between groups.

During recent decades, mathematical statisticians have introduced new, more complex methods that are attributable to the rapid expansion in computing capability, for example, Bayesian methods, artificial neural networks and machine learning. In practice, applied statisticians and medical researchers have applied these tools very rarely; we found only one reference to these methods in any of the evaluated respiratory articles.

TABLE 4 Statistical methods and reporting in the evaluated journals

	Respiratory journals	NEJM+Lancet	Selected medical journals	All articles	p-value of Chi-squared test
Total articles n	160	320	360	840	
Statistical methods described	100 (62.5)	219 (68.4)	221 (61.4)	540 (64.3)	0.140
Reference to statistical literature	38 (23.8)	144 (45.0)	104 (28.9)	286 (34.0)	<0.001
Statistical software reported	126 (78.8)	191 (59.7)	215 (59.7)	532 (63.3)	<0.001
Comparing groups	111 (69.4)	221 (69.1)	199 (55.3)	531 (63.2)	<0.001
Repeated measurements	18 (11.3)	35 (10.9)	36 (10.0)	89 (10.6)	0.884
Correlation coefficient methods	38 (23.8)	25 (7.8)	75 (20.8)	138 (16.4)	<0.001
Regression models	96 (60.0)	205 (64.1)	165 (45.8)	466 (55.5)	<0.001
Other multivariable methods	9 (5.6)	13 (4.1)	35 (9.7)	57 (6.8)	0.010
Intra-cluster correlation methods	24 (15.0)	58 (18.1)	57 (15.8)	139 (16.5)	0.614
Measures of agreement	27 (16.9)	31 (9.7)	69 (19.2)	127 (15.1)	0.002
Bayesian methods	1 (0.6)	5 (1.6)	4 (1.1)	10 (1.2)	0.712
ANN or machine learning	0	2 (0.6)	0	2 (0.2)	0.181

Data are presented as n (%), unless stated otherwise. NEJM: New England Journal of Medicine; ANN: artificial neural networks.

The respiratory journals contained references to a set of statistical techniques that were very similar to those used in the *Lancet* and *NEJM*. One major difference was that the *Lancet* and *NEJM* utilise correlation coefficient methods and measure of agreement analyses less often than these are applied in the respiratory journals.

The distribution of statistical intensity is shown in figure 1. The mean \pm SD statistical intensity score for the respiratory journals was 17.2 \pm 5.6, for the *Lancet* and *NEJM* it was 17.9 \pm 6.7 and for the selected medical journals it was 13.4 \pm 6.8. The mean values differed statistically significantly ($p < 0.001$). The statistical intensity was higher in the respiratory articles compared with the selected medical journals article sample (p -values for Tukey's *post hoc* test was < 0.001). We were unable to detect any statistically significant difference between articles published in the respiratory articles compared with articles published in the *Lancet* and *NEJM* (Tukey's p -value=0.432). In the respiratory journals, the mean \pm SD of statistical intensity score varied as follows: *AJRCCM* 17.5 \pm 5.3, *Chest* 16.5 \pm 5.5, *ERJ* 17.8 \pm 5.8 and *Thorax* 16.8 \pm 6.0, *i.e.* the statistical intensity was at the same level in these prominent respiratory journals (p -value=0.720).

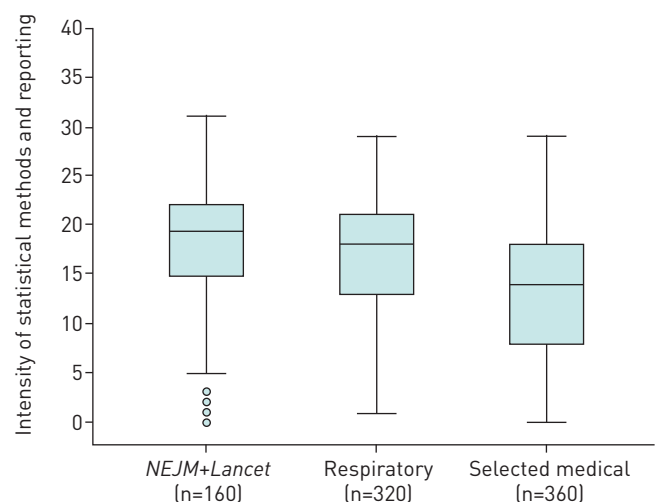
Discussion

We have analysed the statistical intensity of four prominent respiratory journals. We estimated the frequency with which statistical concepts were mentioned and how data analysis methods were reported in the published papers. The present study shows that articles published in the prominent respiratory journals were applying and reporting statistical methods partly differently from articles in selected medical subject-specific journals. Readers of respiratory journals more often encountered basic data analytical methods and regression models than readers of the selected subject-specific journals. However, the statistical intensity was at the same level as in the two eminent medical journals analysed (the *Lancet* and *NEJM*). Our findings suggest that a reader familiar with basic data analysis methods could successfully understand about 95% of the quantitative methods contained in the articles reviewed.

Our results provide up-to-date evidence for the widespread use of statistics in the prominent respiratory journals. Up to 80% of the reviewed respiratory papers displayed some kind of analytical character, applying diverse statistical inferential methods. This proportion was even higher than in the eminent general medical journals. In addition, articles published in the prominent respiratory journals reported statistically significant findings for primary outcomes more often than those in the *Lancet* and *NEJM*. This is an alarming phenomenon, because researchers are likely to have performed several different types of statistical evaluations before finding one capable of producing statistically significant results. In addition, since statistically significant results for the main outcome are more likely to be published in journals, it is possible that publication bias will be a problem when interpreting medical data [23, 24]. Our findings confirm previous reports [25, 26] that the proportion of significant results in published medical studies is high; specifically that the proportion of research reports that reject the null hypothesis ($p < 0.05$) ranges from 71% to 97%. For example, we found that 83.3% of the evaluated articles in *AJRCCM* had used significance testing and reported the main outcome findings with " $p < 0.05$ ".

Almost half of the articles in the *Lancet* and *NEJM* referred to the statistical literature, whereas only 23.8% of the articles in the respiratory journals had such a reference. A research report fails in its task of informing the reader if there are no references to details specific to certain data analysis methods.

FIGURE 1 Intensity of statistical methods and reporting subdivided by the journal groups. *NEJM*: *New England Journal of Medicine*. The horizontal line in the middle of the box is the median value of the scores and the lower and upper boundaries indicate the 25th and 75th percentiles, respectively. The box plot also displays outliers, articles with values more than 1.5 box-lengths from the lower edge of the box are designated with a circle. The largest and smallest observed values that are not outliers are also shown; lines (whiskers) are drawn from the ends of the box to those values.



Furthermore, as well as applying adequate statistical methods, it is essential to describe the statistical methods that have been employed in enough detail. Articles with clearly documented research methods do receive more citations [20]. A complete description of statistical procedures can help readers to evaluate and understand the findings; a knowledgeable reader can even recalculate important study findings. The reporting of statistical information was adequate in only ~64% of the evaluated articles. We suggest that this needs to be improved. Appropriate statistical reporting should include a paragraph on the analytical methods to which the data was subjected; this should be described in the methods section in sufficient detail so that the reader could reproduce the calculations with the available raw data.

There were differences between the journal groups regarding the use of statistical methods. The most frequently used statistical techniques were traditional statistical methods for comparing independent groups (Chi-squared test, t-test, ANOVA and nonparametric tests). These methods were followed by regression methods. Articles in the prominent respiratory journals and eminent general medical journals applied these methods more often than was the case in the other selected medical journals. However, articles in respiratory and other medical subject-specific journals utilised correlation coefficients and agreement coefficients more often than in the eminent general journals. Our findings are in line with earlier findings that eminent medical general journals predominantly publish studies investigating human subjects and make use of risk measures such as relative risk, risk difference or odds ratio when handling categorical data [5, 21].

Previous studies have shown that the use of statistical methods and reporting practices varies between journals, even within medical subfields [20, 27–29]. It is clear that statistical assessments differ between basic, clinical and epidemiological research. Basic science relies on basic analyses, whereas clinical and epidemiological studies require the application of multivariate analyses to adjust for possible confounders. The smaller sample sizes associated with animal studies further lessen the possibility of applying more sophisticated statistical techniques. In some medical sub-fields or related disciplines (public health, health science, nursing and psychology), other multivariable methods such as factor analysis, structural equation models and cluster analysis are also being adopted. Classifying subjects or objects into predefined classes or categories is a rather common activity in certain sub-fields like respiratory medicine, radiology, psychiatry or dentistry. Journals in these sub-fields often publish articles evaluating the agreement between different raters *via* reliability coefficients.

Basic data analysis methods are more familiar to readers of the medical literature, and in addition, it is perhaps easier to publish a manuscript with well-known methods [20]. While some new methods are occasionally introduced in medical research, only regression models and methods for analysing clustered data, where repeated measurements are made in the same individuals over time or individuals are nested within groups, have been more widely accepted by medical researchers [4, 21]. The broader introduction and acceptance of a new analysis method (as useful as the method might be) into medical publications seems to require the method being incorporated into the standard statistical packages generally used by medical researchers. In addition, if readers do not understand the mathematics or reporting style, or the conclusions have been drawn on the basis of advanced mathematics or computationally complex data mining procedures not visible in the data (tables or graphs) presented, then clinicians may not be convinced of the results.

Conclusions

This study provides new information on the status of statistical methods applied for analysing data in journals devoted to respiratory research. Our findings suggest that readers of respiratory journals need to be more familiar with data-analysis methods, including multivariable regression models, than readers of some selected medical subject-specific journals. We encourage all educators in respiratory medicine to review their programmes to ensure that an appropriate level of data analysis capacity is provided to their students, clinicians and colleagues. We also encourage all readers, authors and reviewers of respiratory journals who wish to be more effective consumers of the respiratory literature to review their own statistical skills and to present their results in a manner similar to that advocated and presented in the prominent respiratory journals.

Conflict of interest: None declared.

References

- 1 Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J Assoc Inf Sci Technol* 2015; 66: 2215–2222.
- 2 Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics* 2010; 84: 575–603.

- 3 Petrovecki M. The role of statistical reviewer in biomedical scientific journal. *Biochemia Medica* 2009; 223–230.
- 4 Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med* 2005; 353: 1977–1979.
- 5 Sato Y, Goshi M, Nagashima K, et al. Statistical methods in the journal – an update. *N Engl J Med* 2017; 376: 1086–1087.
- 6 Chatburn RL. Overview of respiratory care research. *Respir Care* 2004; 49: 1149–1156.
- 7 Chinn S. Statistics for the *European Respiratory Journal*. *Eur Respir J* 2001; 18: 393–401.
- 8 Wolfe R, Abramson MJ. Modern statistical methods in respiratory medicine. *Respirology* 2014; 19: 9–13.
- 9 Barquera S, Rico-Méndez FG, Tovar V. Methodology in the epidemiological research of respiratory diseases and environmental pollution. *Rev Saude Publica* 2002; 36: 107–113.
- 10 Nieminen P, Virtanen JI, Vahanikkilä H. An instrument to assess the statistical intensity of medical research papers. *PLoS One* 2017; 12: e0186882.
- 11 Armitage P, Berry G, Matthews JNS. *Statistical Methods in Medical Research*. Oxford, Blackwell Science, 2002.
- 12 Campbell MJ, Machin D. *Medical Statistics. A Commonsense Approach*. Chichester, Wiley, 2003.
- 13 Motulsky H. *Intuitive Biostatistics*. New York, Oxford University Press, 2014.
- 14 Bland M. *An Introduction to Medical Statistics*. Oxford, Oxford University Press, 2015.
- 15 Diggle P, Heagerty P, Liang K, et al. *Analysis of Longitudinal Data*. Oxford, Oxford University Press, 2014.
- 16 Twisk JWR. *Applied Longitudinal Data Analysis for Epidemiology*. Cambridge, Cambridge University Press, 2013.
- 17 Veierod MB, Lydersen S, Laake P. *Medical Statistics in Clinical and Epidemiological Research*. Oslo, Gyldendal, 2012.
- 18 Ertel W. *Introduction to Artificial Intelligence*. London, Springer, 2011.
- 19 Altman DG. Statistics in medical journals: some recent trends. *Stat Med* 2000; 19: 3275–3289.
- 20 Nieminen P, Carpenter J, Rucker G, et al. The relationship between quality of research and citation frequency. *BMC Med Res Methodol* 2006; 6: 42.
- 21 Strasak AM, Zaman Q, Marinell G, et al. The use of statistics in medical research: a comparison of “*The New England Journal of Medicine*” and “*Nature Medicine*”. *Am Stat* 2007; 61: 47–55.
- 22 Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. *Biochem Med (Zagreb)* 2015; 25: 5–11.
- 23 Song F, Eastwood AJ, Gilbody S, et al. Publication and related biases. *Health Technol Assess* 2000; 4: 1–115.
- 24 Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc Ser A* 1988; 151: 419–463.
- 25 Nieminen P, Rucker G, Miettunen J, et al. Statistically significant papers in psychiatry were cited more often than others. *J Clin Epidemiol* 2007; 60: 939–946.
- 26 Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995; 49: 108–112.
- 27 Miettunen J, Nieminen P, Isohanni M. Statistical methodology in major general psychiatric journals. *Nord J Psychiatry* 2002; 56: 223–228.
- 28 Vähänikkilä H, Tjaderhane L, Nieminen P. The statistical reporting quality of articles published in 2010 in five dental journals. *Acta Odontol Scand* 2015; 73: 76–80.
- 29 Vähänikkilä H, Virtanen J, Nieminen P. How do statistics in dental articles differ from those articles published in highly visible medical journals? *Scientometrics* 2016; 108: 1417–1424.