

The Current Use of Statistics in Clinical and Health Psychology Under Review

El Uso de la Estadística en Psicología Clínica y de la Salud a Revisión

Albert Sesé and Alfonso Palmer
Balearic Islands University, Spain

Abstract. The use of statistics in any scientific discipline may be considered a key element in the assessment of the degree of maturity of a field and show the generation of non-speculative knowledge. The aim of this study is to carry out a bibliometric analysis of the use of statistical methods in Clinical and Health Psychology. In order to achieve this aim, a group of 8 journals with an ISI impact index located in quartile 1 or quartile 2 were chosen, and 623 articles published in 2010 were reviewed. The main results show a ranking with the most used techniques and their distribution in each of the journals. This article presents a panoramic view of the degree of use of statistical methodology and its level of diversity and complexity. Finally, a suggestion of the application of statistical models that are currently not present, but which may be very useful for research in Clinical and Health Psychology, is made. This information is most relevant for improving the quality of current research and education of new researchers.

Keywords: statistical methods, clinical and health psychology, bibliometrics.

Resumen. El uso de la estadística en cualquier disciplina científica puede ser considerado como un elemento clave en la evaluación del grado de madurez de un campo y demuestra la generación de conocimiento no especulativo. El objetivo de este estudio es llevar a cabo un análisis bibliométrico del uso de los métodos estadísticos en Psicología Clínica y de la Salud. Para la consecución de este objetivo se escogió un grupo de 8 revistas con índice de impacto ISI, situadas en cuartil 1 o cuartil 2, y fueron revisados 623 artículos publicados durante al año 2010. Los principales resultados muestran un ranking con las técnicas más utilizadas y su distribución en cada una de las revistas. Este artículo presenta una visión panorámica del grado de utilización de la metodología estadística y su nivel de diversidad y complejidad. Finalmente, se sugiere la aplicación de modelos estadísticos que actualmente no tienen presencia, pero que pueden ser muy útiles para la investigación en Psicología Clínica y de la Salud. Esta información es muy relevante para la mejora de la calidad de la investigación actual y del entrenamiento de nuevos investigadores.

Palabras clave: métodos estadísticos, Psicología Clínica y de la Salud, Bibliometría.

Introduction

Psychology, as a behaviour science, bases the generation of knowledge on the use of the scientific method, whose fundamental pillars are observation

and experimentation. All sciences demand results and are aimed at seeking empirical evidence which is favourable toward the hypotheses formulated, in such a way that they ensure predictable results. Thus, psychological research seeks to obtain empirical evidence which will allow the hypotheses derived from the different theories postulated to be contrasted. In order to achieve this aim, some good research

Correspondence on this article should be sent to the first author to the following e-mail: albert.sese@uib.es

designs and appropriate statistical methods must be established. In this respect, it is important to point out that greater statistical complexity does not necessarily have to lead to scientific progress, because for this to happen it is necessary to have correct methodological designs and appropriate and plausible hypothetic models, as necessary, yet not sufficient, conditions (Palmer, Sesé and Montaña, 2005).

Despite this objection, the use of statistical techniques in general empirical research - and in psychology in particular - may be considered an indicator of the degree of scientific progress achieved. In this sense, scientific progress may be more fruitful in as much as the use of statistics may help discover complex relationships between the variables under study. These complex relationships have a greater likelihood of being discovered through the application of advanced statistical models of a multivariate nature. Loftus (1996) explicitly states that Psychology may become a major science if the type of statistical data analysis applied in research or professional practice is improved: and he establishes an important critique of the generation of psychological knowledge with respect to the difficulty involved on many occasions of being able to replicate the results obtained. One possible cause of this lack of consistency lies in a poor choice of the potentially usable statistical tools and of their inappropriate use. Loftus (1996) literally states that, "Sometimes I feel that what we do in research in Psychology is like trying to build a violin with a stone mallet and a chainsaw. The tools we apply to the task are no the appropriate ones and, as a result, we end up building a large quantity of bad quality violins".

Together with Loftus there have been a large number of authors who have tried to establish guidelines and practical advice on the appropriate application of statistical methodology in psychological research, focusing for instance on the concept of statistical significance, effect size, power, confidence intervals, or on the appropriate use and interpretation of specific statistical techniques (Abelson, 1995, 1997; Chow, 1996; Cohen, 1988, Cowles, 1989; Cumming and Finch, 2001; Everitt, 2000; Fritz, 1996; Harlow, Mulaik and Steiger, 1997; Kelley, 2007; Kirk, 1996; Robinson and Wainer, 2001; Rosenthal and Rubin, 1994; Schmidt, 1996;

Smithson, 2003; Snyder and Lawson, 1993; Wainer, 1999, Wainer and Robinson, 2003; Wilkinson, 1999).

Despite the existence of these reference works, some authors such as von Eye and Schuster (2000) have analyzed the development of statistical methodology in psychological research at the beginning of the third millennium and continue to recognize the existence of quite a few obstacles, above all due to concept comprehension, which may have been exacerbated by the easy access to a wide range of computer software for statistical analysis.

Taking all these considerations into account, one of the fundamental factors in order to establish the degree of quality of current research in Psychology consists of determining what statistical methods are being used in order to assess the validity of the main working hypotheses, within the framework of the theoretical models that are being postulated. Following these assumptions and focusing the analysis on the use of statistical methods - as possible signs of progress and development in research - the aim of this study is to assess the degree with which these techniques are being used nowadays, not in Psychology in general, but in the area of Clinical and Health Psychology.

In order to achieve this aim, a sample of 8 relevant journals in this field that fulfill the quality criteria established by ISI Thomson and are included in the lists of the Journal Citation Report of 2009 were considered. The study reviews all 623 articles published in the 8 journals throughout 2010.

The study traces the statistical techniques used in each of the articles published, according to a general taxonomy of tests drawn up by the authors. It is fundamental to point out that a journal must not be inferred to be either better or more important than another just because it has a greater or lower incidence of use of statistical methods, as there may be differences between each of them in terms of scope. From this perspective, a detailed study of the research efforts carried out in this specific, eminently relevant field of Psychology - through statistical use - guarantees an adequate understanding of the research practices and techniques used, and can give us a non-speculative idea of the good and bad aspects of the scientific quality of this field. For the

purposes of intervention and improvement, the results can suggest the implementation of new methods or the enhancement of already known ones for active researchers, and a modification of the learning syllabus for those who are in education.

Statistical methods review

Study methods

For the population of potentially selectable journals, the 93 journals included in 2009 in the category of “Psychology, Clinical” in the Journal Citation Reports were considered. It was decided to reduce the population of journals to those which, due to their impact index, occupied the top positions, specifically those that occupied the first and second quartile. In this way, the total number of journals that made up the reference population was 46. A random selection was made of 8 journals, with different periodicities and issues published per year. To obtain a reading of the most recent research, articles published by the journals in 2010 were considered. Table 1 shows the eight journals selected, in order of impact index in 2009, and shows the quartile occupied, and the amount of issues and articles published per journal in 2010.

To count the different statistical methods used by the 623 articles reviewed, a system of categories was constructed which, with no pretension of com-

prehensiveness, aimed to cover most of the statistical models available to researchers in behavioural and health sciences. Table 3 in the “Results” section shows the system of categories for the statistical techniques employed.

In general, it is much more productive to present the information in the form of a category table with the least possible groupings, both for reasons of simplification, and also to make it easier for readers to create other groupings that are more along the lines of their personal interests. The study also includes the categorization of the use of techniques based on the type of research design applied, as well as the prevalence of use of a set of basic statistical parameters such as: provision of effect size, confidence intervals, power, assessment of statistical assumptions, and, where appropriate, solutions in case of non-compliance.

Results

In order to conduct the study of the use of statistical techniques in the journals analyzed, a typology made up of 46 techniques or groups of techniques was carried out. From the analysis of the 623 articles reviewed, a total frequency of use of these techniques was found to be 1549. Table 2 provides the average number of statistical techniques used in each journal analyzed (the name of each journal appears as an acronym).

Table 1. Bibliometric parameters of this study

Journals	2009 ISI Impact Factor	Quartile	Issues	Articles
Journal of Behavioural Medicine (JBM)	3.084	Q1	6	48
Behaviour, Research and Therapy (BRT)	2.995	Q1	12	160
Depression and Anxiety (DA)	2.926	Q1	12	116
Behavior Therapy (BT)	2.845	Q1	4	48
Journal of Anxiety Disorders (JAD)	2.682	Q1	8	130
International Journal of Clinical and Health Psychology (IJCHP)	1.792	Q2	3	30
British Journal of Clinical Psychology (BJCP)	1.753	Q2	4	37
British Journal of Health Psychology (BJHP)	1.485	Q2	4	54

Table 2. Average number of statistical techniques used per article according to journal

Journal	BT	BJCP	DA	JAD	BRT	BJHP	IJCHP	JBM
Number of techniques	106	72	241	386	392	115	66	171
Number of articles	48	37	114	132	160	54	30	48
Mean Average	2.21	1.95	2.11	2.92	2.45	2.13	2.2	3.56

The largest average number of techniques used corresponds to the journal JBM (3.56) which, on the other hand, is the one with the highest impact index in the set of journals (3.084). To analyze whether there is a sort of pattern between the use of techniques and the impact index achieved, a non-parametric relationship between the average number of techniques used and the value of the impact factor was estimated. The value obtained was 0.619, but it is worth remembering that, as it was based on only 8 observations, it is not significant ($P = 0.102$).

As far as the authorship of the studies is concerned, 58.1% of articles move between 2 and 4 authors which - to our understanding - defines the optimal group for teamwork. Only 3.85% of articles are signed by one author - which can be considered a positive piece of information as it is not advisable to work independently - and the remaining 38% of articles are signed by 5 or more authors. 3.2% of articles are signed by 10 or more authors, reaching the extreme figure in one study signed by 25 authors.

Regarding the number of signatories per journal, DA has the greatest variability in number of authors, as it contains 50% of the articles signed by only one author and, at the same time, it also contains the two articles with the greatest number of signatories (23 and 25 authors). The journal BT is the one with the minimum range as it moves between 2 and 8, followed by BJCP with a range between 1 and 8, BJHP which moves between 1 and 9 and JCHP whose range is from 1 to 10. With respect to the nationality of the first signatory of the article, 33 different countries were counted, although the most productive through the 8 journals are, in this order, the United States (45.58%), United Kingdom (15.41%), Canada (6.42%), Australia (6.1%), Holland (5.94%), Spain (4.5%) and Germany (3.85%).

Concerning the type of article published, of the 623 articles reviewed, 47 were found to be theoretical articles (7.54% of the total), 32 of which (68.1%) are in the journal DA. Other journals that include theoretical articles are BJCP, JAD, BRT and IJCHP with 3 articles each, BJHP with 2 and JBM with 1. 16 articles of a qualitative nature were published, which represents 2.6% of the total, with the journal BJHP, with 9 articles (56.25%) as the journal where most were published.

As far as meta-analysis studies are concerned, 9 articles were published in 2010, among the journals reviewed, except for IJCHP which did not publish any. 20 articles of an instrumental nature were published, with the journals BT and JCHP, with 7 articles each, as the ones that accumulate 70% of this type of article. The journal BJCP with 5 articles and BJHP with 1 article, complete the list.

In order to analyze the incidence of use of the different statistical methods, a frequency table for techniques and journals was devised, showing the distribution of the 1549 statistical techniques used through the 623 articles published in the 8 journals considered (Table 3). In agreement with the frequency count, the technique that appears in the first place is Correlation (207; 13.36%), followed by Between-Subjects T-Test (161; 10.39%), Chi-Square (153; 9.88%), Reliability Analysis-ROC (121; 7.81%), Between-Subjects One-Way Anova (108; 6.97%), and Linear Regression Models (83; 5.36%). The techniques that follow in the list individually reach a percentage use less than 5%. These 6 top techniques with the greatest frequency of use make up 53.78% of the total statistical techniques used.

In order to obtain a more succinct, comprehensive view of the results, below we present some groupings according to families of techniques depending on their affinity or task. Thereby, it can be easier for the reader to obtain a more comprehensive view of the use of different techniques. Thus, for instance, Between-Subjects T-Test (161), Chi-Square (153) and Correlation (207) can be considered three basic techniques that are generally used to assess the degree of prior homogeneity between the different groups or sub-samples used in studies, rather than statistical procedures to contrast fundamental hypotheses. This grouping, with a frequency of 521, makes up 33.63% of the total techniques used.

Another relevant grouping is composed of all the types of Anova that appear in the table (Between-Subjects, Within-Subjects and Mixed), which obtain a joint frequency of 485, and which make up 31.31% of the total techniques used. It is worth pointing out, in the field of Designs, that only 1 article uses a block design, whereas 3 articles use a random design and 3 a mixed design. A third grouping is made up of regression models (linear, hierarchical

Table 3. Count of use of statistical methods according to journal

Statistical methods	BT	BJCP	DA	JAD	BRT	BJHP	IJCHP	JBM	Total
1. Correlation	15	13	25	53	41	22	11	27	207
2. Between-subjects T-Test	11	10	37	42	31	6	9	15	161
3. Chi-Square	6	1	46	48	26	8	4	14	153
4. Reliability Analysis – ROC	1	4	16	51	4	9	4	32	121
5. Anova (A between subjects)	5	5	17	28	36	4	3	10	108
6. Linear Regression Models	6	6	14	23	13	11	3	7	83
7. Hierarchical Regression	1	6	4	19	22	10	3	6	71
8. Logistic Regression	2	1	20	16	13	4	3	7	66
9. Ancova	2	2	8	14	21	3	2	3	55
10. Non-Parametric Analysis	4	2	13	12	13	1	2	7	54
11. Anova (Ax B Mixed)	6	2	1	6	39	0	0	0	54
12. Factor analysis and PCA	6	3	3	17	5	4	9	5	52
13. Structural Equation Modeling	4	0	4	10	10	6	1	5	40
14. Manova	5	8	3	6	7	2	1	4	36
15. Anova (A within subjects)	6	0	4	9	7	2	3	4	35
16. Mediation Analysis	2	0	0	4	12	5	1	6	30
17. Within-subjects T-Test	0	0	3	4	12	3	1	2	25
18. Multilevel Analysis	5	2	1	4	7	0	1	2	22
19. Anova (Ax BxC Mixed)	0	0	0	2	18	2	0	0	22
20. Anova (Ax B between subjects)	1	2	0	2	10	0	1	2	18
21. Linear Mixed Models	1	0	6	0	4	0	0	3	14
22. Qualitative methods	0	4	0	1	0	6	2	0	13
23. Resampling (Jackknife, etc.)	3	0	1	2	3	1	0	2	12
24. Anova (Ax B within subjects)	5	0	2	1	3	1	0	0	12
25. Mancova	2	0	1	1	4	2	1	1	12
26. Anova (Ax BxC between subjects)	0	0	0	3	4	0	0	3	10
27. Survival Analysis	0	0	5	0	3	1	0	0	9
28. Generalized Estimating Equations	1	0	3	0	4	0	0	1	9
29. Cluster Analysis	0	0	0	2	3	1	0	1	7
30. Poisson Models (NB, ZI)	0	0	1	0	2	0	0	1	4
31. Robust techniques	0	0	0	2	2	0	0	0	4
32. Taxometric procedures	3	0	0	1	0	0	0	0	4
33. Discriminant Analysis	0	0	0	2	1	0	0	0	3
34. Anova (Ax BxC within subjects)	0	0	1	0	2	0	0	0	3
35. Anova Type II Random effects	0	0	0	1	2	0	0	0	3
36. Anova Type III Mixed effects	0	0	0	0	2	0	0	1	3
37. Growth Curve Analysis	0	1	0	0	1	1	0	0	3
38. Time Series Analysis	1	0	1	0	0	0	0	0	2
39. Other	0	0	0	0	2	0	0	0	2
40. Logit Models	0	0	1	0	0	0	0	0	1
41. Probit and Tobit Models	1	0	0	0	0	0	0	0	1
42. Log-Linear Models	0	0	0	0	1	0	0	0	1
43. Bloqueo (DBA, DMAG)	0	0	0	0	1	0	0	0	1
44. Factor Mixture Models	1	0	0	0	0	0	0	0	1
45. Chaos Theory	0	0	0	0	0	0	1	0	1
46. Ordinal Regression	0	0	0	0	1	0	0	0	1

1549

and logistic), which obtain a joint frequency of 220, and 14.20% of the total. It is worth noting that other regression models are used very little, as only 4 articles use the Poisson Regression or the Ordinal Regression, with one article.

Other groupings, more minority ones, are the ones composed of Psychometric Analysis and ROC Analysis, which with a frequency of 121 make up

7.81% of the total techniques. Psychometric analyses on the whole have a special incidence when using variables measured by tests and it is necessary to prove their reliability and validity. Therefore, they are not used to contrast fundamental hypotheses, but rather indirectly, trying to ensure the quality of the variables considered in the study.

The group composed of Manova (36), Ancova

(55) and Mancova (12) receives a frequency of use of 103, and makes up 6.65% of the total techniques used, and represents the group of techniques that aim to manage in a multivariate way the possible effect of covariables on the basic outcome variable or variables. Lastly, we would like to emphasize the grouping composed of Structural Equation Modeling (SEM) and the Mediation Model (e.g. Sobel's Test), which obtain a joint frequency of 100, and 6.46% of total techniques. SEM techniques make it possible to test interdependence models, handling multiple variables, observables and latencies, and complex chains of events, whether they be recursive or non-recursive (bidirectional relationships).

The bibliometric review focused on all sorts of methodological designs used and whether or not there was any sort of relationship with respect to the statistical techniques used. Based on the different research designs reviewed through the 623 articles, four categories were established a posteriori: Experimental Designs (123 articles), Quasi-Experimental Designs (177 articles), Surveys (198 articles) and under the category of Others (125 articles), we can find for instance, Meta-Analysis Studies, Qualitative Studies, Instrumental, Observational, Case Reports, or Theoretical. According to this taxonomy, Survey Designs occupies the first place with 31.8%, in second place the Quasi-Experimental Designs with 28.4%, Experimental Designs, with 19.7%, and lastly, the set containing the other types of research obtains 20.1%. Table 4 shows the distribution of techniques based on the type of methodological design used in each article, in accordance with the four categories established. Among the most relevant results obtained concerning the count of use of the statistical techniques based on the type of methodological design used, it is worth noting that the regression procedures are used to a greater extent in Survey type research (41.9%), followed by Quasi-Experimental (32.1%) and Experimental (17.7%). Between-Subjects Design procedures are used, practically to the same extent, in the three types of research: Quasi-Experimental (36.7%), Experimental (30.8%) and Survey (30.0%), whereas the Within-Subjects or Mixed Designs are mainly used in Experimental research (53.6%), followed by the Quasi-

Experimental type (27.2%), and to a much lesser extent, the Survey type (16.6%).

Likewise, basic procedures, such as a two-mean comparison and contingency tables (chi-square), are fundamentally used in Quasi-Experimental research (40.9%) and Surveys (38.6%), and have a lesser incidence in Experimental designs (13.4%). The reason for this distribution probably lies in the fact that both Quasi-Experimental and Survey designs, due to their lack of initial control compared to Experimental ones, need the application of basic methods that will make it possible to analyse the lack of randomization of the sample or samples under study. In this same situation we find psychometric analyses, which are more linked with Quasi-Experimental designs (41.32%) and Surveys (39.67%), whereas their incidence in Experimental designs is only 3.3%.

As far as the use of Structural equation modelling techniques is concerned, Survey designs (52.5%) are the ones that monopolize this practice, followed by Quasi-Experimental designs (20%), while they are practically non-existent among articles that applied an Experimental design (0.75%). As regards dimensionality reduction procedures of an exploratory nature, their use is strongly linked to Survey designs (59.62%), to a lesser extent to Quasi-Experimental designs (11.54%), while their presence is practically testimonial in articles with Experimental designs (3.85%).

Finally, concerning less prevalent statistical techniques in the study, it is worth noting that non-parametric techniques (54 in all) are mainly used for Quasi-Experimental designs (55.6%), followed by Surveys (22.2%) and Experimental ones (20.4%). Regarding robust techniques, which on the whole obtained a practically null overall percentage use (0.26%), these were mainly used by Quasi-Experimental designs (75%), followed by Experimental ones (25%), but with no use in Survey designs. As far as resampling techniques are concerned (Jackknife, Monte Carlo, Bootstrap), Experimental designs are the ones that most use them (33.3%), as opposed to Surveys (25%) and Quasi-Experimental (16.7%).

The study also assessed, based on the type of research applied (Experimental, Quasi-Experimen-

Table 4. Count of use of statistical methods according to methodological design

Statistical methods	Others	Experimental	Quasi Experim	Surveys	Total
1. Correlation	22	20	70	95	207
2. T-test between-subjects	9	27	66	59	161
3. Chi-Square	6	23	77	47	153
4. Reliability Analysis – ROC	19	4	50	48	121
5. Anova (A between subjects)	4	29	40	35	108
6. Linear Regression Models	8	9	30	36	83
7. Hierarchical Regression	5	9	18	39	71
8. Logistic Regression	4	10	25	27	66
9. Ancova	1	23	18	13	55
10. Non-Parametric Analysis	1	11	30	12	54
11. Anova (AxB Mixed)	0	38	6	10	54
12. Factor analysis and PCA	13	2	6	31	52
13. Structural Equation Modeling	8	3	8	21	40
14. Manova	1	10	13	12	36
15. Anova (A within subjects)	2	11	18	4	35
16. Mediation Analysis	5	11	7	7	30
17. T-test within	2	7	10	6	25
18. Multilevel Analysis	1	8	6	7	22
19. Anova (AxBxC Mixed)	0	16	2	4	22
20. Anova (AxB between subjects)	0	7	6	5	18
21. Linear Mixed Models	1	5	8	0	14
22. Qualitative methods	12	0	0	1	13
23. Resampling (Jackknife, etc.)	3	4	2	3	12
24. Anova (AxB within subjects)	0	7	4	1	12
25. Mancova	0	1	5	6	12
26. Anova (AxBxC between subjects)	0	3	6	1	10
27. Survival Analysis	3	3	3	0	9
28. Generalized Estimating Equations	0	4	5	0	9
29. Cluster Analysis	1	1	2	3	7
30. Poisson Models (NB, ZI)	0	2	2	0	4
31. Robust techniques	0	1	3	0	4
32. Taxometric procedures	3	0	0	1	4
33. Discriminant Analysis	0	0	1	2	3
34. Anova (AxBxC within subjects)	0	2	1	0	3
35. Anova Type II Random effects	1	0	0	2	3
36. Anova Type III Mixed effects	0	2	1	0	3
37. Growth Curve Analysis	0	2	0	0	2
38. Time Series Analysis	0	1	1	0	2
39. Other	2	0	0	1	3
40. Logit Models	0	0	1	0	1
41. Probit and Tobit Models	0	0	0	1	1
42. Log-Linear Models	0	1	0	0	1
43. Bloqueo (DBA, DMAG)	0	1	0	0	1
44. Factor Mixture Models	1	0	0	0	1
45. Chaos Theory	1	0	0	0	1
46. Ordinal Regression	0	0	1	0	1
Total	139	318	552	540	1549

tal, Surveys and Other), some basic parameters in order to analyze the quality of the statistical information provided by the authors, such as: effect size, use of confidence intervals, calculation of a priori power and observed power, assessment of the assumptions of the statistical models to be applied, and the solutions implemented when faced with non-compliance of statistical assumptions.

The most used effect size in the set of journals reviewed is the R squared coefficient of determination (111), followed by the eta squared coefficient (94) and a Cohen's effect size measure (82). An effect size measure is provided in 304 out of the 498 studies in which it was feasible to provide such information (Experimental study, Quasi-Experimental or Survey) during 2010 in the 8 journals consid-

ered, 61% of which appear in 71 Experimental type studies, in which the most frequent index is *etasquared*. In the 131 Quasi-Experimental type studies that provide an effect size index (74%), the most frequent is *R squared*, which is also the most frequent in the 102 Survey type studies (51.5%) which provide an effect size index. Thus, Survey designs have a lower incidence with respect to the contribution of effect size, in such a way that only half of the articles reviewed do so.

As far as the use of confidence intervals on the estimation of the parameters of different statistical models is concerned, 18.87% of articles include this information. According to types of design, the Quasi-Experimental ones provide confidence intervals of 88.14%, whereas Surveys do so with 49.49%, and to a lesser extent in Experimental designs, with 9.76%. This lower incidence in Experimental designs may take place because the experimentalist tradition, related with variance analysis techniques, generally opts for effect size indexes.

Regarding power analysis, we differentiated between the calculation of a priori power, and the calculation of observed power. Concerning the former, only 18 studies included this information, 11 of which corresponded to Quasi-Experimental type designs, and to a lesser extent to Experimental ones (3) and Surveys (4). In reference to the observed power, its prevalence of use is no better than the poor indicators of a priori power, as only 21 studies include this information, divided between Surveys (8) and Quasi-Experimental designs (8), and to a lesser extent Experimental ones (5).

Lastly, another of the basic parameters of adequacy in statistical use is the prospective assessment of the assumptions in the different statistical models. Despite the importance of this practice, only 17.27% of articles use it; depending on the type of design, Experimental ones show 66.67% use, Quasi-Experimental ones 30%, and Surveys 29.41%. If the percentage studies that include the verification of statistical assumptions associated with each technique is low, neither do the solutions when faced with non-compliance offer a good performance, as only 65 articles claim to have applied some sort of solution in case of non-compliance of

assumptions. Specifically, 22 carried out a change of statistical technique (33.85%), 18 a transformation (ordinal, logarithmic, etc.) (27.69%), 13 a correction (for instance, Greenhouse-Geisser's epsilon) (20%), whereas 12 applied a robust estimation (18.46%).

Discussion

This study, of a bibliometric nature, aims to conduct a review, with no intention of being comprehensive, of the use of statistical methodology in Clinical and Health Psychology research. In this way we hope to characterize what sort of statistical models are being applied in recent research in this field, through an analysis of all the articles published during 2010 in 8 journals with an impact factor, considering this use an acceptable empirical indicator of the degree of statistical maturity in the field. Although it is true that a greater quantitative or qualitative use of statistical methodology does not necessarily lead to greater development of scientific knowledge, it is no less true that the empirical contrast of research hypotheses can be improved insofar as the statistical models are applied appropriately, whether they be more simple or more advanced, within the wide range of techniques that are currently available to researchers, even with acceptably friendly software. Loftus (1996) clearly points out that Psychology will be a better science in as much as it changes its way of analyzing data. Data analysis must involve the consideration of any set of techniques that will optimize conditions for contrasting the hypotheses the study was designed to test, and not putting into practice a memorized set of steps or rules, in the style of a cookery book, which is probably condemned to failure.

Obviously this study does not put statistical analysis before substantive or clinical analysis of the reality under study, but rather its precise aim is to stress the impelling need to establish a link of adequation between research designs and the statistical tools to be used. It is not, therefore, a question of only assessing the quality of an article according to what statistical techniques are applied or whether it possesses more or less algorithmic complexity of

estimation. The practice of trying to apply models - the more complex the better - in research is usually a well-known phenomenon among doctorate students, in as much as this is how they seek to make their thesis more brilliant. However, the application of complex statistical models is not always the most appropriate in certain research hypotheses. Even if an analysis of the adequation between research design and statistical use was not the aim of this study, we did seek to highlight the value of different prevalences of use of a wide range of statistical techniques that are available to research in Clinical and Health Psychology.

At an empirical level, the study aimed to analyse whether a greater use of statistical techniques corresponded with the journals with a greater impact factor value. The results effectively show a non-parametric correlation of 0.619 in the sample of 8 journals considered, although at a populational level, the correlation value is not significant, given the small number of publications reviewed. Thus we cannot assure that there is a conclusive pattern concerning the fact of a greater statistical use correlating well with the obtention of a greater impact factor. This is probably the way it should be, as quantity should not be confused with adequation, that is, it is not the number of techniques used which should be relevant, but rather the use of the most appropriate technique, in other words the most powerful one, on each occasion.

As far as the techniques used is concerned (a total of 1549), the main results obtained through the 623 articles reviewed, point towards a prevalent use of the more conventional statistical techniques, as the top 8 most frequently used techniques are, in order, Correlation (13.36%), Between-Subjects T-Test (10.39%), Chi-Square (9.88%), Reliability Analysis-ROC curves (7.81%), Between-Subjects One-Way Anova (6.97%), Linear Regression Models (5.36%), Hierarchical Regression (4.58%) and Logistic Regression (4.26%). On the whole they make up 62.61% of all statistical techniques applied. These are well-known techniques, with a long tradition in psychological research, generally in the statistical field of the General Linear Model.

Introducing a logical pattern of groups or families of techniques, the techniques that are applied in

order to assess the degree of a priori homogeneity between the groups or sub-samples used - such as Between-Subjects T-Test, Chi-square test or Correlation - make up a third of the total techniques used (33.63%), and are not generally used for empirical contrast of the basic hypotheses of each study. The most prevalent techniques to test the main hypotheses of the articles reviewed are, on the one hand, the group of techniques related to Variance Analysis, which make up the second third of statistical models applied (31.31%), and, on the other hand, techniques related to Regression Models (Linear, Hierarchical, Logistic, Poisson, Ordinal), with 14.20%. That is to say, if we consider the so-called basic techniques, Anova and Linear regression models make up 79.14% of all techniques applied. And if, as well as these three blocks, we include the classic psychometric analysis techniques which aim to assess the quality of the measures used, and which make up 7.81%, together they represent 86.95% of all the statistics used.

Characterization of statistical use according to the type of methodological design used did not offer an excessively disparate association pattern, although the most important results indicate that Survey designs make a greater use of the techniques related with Regression Models, whereas the techniques related with Variance Analysis appear to be more linked to Experimental and Quasi-Experimental designs.

All these results show that the main set of techniques is located within the framework of the General Linear Model, and only 13% of techniques represent more complex statistical methods or, perhaps, lesser known ones. Thus, for instance, the use of Structural Equation Modeling represents only 2.58%, Multilevel Analysis only 1.42%, and GEE models (Generalized Estimating Equations) 0.58%.

Lastly, as far as statistical methods that handle categorical variables, the results show a practically irrelevant use (0.19% of all techniques applied). This highlights a clear tendency towards the use of variables or quantitative indicators, to the detriment of those of a categorical nature. Underlying this low incidence, there is probably an important degree of ignorance, but above all, difficulty in handling or interpretation. The same happens in relation to other

more complex newfangled methods, which have not been used, such as Artificial Neural Networks, Support Vector Machine, Latent Class or Mixture Models. In this respect, and following the recommendations of von Eye and Schuster (2000), there is a bright future opening up for the integration of these research techniques, insofar as there is a close interdisciplinary collaboration between psychologists specialized in statistical methodology and psychologists in Clinical and Health Psychology. It does not seem worth a researcher in this field devoting an enormous amount of time on such specialized statistical training, when it is possible to establish important synergy with behaviour and health science methodologists. Supposing this collaboration is feasible, and lesser known statistical methods - because of their complexity or their novelty - can be included, Palmer et al. (2005) recommend the authors endeavour to give their articles a didactic slant, in order to make them more accesible to would-be readers and, thereby, encourage their use and the development of their potential in applied research.

Von Eye and Schuster (2000) endorse the idea of interdisciplinary synergy with methodologists as, otherwise, the choice between methods of statistical analysis is becoming more and more difficult, and the cases of improper use of statistics are increasingly more frequent. Anyway, for the synergy to be completely efficient, the role of methodologists must change towards powerful education in statistics but applied to psychological research.

In relation to the possible improper use of statistical techniques, the study highlights some important shortcomings concerning relevant statistical information that is not provided. A clear example of this shortcoming is the provision of effect size, which only takes place in 52.78% of the studies published through the 8 journals during 2010. Another alternative way of offering effect sizes is to provide the confidence intervals for the estimated parameters, which only appear in 18.87% of articles. It must be taken into account that without this information it is much more complicated to conduct an empirical analysis of the substantive or clinical significance of a certain effect, correlation, difference, discrepancy, etc. Hence, it would seem

to be necessary for research teams to make an effort to clearly include this information, not only concerning statistical, but also substantive, significance.

Another important informative shortcoming refers to power analysis, as only 3% of the studies include the estimated value of a priori power, and 3.64% that of the observed or empirical power. The power of a statistical test is the likelihood of the null hypothesis being rejected when it is false, and is associated with a type II error, which occurs when the researcher does not reject the null hypothesis when it is false in the population. As the power increases, the type II error decreases. Therefore, given the importance of power for any statistical test, power analysis can be used to calculate the minimum sample size required in order to obtain a reasonable likelihood of detecting an effect of a certain size. Besides, power analysis can also be used to calculate the minimum effect that may be detected in a study with a certain sample size. The small percentage of studies that mention carrying out power analysis is truly worrying and it would be necessary to drastically improve these figures.

As far as the assessment of assumption compliance corresponding to each statistical test, only 17.27% of the studies mention carrying out an analysis of assumption compliance. This low incidence is as or even more worrying than the shortcomings referred to above, as the applications of techniques when faced with possible non-compliance of assumptions may compromise the veracity of the statistical conclusions obtained. Finally, it is worth noting that only 11.28% of the studies refer to having applied some sort of solution in the face of non-compliance of statistical assumptions. With these results we cannot assert that the real situation regarding the quality of the generation of statistical inference is negative, but in the absence of information concerning assumption analysis, many of the statistical conclusions generated may be compromised. Hence the importance of implementing and referencing assumption analysis and their result in each study.

With the results obtained by this study, we hope to be able to contribute to offering a general view of the degree of statistical prowess possessed by cur-

rent research in Clinical and Health Psychology, and by keeping an uncensored, critical attitude, we expect, in the short to medium term, the shortcomings detected will act as a motivating element for quantitative and qualitative improvement of the application of existing statistical methodology. In fact, we believe that progress in understanding the phenomena that are the object of study in Clinical and Health Psychology is a triply sophisticated demand for field researchers: reflexive articulation of theoretical models, experience in designing research methodology and statistical rigour. Therefore, the active implication of methodologists in interdisciplinary research teams is essential. To close, we make our own the statement by Treat and Weersing (2005), who claim that the next generation of clinical and health psychologists should probably be known in part for their degree of sophistication in statistical usage.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, New Jersey: Erlbaum.
- Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, 23, 12-15.
- Chow, S. L. (1996). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Edition)*. Hillsdale, New Jersey: Erlbaum.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, New York: Lawrence Erlbaum Associates.
- Cumming, G. and Finch, S. (2001). A primer on the understanding use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61, 530-572.
- Everitt, B. S. (2000). Latent variables, factor analysis and causal modeling. *Comprehensive Clinical Psychology*, 3, 287-311.
- Fritz, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1, 379-390.
- Harlow, L. L., Mulaik, S. A. and Steiger, J. H. (1997). *What if there were no significance tests?* Hillsdale, New Jersey: Erlbaum.
- Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application and implementation. *Journal of Statistical Software*, 20, 1-24.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746-759.
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current directions in psychological science*, 5, 161-171.
- Palmer, A., Sesé, A. y Montaña, J. J. (2005). Tourism and statistics: Bibliometric study 1998-2002. *Annals of Tourism Research*, 32, 167-178.
- Robinson, D. H. and Wainer, H. (2001). *On the past and future of null hypothesis significance testing*. Princeton, New Jersey: Educational Testing Service.
- Rosenthal, R. and Rubin, D. B. (1994). The counternull value of an effect size: A new statistic. *Psychological Science*, 5, 329-334.
- Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, 1, 115-129.
- Smithson, M. (2003). *Confidence intervals*. Quantitative Applications in the Social Sciences Series, n° 140. Belmont, CA: SAGE Publications.
- Snyder, P. and Lawson, S. (1993). Evaluating results using corrected and uncorrected effect sizes estimates. *Journal of Experimental Education*, 61, 334-349.
- Treat, T. A. and Weersing, V. R. (2005). Clinical Psychology. In B. S. Everitt and D. C. Howell *Encyclopedia of Statistics in Behavioral Science*. New York John Wiley and sons.
- von Eye, A. and Schuster, C. (2000). The road to freedom: Quantitative developmental methodology in the third millennium. *International Journal of Behavioral Development*, 24, 35-43.
- Wainer, H. (1999). One cheer for null hypothesis

significance testing. *Psychological Methods*, 4, 212-213.

Wainer, H. and Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance test-

ing. *Educational Researcher*, 32, 22-30.

Wilkinson, L. (1999). Statistical methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, 54, 594-604.

Artículo recibido: 01/07/2011

Revisión recibida: 15/09/2011

Aceptado: 22/09/2011