Received: 11 February 2019 | Revised: 3 April 2019 | Accepted: 12 April 2019
DOI: 10.1002/mpr.1784

ORIGINAL ARTICLE

WILEY

Reporting of data analysis methods in psychiatric journals: Trends from 1996 to 2018

Pentti Nieminen[1] | **Jasleen Kaur**[2]

[1] Medical Informatics and Data Analysis Research Group, University of Oulu, Oulu, Finland

[2] Department of Internal Medicine, Hurley Medical Center/Michigan State University, Flint, Michigan

Correspondence
Pentti Nieminen, Medical Informatics and Data Analysis Research Group, University of Oulu, Oulu 90014, Finland.
Email: pentti.nieminen@oulu.fi

Abstract

Objectives: The article aims to evaluate how study designs and data analysis methods in psychiatric studies have changed over the last 22 years.

Methods: This study involved a total of 320 papers published in 1996 and 2018 in the American Journal of Psychiatry, Acta Psychiatrica Scandinavica, British Journal of Psychiatry, and JAMA Psychiatry. We manually reviewed the articles to determine the way in which they reported the study characteristics and the methods applied in data analysis.

Results: The statistical intensity in psychiatric journals has changed over the past 20 years. Traditional methods of testing statistical significance were widely used both in 1996 and in 2018. In 2018, there was an increase in reporting more complex methods, such as multivariable regression models, multilevel modelling, and intracluster correlation methods. However, computationally complex data mining or machine learning procedures were not adopted by psychiatric researchers.

Conclusion: The increase in statistical intensity in the literature suggests that readers of prominent psychiatric journals must possess a substantial level of statistical expertise if they wish to critically evaluate the findings published in these journals. It is essential to include an awareness of this substantial change in data analysis methods in psychiatric undergraduate and postgraduate education.

KEYWORDS
data analysis, psychiatry, publications, statistics

1 | INTRODUCTION

The most important function of any medical journal is the effective dissemination of research findings to its target audience. The amount of information being published has been exponentially increasing (Bornmann & Mutz, 2015; Larsen & von Ins, 2010). Researchers and practitioners who are determined to maintain their professional skills are forced to decide which items they will focus on and which will be ignored. To evaluate the scientific merit of a published article, readers should be familiar with the methodology described and especially with the kinds of issues that can arise when quantitative data analysis techniques are utilized to clarify findings or to draw conclusions from raw data. Statistical methods also play an important role in medical publications (Petrovecki, 2009). 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, or interpretation (Horton & Switzer, 2005; Sato et al., 2017).

Information literacy is a crucial skill in medical education today. Medical education has placed emphasis on evidence-based medicine and has taught students and residents to critically evaluate the literature and then use this evidence in clinical expertise to make diagnostic and treatment decisions (Straus, Glaziou, Richardson, & Haynes, 2018). It is important to include an awareness of data analysis methods in medical (undergraduate and postgraduate) training. This can help clinicians and graduate student readers of medical journals to identify the major statistical skills needed to critically evaluate the published literature.

There have not been many systematic assessments of data analysis methods in psychiatry. Previous evaluations include bibliometric studies of general psychiatric journals (Hokanson et al., 1986; Miettunen, Nieminen, & Isohanni, 2002; Nieminen, 1995; Pincus, Henderson, Blackwood, & Dial, 1993). Hokanson et al. (1986) inventoried the statistical techniques described in 15 major psychiatric journals between 1983 and 1984. A dozen procedures, typically encountered in intermediate-level statistics courses, accounted for approximately 95% of all statistical methods reported. Traditional research with systematic data collection and statistical data analysis was the most common type of study in research reports published in Medline indexed journals. Miettunen et al. (2002) described the frequency with which various data analysis designs and methods were reported in 448 articles published in four general psychiatric journals in 1996. The study revealed that the selected journals differed in their use of sophisticated multivariable methods. In addition, the quality and appropriateness of the statistical analysis were not associated with the number of received citations (Nieminen, Carpenter, Rucker, & Schumacher, 2006). White (1979) and McGuigan (1995) have reviewed the British Journal of Psychiatry (BJP) for statistical errors. The proportion of papers found to have committed statistical errors was 45% in 1977–1978 and 40% in 1993. The reporting of statistical inference in abstracts of visible psychiatric journals has also been analysed (Baethge, Deckert, & Stang, 2018). The authors note that there is a shift to provide precise *p* values without invoking thresholds, such as $p < .05$. In addition, reporting confidence intervals has not replaced *p* values in the abstracts.

Textbooks of medical statistics cover commonly used data analysis methods in medical studies (Bland, 2015; Campbell & Machin, 2003; Indrayan & Malhotra, 2018; Motulsky, 2014). These basic methods include descriptive statistics, inferential methods for comparing groups, methods for repeated measurement, correlation coefficient methods, and multivariable regression methods. However, basic textbooks do not provide a detailed overview of all the statistical techniques readers encounter in psychiatric research reports. Psychiatric studies often include longitudinal data where individuals are repeatedly measured 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 that allow for the use of all available longitudinal data (Diggle, Heagerty, Liang, & Zeger, 2014; Goldstein, 2011; Twisk, 2013). In psychiatry, it is also common in methods to evaluate the agreement among raters or to estimate the reliability of developed questionnaires or test tools (Streiner, Norman, & Cairney, 2015). During recent decades, mathematical statisticians have introduced new computational methods that are compatible with the rapid expansion of computing efficiency (Veierod, Lydersen, & Petter, 2012). Bayesian methods and machine learning are examples of these approaches (Ertel, 2011). However, it is unclear how widely these methods are applied in the psychiatric field (Nieminen, Virtanen, & Vahanikkila, 2017; Vayena, Blasimme, & Cohen, 2018).

Our aim in this study was to evaluate the application and complexity of data analytical methods in high impact psychiatric journals and to assess how authorship, study designs, sample sizes, and data analysis methods have changed in the last 20 years. One object of this study was also to determine whether there were differences between the four journals over the 20-year span in the statistical content of published papers.

2 | MATERIAL AND METHODS

2.1 | Set of articles

For our investigation, we selected four general English-language psychiatric journals: The American Journal of Psychiatry (AJP), the Acta Psychiatrica Scandinavica (APS), the BJP, and JAMA Psychiatry (JAMA PSY; formerly the Archives of General Psychiatry). AJP and JAMA PSY are the two leading medical journals covering psychiatric research and have consistently been among the top journals ranked by Garfield's impact factor, whereas BJP and APS are visible psychiatric journals outside the United States. The four journals had the following impact factors in 2017: JAMA PSY 16.642, AJP 13.396, BJP 5.867, and APS 4.984. We anticipated that 40 articles per journal per year would be sufficient to make comparisons between the journals and publication years. The starting articles were chosen randomly, with the only criteria being that there would 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.

The journals were scrutinized for original research articles published in 1996 and 2018. The year 1996 was chosen because one of the authors (P. N.) has previously applied bibliometric methods to three of the same psychiatric journals published in 1996 and studied the use of statistical methods and quality of reporting in those papers (Nieminen et al., 2006; Nieminen, Rucker, Miettunen, Carpenter, & Schumacher, 2007). The year 2018 was chosen to evaluate the most recent data analysis trends in psychiatric research. The total number of articles reviewed was 320, and each paper underwent careful scrutiny for the use of statistical methods and reporting.

2.2 | Number of authors and international collaboration

The following properties of the informed author(s) of an article were measured: the number of authors and whether there were international collaborations. The collaboration was assessed from the informed author's institutions. If the addresses of the institutions included at least two different countries, then we classified the paper as international.

2.3 | Classification of the study design, sample size, and statistical significance

We classified the articles into the following six groups: cross-sectional surveys, cohort studies, case–control studies, randomized clinical trials (RCTs), meta-analysis, and other designs (including case reports, nonrandomized intervention studies, reliability evaluations, and basic science studies). Cross-sectional studies are observational studies in which the investigator stands apart from the events taking place involving the study subjects. In a cross-sectional survey, each subject is examined on only one occasion, and in longitudinal cohort study, each subject is followed over a period of time. In a case–control study, two populations (typically those with and without a disorder) are sampled. An experimental RCT was defined as a study in which one or more variables were controlled by the investigator to monitor the effect on a process or an outcome, and the study subjects were randomly assigned to two or more treatments controlled by the investigator.

We then further divided the studies into four groups based on the following sample sizes: under 30, 30–99, 100–300, and over 300. Information extracted from the articles also included whether the authors estimated the statistical significance of the primary outcome and whether they reported a statistically significant result (at the $p < .05$ level). 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.

2.4 | Categories of statistical procedures

We manually reviewed all 320 papers for their statistical content. We used the statistical intensity of medical articles (SIMA) tool (Nieminen et al., 2017) 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 68 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. The updated version of the instrument is included in Appendix A. In the instrument, the items are categorized into 16 groups. These subgroups are denoted with capital letters (from A to P) in the evaluation form. Each group includes items measuring the usage of specific statistical analysis methods or reporting styles. We calculated the sums of these subgroup items and a sum of all 68 items for a total score. We designated the total score as the statistical intensity score that ranged from 0 to 76. 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 very high number of tables and figures can be awarded a high SIMA score, but readers with a medical background might find these papers difficult to read. An article with a low statistical intensity value means that it has hardly utilized any statistical methods (e.g., laboratory studies or narrative studies). The reliability of the statistical intensity score has previously been shown to be high, and the interobserver agreement and test–retest reliability as measured by the ICC are >0.80 (Nieminen et al., 2017).

2.5 | Data analysis

IBM SPSS Statistics version 25 was used for the data analysis. Median and quartile values for the number of authors per article are reported. Cross-tabulation was used as a main tool to analyse and present the changes in basic article characteristics and in the use of data analysis methods. The distributions of SIMA scores by publication year and journal were visualized with box plots. In addition, mean and standard deviation values of the SIMA scores were reported to evaluate the differences in the SIMA scores of 1996 and 2018 and to compare the differences of articles published in 2018 between the journal groups. There were no missing data points. If there were problems interpreting a specific paper, its classification was appraised by the authors.

3 | RESULTS

3.1 | Authors

All of the journals demonstrated an increase in the number of authors per article over the years (Table 1). The median number of authors increased from four authors to eight authors from 1996 to 2018 in the total article set. Over 25% of the articles included more than 10 co-authors in 2018 (lower quartile Q_1 = 6 and upper quartile Q_3 = 11). There was a clear trend of increased international collaboration among the co-authors in all evaluated journals. In 1996, lists of authors included authors from at least two different countries as follows: AJP 4 articles (10%), APS 5 (12.5%), BJP 7 (17.5%), and JAMA PSY 10 (25.0%). Two decades later in 2018, the international collaboration among authors was more evident: AJP 20 articles (50%), APS 20 (50.0%), BJP 16 (40.0%), and JAMA PSY 21 (52.5%).

TABLE 1 Median and quartile values of number of authors per article in 1996 (n = 160) and 2018 (n = 160) by the publishing journals

	1996			2018		
	Number of articles	Median	(Q_1, Q_3)	Number of articles	Median	(Q_1, Q_3)
AJP	40	5	(3, 7)	40	9	(6, 11)
APS	40	3	(2, 6)	40	8	(6, 11)
BJP	40	4	(3, 6)	40	8	(5, 10)
JAMA PSY	40	6	(3, 7)	40	8	(6, 11)

Note. Q_1 = lower quartile and Q_3 = upper quartile.

Abbreviations: AJP, American Journal of Psychiatry; APS, Acta Psychiatrica Scandinavica; BJP, British Journal of Psychiatry; JAMA PSY, JAMA Psychiatry.

3.2 | Study design, sample size, and label of main outcome

Table 2 shows the study design of the evaluated articles. Most of the psychiatric articles applied an observational approach both in 1996 and 2018. An experimental design was used in 12.5% of the 1996 articles and in 15.0% of the 2018 articles. Longitudinal cohort studies were more common in 2018 (37.5%) than in 1996 (22.5%) articles. Sample sizes increased in all four journals: sample size exceeded 300 in 63.1% of the articles published in 2018, but only in 30.6% of the articles published in 1996.

The reported "statistical significance" of the primary outcome of articles in the four psychiatric journals is summarized also in Table 2. We found that 127 (79.4%) of psychiatric articles reported a statistical test that had been conducted on the primary outcome in 1996. Of these, 103 (81.1%) articles labelled their findings as "statistically significant." This proportion slightly increased in 2018, with 122/134 (83.1%) articles that were published in the selected journals reporting "statistically significant" results. The proportion of studies that reported "statistically significant or positive results" seems to be higher in the more highly visible psychiatric journals AJP and JAMA PSY than in the other peer-reviewed journals, APS and BJP. In APS, there was a notable change in declaration of "statistical significance"; in 1996, only 60% of the published article reported p values, but in 2018, this proportion had increased to 85% (Table 2).

3.3 | Statistical intensity

The distribution of statistical intensity is shown in Figure 1. The statistical intensity increased from 1996 to 2018 in the evaluated articles. The mean (*SD*) of statistical intensity in all 160 psychiatric articles published in 1996 was 11.8 (4.8) and increased to 18.2 (4.9) in 2018. This trend was clear in all four journals. The SIMA score increased most in APS from 8.2 to 17.9 from 1996 to 2018. In the APS journal, the description of methods and reporting of statistical inference methods and ancillary analyses improved notably. The use of new computational and more complex methods did not increase in any of the evaluated psychiatric journals.

In 2018, the mean (*SD*) of statistical intensity among the evaluated journals varied as follows: AJP 17.1 (4.8), APS 17.9 (5.2), BJP 17.6 (5.4), and JAMA PSY 20.3 (3.4), that is, the statistical intensity was not at the same level among these prominent psychiatric journals. In particular, JAMA PSY published articles with higher statistical intensity than other journals in 2018.

TABLE 2 Distribution of study design, sample size, and main outcome in 1996 (n = 160) and 2018 (n = 160) by the evaluated journals

	AJP		APS		BJP		JAMA PSY		All	
	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)
Study design										
Cross-sectional survey	11 (25.5)	9 (22.50)	19 (47.5)	14 (35.0)	19 (47.5)	7 (17.5)	8 (20.0)	8 (20.0)	57 (35.6)	38 (23.8)
Longitudinal cohort study	8 (20.0)	18 (45.0)	7 (19.5)	13 (32.5)	9 (22.5)	11 (27.5)	12 (30.0)	18 (45.0)	36 (22.5)	60 (37.5)
Case–control	13 (32.5)	1 (2.5)	5 (12.5)	7 (17.5)	4 (10.0)	1 (2.5)	8 (20.0)	1 (2.5)	30 (18.8)	9 (5.6)
Interventional study (clinical trial)	2 (5.0)	8 (20.0)	2 (5.0)	0 (0.0)	5 (12.5)	7 (17.5)	11 (27.5)	9 (22.5)	20 (12.5)	24 (15.0)
Meta-analysis	0 (0.0)	3 (7.5)	0 (0.0)	6 (15.0)	0 (0.0)	9 (22.5)	0 (0.0)	4 (10.0)	0 (0.0)	22 (13.8)
Other	6 (15.0)	2 (5.0)	7 (17.5)	0 (0.0)	3 (7.5)	5 (12.5)	1 (2.5)	0 (0.0)	17 (10.7)	7 (4.4)
Sample size										
<30	8 (20.0)	0	1 (2.5)	1 (2.5)	4 (10.0)	1 (2.5)	5 (12.5)	0	18 (11.3)	2 (1.3)
30–99	15 (37.5)	11 (27.5)	16 (40.0)	10 (25.0)	11 (27.5)	3 (7.5)	13 (32.5)	3 (7.5)	56 (34.4)	27 (16.9)
100–300	6 (15.0)	5 (12.5)	11 (27.5)	4 (10.0)	10 (25.0)	11 (27.5)	9 (22.5)	8 (20.0)	36 (22.5)	28 (17.5)
>300	11 (27.5)	24 (60.0)	10 (25.0)	24 (60.0)	15 (37.5)	24 (60.0)	13 (32.5)	29 (72.5)	49 (30.6)	101 (63.1)
Missing	0 (0.0)	0	2 (5.0)	1 (2.5)	0	1 (2.5)	0	0	2 (1.3)	2 (1.3)
Label of the primary outcome										
Declaration of "statistically not significant"	6 (15.0)	3 (7.5)	7 (17.5)	5 (12.5)	8 (20.0)	2 (5.0)	3 (7.5)	1 (2.5)	24 (15.0)	11 (6.9)
Declaration of "statistically significant"	29 (72.5)	31 (77.5)	17 (42.5)	29 (72.5)	25 (62.5)	27 (67.5)	32 (80.0)	35 (87.5)	103 (64.4)	122 (76.3)
Statistical significance not considered	5 (12.5)	6 (15.0)	16 (40.0)	6 (15.0)	7 (17.5	11 (27.5)	5 (12.5)	4 (10.0)	33 (20.6)	27 (16.9)

Abbreviations: AJP, American Journal of Psychiatry; APS, Acta Psychiatrica Scandinavica; BJP, British Journal of Psychiatry; JAMA PSY, JAMA Psychiatry.

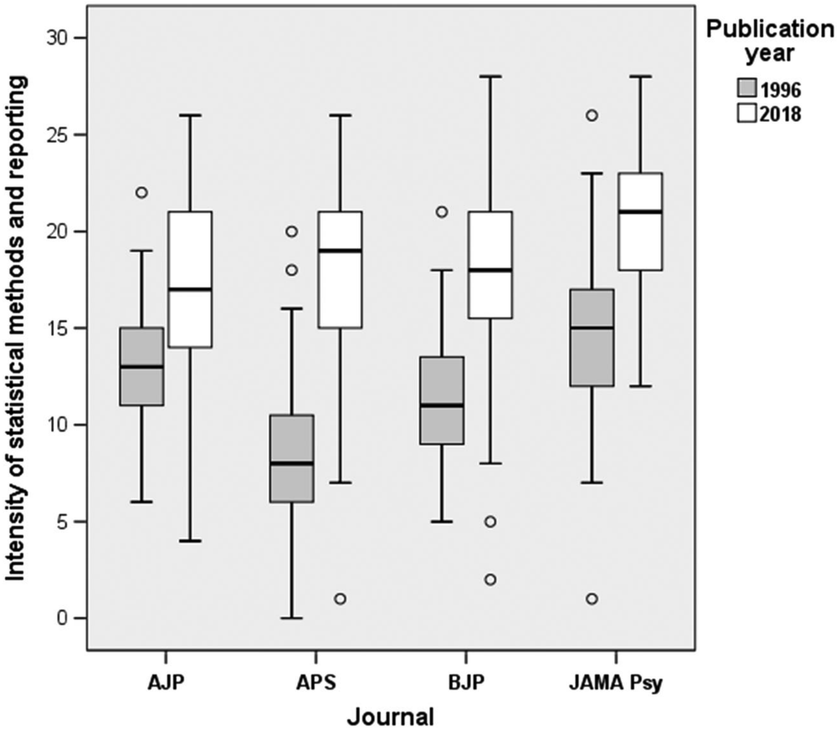

FIGURE 1 Intensity of statistical methods and reporting in psychiatric articles published in 1996 (*n* = 160) and 2018 (*n* = 160) subdivided by journal groups

Colour online, B&W in print

TABLE 3 Percentage of statistical methods and reporting in 1996 (*n* = 160) and 2018 (*n* = 160) in five psychiatric journals

	AJP		APS		BJP		JAMA PSY		ALL	
	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)	1996 *n* (%)	2018 *n* (%)
Statistical methods described	36 (90.0)	29 (72.5)	18 (45.0)	37 (92.5)	25 (62.5)	24 (60.0)	34 (85.0)	31 (77.5)	113 (70.6)	121 (75.6)
Reference to statistical literature	22 (55.0)	17 (42.5)	10 (25.0)	20 (50.0)	10 (25.0)	24 (60.0)	17 (42.5)	24 (60.0)	59 (36.9)	85 (53.1)
Statistical software reported	8 (20.0)	23 (57.5)	9 (22.5)	34 (85.0)	19 (47.5)	33 (82.5)	8 (20.0)	33 (82.5)	44 (27.5)	123 (76.9)
Confidence intervals reported	4 (10.0)	26 (65.0)	5 (12.5)	25 (62.5)	16 (40.0)	30 (75.0)	15 (37.5)	30 (75.0)	40 (25.0)	111 (69.4)
Comparing groups	33 (82.5)	19 (47.5)	24 (60.0)	15 (37.5)	34 (85.0)	20 (50.0)	33 (82.5)	15 (37.5)	124 (77.5)	69 (43.1)
Repeated measurements	22 (55.0)	4 (10.0)	7 (17.5)	0 (0.0)	6 (15.0)	1 (2.5)	13 (32.5)	2 (5.0)	48 (30.0)	7 (4.4)
Correlation coefficient methods	18 (47.5)	9 (22.5)	8 (20.0)	14 (35.0)	17 (42.5)	7 (17.5)	19 (47.5)	8 (20.0)	62 (38.7)	38 (23.7)
Regression models	18 (45.0)	33 (82.5)	11 (27.5)	32 (80.0)	11 (27.5)	26 (65.0)	18 (45.0)	34 (85.0)	58 (36.2)	125 (78.1)
Factor analysis or SEM	6 (15.0)	1 (2.5)	0	2 (5.0)	7 (17.5)	5 (12.5)	1 (2.5)	1 (2.5)	14 (8.7)	9 (5.6)
Other multivariable methods	10 (25.0)	2 (5.0)	1 (2.5)	4 (10.0)	2 (5.0)	2 (5.0)	2 (5.0)	7 (17.5)	16 (10.0)	15 (9.4)
Intracluster correlation methods	1 (2.5)	20 (50.0)	0 (0.0)	8 (20.0)	2 (5.0)	13 (32.5)	3 (7.5)	17 (42.5)	6 (3.7)	58 (36.2)
Measures of agreement	17 (42.5)	3 (7.5)	7 (17.5)	2 (5.0)	9 (22.5)	5 (12.5)	9 (22.5)	6 (15.0)	42 (26.3)	16 (10.0)
Bayesian methods	0	0	0	1 (2.5)	0	0	0	1 (2.5)	0 (0.0)	2 (1.2)
ANN or machine learning	0	0	0	2 (5.0)	0		0	1 (2.5)	0 (0.0)	3 (1.9)

Note. Data are presented as *n* (%).

Abbreviations: AJP, American Journal of Psychiatry; APS, Acta Psychiatrica Scandinavica; BJP, British Journal of Psychiatry; JAMA PSY, JAMA Psychiatry; SEM, structural equation modelling; ANN, artificial neural network.

3.4 | Use of methods

The statistical methods and reporting used in the journals are shown in Table 3. Close to one third of the psychiatric papers provided an incomplete description of their statistical procedures. This serious reporting problem seems to still be prevalent in 2018. Only in APS 2018 and AJP 1996 were the variables and methods used mentioned with enough detail in over 90% of the published articles. In 2018, authors more often gave references to the statistical literature and more frequently detailed, which statistical software had been used. There was also a noteworthy growth in the use of confidence intervals in all evaluated journals. Almost 70% of the evaluated papers reported confidence intervals in 2018 to allow readers to make more informed judgements about the findings.

The percentage frequencies of statistical procedures used in the evaluated article sample of published psychiatric research highlights the broad use of statistical methods (Table 3). However, there has been a shift from applying traditional methods to regression modelling. The procedures most commonly reported in psychiatric papers in 1996 were comparisons of various independent groups. Almost 80% of the articles compared frequencies, mean values, median values, or time-to-event curves between groups. Two decades later, readers of psychiatric articles most often encountered regression models. In addition, various approaches to analyse repeated measurements and hierarchical multilevel data and to estimate intracluster correlation as an extension to regression models have been widely adopted in the analysis of psychiatric data. More than one third of the evaluated papers applied these methods in 2018 (Table 3). Application of techniques such as exploratory and confirmatory factor analysis and structural equation modelling had not increased in 2018, rather they were less commonly reported (8.7% in 1996 and 5.6% in 2018). The trends in the use of statistical techniques were similar in all evaluated psychiatric journals.

In recent decades, mathematical statisticians have introduced newer, more complex methods that can be attributed to the rapid expansion in computing capability including Bayesian methods, artificial neural networks, and machine learning. In practice, applied statisticians and medical researchers apply these tools very rarely; we found only five references to these methods in any of the evaluated respiratory articles (Table 3).

4 | DISCUSSION

We have analysed trends in the statistical intensity of four prominent psychiatric journals. We estimated the frequency with which statistical concepts were mentioned and how data analysis methods were reported in the published paper. The present study shows that articles published in the prominent psychiatric journals in 2018 were applying and reporting statistical methods differently from articles published in 1996 in the same journals. In 2018, readers of psychiatric journals more often encountered regression models with intracorrelation extensions. In addition, the statistical intensity had increased from 1996 to 2018.

To the best of our knowledge, this study represents the most thorough investigation of data analysis trends to date in psychiatry. Previous evidence concerning the extent of methodological shifts mainly came from the limited scope of prominently visible general medical journals (Arnold, Braganza, Salih, & Colditz, 2013; Sato et al., 2017) and showed that the proportion of papers reporting multilevel modelling results, multivariable regression, and time to event methods had significantly increased between 1990 and 2010. Our findings are in line with these earlier studies, but we have additionally widened this observation to psychiatry and applied new tools to estimate the intensity of statistical reporting.

The overall increase in the number of multiple authors in psychiatric journals was similar to the findings in other medical disciplines (Brunson, Wang, & Laubenbacher, 2017; Geminiani, Ercoli, Feng, & Caton, 2014; Hammad, Shaban, & Abu-Zidan, 2012; McDonald, Neff, Rethlefsen, & Kallmes, 2010; Shaban & Aw, 2009). Having multiple authors is a phenomenon that has been observed widely in recent years (Brunson et al., 2017). Reasons for the increase in the number of authors include the increased complexity of psychiatric research that require collaboration between various individuals who bring their own particular expertise to the research projects. Another reason is the increase in the number of multi-institutional projects at both the national and international levels due to improvements in communication and technology. A debatable reason for having multiple authors is the pressure of academia, which has been summarized with the phrase "publish or perish" (Hammad et al., 2012). This has resulted in the inclusion of "guest" authors such as senior individuals and chairpersons who do not fulfil the generally accepted criteria for authorship (Persson & Glänzel, 2014).

The increase in the number of authors is paired with an increase in internationally co-authored papers (National Science Board, 2016). Based on our study results, we can say that co-authorship in psychiatry has also diversified, with the share of publications appearing as a result of international cooperation, cross-country involvement, and domestic interorganizational cooperation. International co-authorship has been found to be more useful for papers from small countries. In large countries, such as the United States, the effect of international papers on impact ranking is much weaker (Persson, 2010).

We observed that traditional observational study designs (cross-sectional surveys and case–control studies) occurred with less frequency in 2018 compared with 1996, with a corresponding increase in the proportion of longitudinal cohort studies and meta-analyses. Our finding is in line with previous findings from highly visible general medical journals (Arnold et al., 2013; Yi et al., 2015). This suggests that study designs with stronger evidence are required in and applied to psychiatric papers published in 2018.

Sample size is an important consideration for research. The number of studies that used large sample sizes increased in all four journals. In 2018, over 60% of the evaluated articles had data that included at least 300 study subjects. The increased frequency of cohort studies with very large sample sizes is a likely explanation for this development. The larger the sample size, the greater the precision and thus the more power for a given study to detect the effect of a given size.

The psychiatric literature shows a strong tendency to accentuate significance testing, and specifically "statistically significant" outcomes. Over the time frame of this study, 1996 to 2018, we noted that there has been no change in the use of significance testing. Psychiatric researchers still likely perform several tests to label their findings as "statistically significant," and this phenomenon can lead to multiple testing problems and possible publication bias. Small p values have a hypnotic effect on editors and readers alike. Altman (2000) stated that in all medical fields, "p-values and p < 0.05 rules most." Studies with "p < .05" are more likely to be reported and published. Our results reinforce increasing concern for the misuse of significance testing in interpreting medical data (Mlinarić, Horvat, & Šupak Smolčić, 2017; Wasserstein, Schirm, & Lazar, 2019). However, it should be noted that there was a substantial increase in the reporting of confidence interval in all four evaluated journals.

There were changes between the articles published in 1996 or 2018 regarding the reporting and use of data analysis methods. Our results show that reporting the statistical software thoroughly is almost a standard in 2018 in the selected visible psychiatric journals. Reporting the software used in multivariable analyses is vital because different software utilize different calculation algorithms (Okunade, Chang, & Evans, 1993). Reporting the software also helps readers evaluate and understand the details specific to certain statistical software. Regarding trends of statistical methods, the proportion of papers that applied traditional statistical methods to compare independent groups or to evaluate repeated measurements decreased from 1996 to 2018. Multivariable regression modelling has replaced the univariable comparison methods in observational studies. However, in RCTs, the evaluation of the primary outcome between treatment groups should be based on the basic inferential methods for comparing between groups. In addition, intracluster correlation methods (hierarchical linear modelling, regression models with generalized estimation equation, or random effect extensions) almost completely overtook the outdated analysis of variance for repeated measurements.

Factor analysis and structural equation modelling are commonly used in psychology and health care sciences. However, our study shows that these methods are not widely used in psychiatry. It can be argued that researchers who publish in the visible psychiatric journals seem closely follow the data analysis methods and reporting practices common in the visible medical journals (Arnold et al., 2013; Strasak, Zaman, Marinell, Pfeiffer, & Ulmer, 2007).

Our findings indicate that the last 20 years has seen a drastic increase in the use of regression methods beyond the basic methods in psychiatric research. The availability of statistical software packages has greatly facilitated extensive data analysis, increasing their frequency of use and complexity. Altman and Goodman (1994) suggested that the following methods are likely to be seen more often in the coming years: (a) bootstrap methods, (b) Bayesian methods, (c) generalized additive methods, (d) classification and regression trees, (e) general estimation equations, (f) models for hierarchical data, and (g) neural networks. Although these methods are now sometimes used in psychiatric research, our findings show that only models for hierarchical data are widely used in 2018. The broader introduction and acceptance of a new analysis method (as useful as the method might be) into psychiatric publications seems to require the method being incorporated into the standard statistical packages generally used by researchers. In addition, if readers do not understand the mathematics or reporting style, or if 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.

The increased statistical intensity in psychiatric papers requires strong data analysis understanding to interpret the applications and the results. The phenomenon that statistical intensity has increased indicates that journals are more strict regarding the accuracy and type of data analysis that are reported in articles (Arnold et al., 2013; Hellems, Gurka, & Hayden, 2007; Welch & Gabbe, 2002; Yi et al., 2015) Authors, reviewers, and journal editors have a greater responsibility for ensuring that statistical procedures are used appropriately, as it may be increasingly unrealistic to expect readers to fully understand the statistical analyses used in journal articles.

5 | CONCLUSIONS

This study provides new information on the status of statistical methods applied for analysing data in journals devoted to psychiatric research. Our findings suggest that readers of psychiatric journals need to be more familiar with data-analysis methods including multivariable regression models than they were 20 years ago. We encourage all educators in psychiatric medicine to review their programmes to ensure that an appropriate level of data analysis understanding is provided to their students, clinicians, and colleagues. We also encourage all readers, authors, and reviewers of psychiatric journals who wish to be more effective consumers of the psychiatric literature to review their own statistical skills and to present their results in a manner similar to that advocated and presented in the prominent psychiatric journals.

DECLARATION OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

ORCID

Pentti Nieminen https://orcid.org/0000-0003-0591-586X

REFERENCES

Altman, D. G. (2000). Statistics in medical journals: Some recent trends. *Statistics in Medicine*, *19*(23), 3275–3289. https://doi.org/10.1002/1097-0258(20001215)19:23<3275::AID-SIM626>3.0.CO;2-M

Altman, D. G., & Goodman, S. N. (1994). Transfer of technology from statistical journals to the biomedical literature: Past trends and future predictions. *JAMA*, *272*(2), 129–132. https://doi.org/10.1001/jama.1994.03520020055015

Arnold, L. D., Braganza, M., Salih, R., & Colditz, G. A. (2013). Statistical trends in the journal of the American Medical Association and implications for training across the continuum of medical education. *PLoS ONE*, *8*(10), e77301. https://doi.org/10.1371/journal.pone.0077301

Baethge, C., Deckert, M., & Stang, A. (2018). Tracing scientific reasoning in psychiatry: Reporting of statistical inference in abstracts of top journals 1975–2015. *International Journal of Methods in Psychiatric Research*, *27*(3), e1735. https://doi.org/10.1002/mpr.1735

Bland, M. (2015). *An introduction to medical statistics*. UK: Oxford University Press.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Brunson, J. C., Wang, X., & Laubenbacher, R. C. (2017). Effects of research complexity and competition on the incidence and growth of coauthorship in biomedicine. *PLoS ONE*, *12*(3), e0173444. https://doi.org/10.1371/journal.pone.0173444

Campbell, M. J., & Machin, D. (2003). *Medical statistics. A commonsense approach* (3. ed., reprint. ed.). Chichester: Wiley.

Diggle, P., Heagerty, P., Liang, K., & Zeger, S. (2014). *Analysis of longitudinal data* (2nd ed.). Oxford: OUP Oxford.

Ertel, W. (2011). *Introduction to artificial intelligence*. London: Springer. https://doi.org/10.1007/978-0-85729-299-5

Geminiani, A., Ercoli, C., Feng, C., & Caton, J. G. (2014). Bibliometrics study on authorship trends in periodontal literature from 1995 to 2010. *Journal of Periodontology*, *85*(5), e143. https://doi.org/10.1902/jop.2013.130354

Goldstein, H. (2011). *Multilevel statistical models* (4. ed. ed.). Chichester: Wiley.

Hammad, F. T., Shaban, S., & Abu-Zidan, F. (2012). Multiple authorship and article type in journals of urology across the Atlantic: Trends over the past six decades. *Medical Principles and Practice*, *21*(5), 435–441. https://doi.org/10.1159/000339884

Hellems, M. A., Gurka, M. J., & Hayden, G. F. (2007). Statistical literacy for readers of pediatrics: A moving target. *Pediatrics*, *119*(6), 1083. https://doi.org/10.1542/peds.2006-2330–1088.

Hokanson, J. A., Bryant, S. G., Gardner, J. R., Luttman, D. J., Guernsey, B. G., & Bienkowski, A. C. (1986). Spectrum and frequency of use of statistical techniques in psychiatric journals. *American Journal of Psychiatry*, *143*(9), 1118–1125. https://doi.org/10.1176/ajp.143.9.1118

Horton, N. J., & Switzer, S. S. (2005). Statistical methods in the journal. *The New England Journal of Medicine*, *353*(18), 1977–1979. https://doi.org/10.1056/NEJM200511033531823

Indrayan, A., & Malhotra, R. K. (2018). *Medical biostatistics* (Fourth Edition ed.). Milton: CRC Press.

Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*, *84*(3), 575–603. https://doi.org/10.1007/s11192-010-0202-z

McDonald, R. J., Neff, K. L., Rethlefsen, M. L., & Kallmes, D. F. (2010). Effects of author contribution disclosures and numeric limitations on authorship trends. *Mayo Clinic Proceedings*, *85*(10), 920–927. https://doi.org/10.4065/mcp.2010.0291

McGuigan, S. M. (1995). The use of statistics in the British Journal of Psychiatry. *The British Journal of Psychiatry*, *167*(5), 683–688. https://doi.org/10.1192/bjp.167.5.683

Miettunen, J., Nieminen, P., & Isohanni, M. (2002). Statistical methodology in general psychiatric journals. *Nordic Journal of Psychiatry*, *56*(3), 223–228. https://doi.org/10.1080/080394802317607219

Mlinarić, A., Horvat, M., & Šupak Smolčić, V. (2017). Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia Medica*, *27*(3), 030201. https://doi.org/10.11613/BM.2017.030201

Motulsky, H. (2014). *Intuitive biostatistics* (3. Ed. ed.). New York: Oxford University Press.

National Science Board (2016). *Science & engineering indicators 2016*. Washington, DC: National Science Foundation.

Nieminen, P. (1995). Statistical content of published therapeutic community research. *Therapeutic Communities*, *16*(4), 239–251.

Nieminen, P., Carpenter, J., Rucker, G., & Schumacher, M. (2006). The relationship between quality of research and citation frequency. *BMC Medical Research Methodology*, *6*, 42. https://doi.org/10.1186/1471-2288-6-42

Nieminen, P., Rucker, G., Miettunen, J., Carpenter, J., & Schumacher, M. (2007). Statistically significant papers in psychiatry were cited more often than others. *Journal of Clinical Epidemiology*, *60*(9), 939–946. https://doi.org/10.1016/j.jclinepi.2006.11.014

Nieminen, P., Virtanen, J. I., & Vahanikkila, H. (2017). An instrument to assess the statistical intensity of medical research papers. *PLoS ONE*, *12*(10), e0186882. https://doi.org/10.1371/journal.pone.0186882

Okunade, A. A., Chang, C. F., & Evans, R. D. (1993). Comparative analysis of regression output summary statistics in common statistical packages. *American Statistician*, *47*(4), 298–303.

Persson, O. (2010). Are highly cited papers more international? *Scientometrics*, *83*(2), 397–401. https://doi.org/10.1007/s11192-009-0007-0

Persson, O., & Glänzel, W. (2014). Discouraging honorific authorship. *Scientometrics*, *98*(2), 1417–1419. https://doi.org/10.1007/s11192-013-1042-4

Petrovecki, M. (2009). The role of statistical reviewer in biomedical scientific journal. *Biochemia Medica*, *19*(3), 223–230. https://doi.org/10.11613/BM.2009.020

Pincus, H. A., Henderson, B., Blackwood, D., & Dial, T. (1993). Trends in research in two general psychiatric journals in 1969-1990: Research on research. *American Journal of Psychiatry*, *150*(1), 135–142. https://doi.org/10.1176/ajp.150.1.135

Sato, Y., Gosho, M., Nagashima, K., Takahashi, S., Ware, J. H., & Laird, N. M. (2017). Statistical methods in the journal—An update. *The New England Journal of Medicine*, *376*(11), 1086–1087. https://doi.org/10.1056/NEJMc1616211

Shaban, S., & Aw, T. (2009). Trend towards multiple authorship in occupational medicine journals. *Journal of Occupational Medicine and Toxicology (London, England)*, *4*(1), 3. https://doi.org/10.1186/1745-6673-4-3

Strasak, A. M., Zaman, Q., Marinell, G., Pfeiffer, K. P., & Ulmer, H. (2007). The use of statistics in medical research: A comparison of “the new England journal of medicine” and “nature medicine”. *The American Statistician*, *61*(1), 47–55. https://doi.org/10.1198/000313007X170242

Straus, S. E., Glaziou, P., Richardson, W. S., & Haynes, R. B. (2018). *Evidence-based medicine: How to practice and teach EBM*. Edinburgh: Elsevier.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5. ed. ed.). Oxford: Oxford University Press.

Twisk, J. W. R. (2013). *Applied longitudinal data analysis for epidemiology* (2. ed. ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139342834

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, *15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Veierod, M. B., Lydersen, S., & Petter, L. (2012). *Medical statistics in clinical and epidemiological research*. Oslo: Gyldendal.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05". *American Statistician*, *73*(S1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Welch, G. E., & Gabbe, S. G. (2002). Statistics usage in the American Journal of Obstetrics and Gynecology: Has anything changed? *American Journal of Obstetrics and Gynecology*, *186*(3), 584–586. https://doi.org/10.1067/mob.2002.122144

White, S. J. (1979). Statistical errors in papers in the British Journal of Psychiatry. *The British Journal of Psychiatry*, *135*(4), 336–342. https://doi.org/10.1192/bjp.135.4.336

Yi, D., Ma, D., Li, G., Zhou, L., Xiao, Q., Zhang, Y., … Wu, Y. (2015). Statistical use in clinical studies: Is there evidence of a methodological shift? *PLoS ONE*, *10*(10), e0140159. https://doi.org/10.1371/journal.pone.0140159

How to cite this article: Nieminen P, Kaur J. Reporting of data analysis methods in psychiatric journals: Trends from 1996 to 2018. *Int J Methods Psychiatr Res*. 2019;28:e1784. https://doi.org/10.1002/mpr.1784

APPENDIX A

STATISTICAL ANALYSES AND METHODS EVALUATION FORM

Article id number: ________________________

Reviewer: ________________________ Evaluated: _________________

Journal name: ____________________________

Volume and pages: ___________________ Year: 2018 ______________

DOI: ____

Authors

Name of the first author: __________________________________

Design, software, and primary outcome

Study design:
1. Cross-sectional survey
2. Longitudinal or cohort study
3. Case–control
4. Randomized clinical trial
5. Reliability/diagnostic study
6. Laboratory work
7. Meta-analysis
8. Case study
9. Intervention study (not RCT)
10. Other

Sample size: __________________

First software: ________________ Second software: ________________

Statistical significance of the primary outcome:

0 = *Not significant*

1 = *Significant*

2 = *Not evaluated*

	No	Yes
A. Description of methods		
Item A.1 Statistical methods described with enough detail	0	1
Item A.2 Extended description of some specific procedures	0	1
Item A.3 Software reported	0	1
Item A.4 Reference to statistical literature	0	1

(Continued)

		No	Yes		
B. Ancillary analyses and computations					
Item B.1 Power analysis and sample calculations		0	1		
Item B.2 Methodological considerations		0	1		
Item B.3 Variable transforms, recodes, or constructs		0	1		
Item B.4 Sensitivity or influence		0	1		
Item B.5 Stratification or subgroup analyses		0	1		
Item B.6 Addressing subject attrition or exclusion		0	1		
C. Reporting of statistical inference					
Item C.1 *P* values		0	1		
Item C.2 Confidence intervals		0	1		
Item C.3 *P* values or confidence intervals in abstract		0	1		
Item C.4 Actual *p* values reported		0	1		
Item C.5 Adjustment of *p* values (multiple comparisons)		0	1		
D. Number of *p* values and CIs in tables and figures		0	1–9	10–29	≥30
Item D.1 Number of *p* values	[N1=]	0	1	2	3
Item D.2 Number of CIs	[N2=]	0	1	2	3
E. Number of statistical tables and figure		0	1–2	3–4	≥5
Item E.1 Number of tables	[N3=]	0	1	2	3
Item E.2 Number of figures	[N4=]	0	1	2	3
F. Basic inferential methods for comparing groups					
Item F.1 Comparison of means		0	1		
Item F.2 Methods for proportions and cross-tabulations		0	1		
Item F.3 Nonparametric methods		0	1		
G. Basic methods for repeated measurements					
Item G.1 Comparison of means		0	1		
Item G.2 Methods for proportions and cross-tabulations		0	1		
Item G.3 Nonparametric methods		0	1		
H. Correlation coefficient methods					
Item H.1 Pearson's correlation coefficient or related		0	1		
Item H.2 Rank correlation coefficients		0	1		
I. Other basic methods					
Item I.1 Kaplan–Meir curves		0	1		
Item I.2 Epidemiological measures of disease frequency		0	1		
Item I.3 Evaluation of normality		0	1		
Item I.4 Comparison of variances		0	1		
Item I.5 Missing data addressed		0	1		
J. Regression methods					
Item J.1 Linear regression		0	1		
Item J.2 Logistic regression		0	1		
Item J.3 Cox regression		0	1		
Item J.4 Poisson or negative binomial regression		0	1		
Item J.5 Other regression model		0	1		
K. Multivariable model building					
Item K.1 Assessing assumptions		0	1		
Item K.2 Methods for stepwise variable selection		0	1		

(Continued)

	No	Yes
Item K.3 Covariate adjustments	0	1
Item K.4 Goodness of fit statistics and model validation	0	1
Item K.5 Methods for analysing interaction	0	1
Item K.6 Influence analysis and other diagnostic statistics	0	1
L. Intracluster correlation methods		
Item L.1 GEE	0	1
Item L.2 Random effects models or mixed models	0	1
Item L.3 Complex surveys, robust estimates, or other	0	1
M. Measures of agreement and diagnostic tests		
Item M.1 Assessing agreement with continuous outcomes	0	1
Item M.2 Agreement for categorical classifications	0	1
Item M.3 Methods for diagnostic tests	0	1
N. Meta-analysis		
Item N.1 Overall CI or *p* value	0	1
Item N.2 Test of homogeneity or meta-regression	0	1
Item N.3 Other meta-analysis procedure	0	1
O. Other multivariable methods		
Item O.1 Factor analysis	0	1
Item O.2 Principal component analysis	0	1
Item O.3 SEM	0	1
Item O.4 Cluster analysis	0	1
Item O.5 Other multivariable method not classified above	0	1
P. Other techniques		
Item P.1 Methods for analysing time series	0	1
Item P.2 Bayesian methods	0	1
Item P.3 Curve fitting, spline functions, or GAM	0	1
Item P.4 Statistical genetics and microarray data	0	1
Item P.5 Artificial neural networks or machine learning	0	1
Item P.6 Simulations	0	1
Item P.7 Bootstrap or jack-knife estimates	0	1
Item P.8 Interim analyses	0	1
Item P.9 Cost-effectiveness analyses	0	1
Item P.10 Propensity score estimation	0	1
Item P.11 Other technique not classified in earlier items	0	1