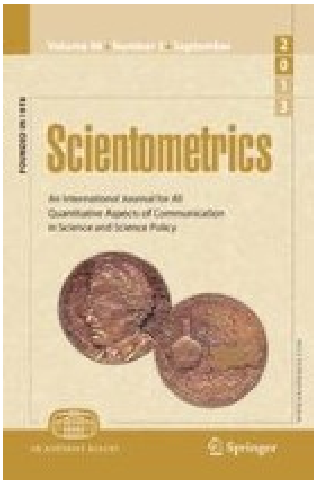
A mixed longitudinal and cross-sectional model to forecast the journal impact factor in the field of Dentistry

- Pilar Valderrama, Manuel Escabias, Evaristo Jiménez-Contreras, Mariano J. Valderrama, Pilar Baca

- A mixed longitudinal and cross-sectional model to forecast the journal impact factor in the field of Dentistry

- *Scientometrics 116, 1203–1212 (2018)*
- DOI: https://doi.org/10.1007/s11192-018-2801-z



Scientometrics (2018) 116:1203–1212
https://doi.org/10.1007/s11192-018-2801-z

A mixed longitudinal and cross-sectional model to forecast the journal impact factor in the field of Dentistry

Pilar Valderrama[1] · Manuel Escabias[1] · Evaristo Jiménez-Contreras[2] · Mariano J. Valderrama[1] · Pilar Baca[3]

Received: 4 April 2018 / Published online: 8 June 2018

Abstract In order to estimate the impact factor value for a journal in Dentistry, two sets of variables were considered in this study: the first takes in the longitudinal behavior of the process specified in the slope and intercept of the straight line fitted to the trend of the last years, whereas the second considers the percentage of review papers published each year and the adhesion degree of the journal to ICMJE guidelines. The final estimated model showed a high determination coefficient (99.3%) and its performance was tested on a new set of journals randomly sampled from the list of journal citation reports.

Keywords Journal impact factor · Multiple regression · Longitudinal parameters · ICMJE

✉ Mariano J. Valderrama
valderra@ugr.es

Pilar Valderrama
piluvb95@ugr.es

Manuel Escabias
escabias@ugr.es

Evaristo Jiménez-Contreras
evaristo@ugr.es

Pilar Baca
pbaca@ugr.es

1 Department of Statistics and Operations Research, University of Granada, Campus de Cartuja, 18071 Granada, Spain

2 Department of Information and Communication, University of Granada, Campus de Cartuja, 18071 Granada, Spain

3 Department of Dentistry, University of Granada, Campus de Cartuja, 18071 Granada, Spain

Introduction

Clinical fields rely on informative guidelines for publishing the outcomes of research in journals. These guidelines specify the minimum information to be included in a research report so that the reader can assess the study and its results. Among them we can cite the CONSORT Declaration for randomized trials, STROBE for observational studies, PRISMA for systematic reviews, CARE for case reports, SRQR for qualitative research, SPIRIT for study protocols, and SQUIRE for quality improvement studies. Furthermore, in a general context, the International Committee of Medical Journal Editors (ICMJE) provides a list of recommendations for conducting, reporting, editing, and publishing scholarly work in medical journals; the most recent list came out in December 2017 (www.ICMJE.org). It is logical that journals specifying the obligation or, at least, the recommendation to adhere to such norms (in their instructions to authors) should have a higher impact than those not including any reference to them.

It is a well-known fact that journals including review articles, especially systematic reviews, by prominent authors receive more citations than others. However, we lack studies that quantify the actual influence of review papers on the impact factor itself. Furthermore, a very important facet of the Journal Impact Factor (JIF) is its own evolution over a previous interval of time. The inclusion of variables on information relative to this characteristic, such as trend or initial value in the study period, provide for an important longitudinal nature that deserves consideration.

In this framework, a previous paper by Valderrama et al. (2018) developed an ordinal regression model to estimate the tercil occupied by the JIF in the field of Dentistry in terms of the set of covariables: *h*-index of the Editor-in-chief, percentage of papers published in the journal whose research received public or private financial support, and the average number of papers yearly published in a journal, and the factors: scope of the journal (specialized in a concrete topic or generalistic), and its internal structure (including survey papers, theoretical, applied, etc.). Earlier work along this research line (Valderrama et. al 2017) involved the introduction of a binary level variable and the estimation of a logistic model. Previously, Park (2015) used logistic regression and empirical analysis to verify whether a national technology innovation R&D program's performance followed the stepwise chain structure of typical R&D logic models. Contreras et al. (2006) estimated the long-term impact of journals aggregated in 24 different fields using a simple logistic diffusion model, relating their results to the current impact factor. Recently, Li et al. (2017) investigated the degree of personal citation in Chemistry, Mathematics, and Physics, as well as the factors influencing it, such as total citations, *h*-index, and citations per publication, applying binary logistic regression.

In the field of clinical medicine, an alternative means of appraising research collaboration and authorship trends in Malaysia from 2001 to 2010 was proposed by Low et al. (2014), using regression trees. More concretely in Dentistry, Cheng et al. (2017) estimate and identify by multiple regression those factors associated with citation rates in oral and maxillofacial surgery literature, adopting as predictor variables the authorship and specific article features, whereas the outcome variable was the citation rate defined as the total number of citations for each article over a 4-years period.

More sophisticated regression techniques to predict metrics, namely JIF, were used by Yu and Yu (2016), who applied multivariate regression and quantile regression to study the relationship between average JIF percentile and other bibliometric indicators. Qian et al. (2017) drew on GLIM regression, concretely Poisson or negative binominal regression, to deal with the citation rate as outcome, as it is a counting variable, being the classification of a publication, number of authors, maximum *h*-index of all authors and average number of papers published, the set of explicative variables.

On the basis of 80 selected journals records from 1992 to 2003 in the category *Public, Environmental and Occupational Health,* an early effort along these lines was that of López-Abente and Muñoz-Tinoco (2005) that estimated a linear regression model where the dependent variable was the JIF and the independent variable the year. Then, the slope of the model and its statistical significance were taken as the indicator of annual change. On the basis of this work, Smith (2008) performed a longitudinal citation-based analysis on 5 core journals belonging to the field *Occupational Medicine* published between 1985 and 2006, and confirmed that the absolute number of citations received each year is steadily increasing.

Yu et al. (2014) used stepwise multiple regression analysis to select appropriate features (external, authors, citations, etc.) to derive a regression model that would explain the relationship between citation impact and the chosen features; they tested the validity of this model in the subject area of *Information Science & Library Science*. Even more recently, Ayaz et al. (2018) evaluated different *h*-index prediction models for the field of *Computer Science* by means of regression models with parameters comprising current *h*-index, average citations per paper, number of coauthors, years since publishing first article, number of publications, number of impact factor publications, and number of publications in distinct journals.

The current paper goes one step beyond, and proposes quantitative estimation of the JIF by means of some new variables representing quality characteristics while also considering the dependence over time of the stochastic process involved. Specifically, we consider data on Dentistry journals in 2016 and the model will have two parts: the first includes a set of cross-sectional variables recorded in the same year, such as degree of adaptation of guidelines to authors in 2016 to certain publication guides (e.g. ICMJE, CONSORT, etc.), and percentage of literature and systematic reviews published in the issues of that year; and the second one takes into account characteristics reflecting the evolution of the process, such as the trend and starting point of the time series of the journals sampled since 2007.

Once the model is estimated by means of a stepwise procedure, selecting the influential variables and excluding the non-significant ones, it is tested with new journal data not included in the training phase.

Methodology

In developing a multiple quantitative model to explain the JIF in the field of *Dentistry, Oral Surgery and Medicine* from the Journal Citation Reports by the Institute for Scientific Information (Thompson Reuters) corresponding to 2016, we performed a stratified random sampling by quartiles that provided 30 journals for the study (Table 1) besides 10 more journals (Table 2) that were used as the test sample. The following explicative covariables were initially considered:

- *Slope* and *intercept* In view of the JIF over time from 2007 to 2015, a linear equation was fitted to the journal series through a least squares procedure, and the slope and intercept of each straight line were calculated. Some journal records were incomplete in this interval because they were included in JCR after 2007.
- *Percentage of review articles (PRev)* From the *PubMed* database we extracted the number of reviews in relation to total papers published in each journal in 2016. Because

Table 1 Slope, intercept and linear determination coefficient (R^2) for each journal in the training sample, percentage of review papers (% Rev) in 2016 and their Anscombe transformation (AnsPRev), adhesion degree to ICMJE and CONSORT Declaration and scope of the journal

Journal	Slope	Intercept	R^2 adj.	% Rev	AnsPRev	ICMJE	CONS	Scope
J. Dental Research	0.160	3.002	0.840	17.0	0.425	2	2	0
Dental Materials	0.154	2.609	0.758	3.5	0.187	2	0	1
Int. J. Oral Sciences	0.431	0.000	0.887	6.3	0.253	2	0	1
Clinical Oral Implants Research	0.151	2.237	0.686	5.5	0.237	2	2	0
J. Clinical Periodontology	0.000	3.029	0.285	14.9	0.397	2	2	1
J. Dentistry	0.163	1.749	0.748	17.3	0.429	2	2	1
Int. Endodontic J.	0.083	1.998	0.430	10.0	0.322	2	2	1
J. Endodintics	0.000	3.108	0.000	8.0	0.286	2	2	0
J. Oral Facial Pain Headache	0.254	0.000	0.426	7.0	0.267	1	0	0
J. Evidence-Based Dental Practice	0.198	0.000	0.412	34.9	0.632	2	0	1
Clinical Oral Investigations	0.033	2.028	0.409	6.5	0.257	1	0	1
Int. J. Oral & Maxillofacial Implants	0.000	1.800	0.000	13.4	0.375	1	0	0
J. Oral Maxillofacial Surgery	0.046	1.237	0.484	17.8	0.435	2	2	1
Gerodontology	0.150	0.000	0.493	9.2	0.308	2	2	0
European J. Orthodont.	0.076	0.769	0.695	16.7	0.421	0	2	1
Odontology	0.000	0.000	0.281	10.0	0.322	0	0	1
J. Cranio-Maxillofacial Surgery	0.000	1.079	0.246	13.2	0.372	2	2	1
Dental Traumatology	0.000	1.127	0.131	8.2	0.291	2	2	1
Int. J. Prosthodontics	0.000	1.417	0.000	5.8	0.243	1	0	0
J. Public Health Dentistry	0.065	0.879	0.439	4.1	0.203	1	2	0
Head & Face Medicine	0.161	0.000	0.777	5.6	0.238	0	2	1
Int. Dental J.	0.073	0.557	0.714	5.4	0.234	2	2	1
Int. J. Dental Hygiene	0.152	0.000	0.779	19.2	0.454	0	1	1
J. Esthetic & Restorative Dentistry	0.124	0.000	0.691	2.0	0.142	0	0	1
Int. J. Periodontics & Restorative Dentistry	0.000	1.596	0.312	0.8	0.089	0	0	1
J. Advanced Prosthodont.	0.126	0.000	0.855	2.9	0.170	0	0	0
British Dental J.	0.000	0.984	0.000	6.0	0.247	0	2	0
Quintessence International	0.026	0.637	0.356	24.5	0.517	0	2	1
J. Oral Sciences	0.110	0.000	0.581	2.5	0.160	0	0	0
Australian Endodontic J.	0.000	0.000	0.290	5.7	0.241	2	1	0

the proportion varies in the interval [0,1], we applied a transformation (Anscombe 1948) that stabilizes the variance and converts it into a new variable *AnsPRev*, with range $(-\infty,\infty)$.

Table 2 Regression coefficients and significance (p values) when introducing all the variables and after the stepwise estimation

Variables	Introducing all variables	Stepwise regression
Constant	− 0.178 (0.431)	
ICMJE	0.225 (0.000)	0.212 (0.000)
CONSORT	− 0.047 (0.309)	
Slope	8.185 (0.000)	8.052 (0.000)
Intercept	0.856 (0.000)	0.834 (0.000)
AnsPRev	1.048 (0.005)	0.613 (0.008)
Scope	0.020 (0.831)	
R^2 adjusted	0.967	0.993

As factors we include:

- *Compliance to ICMJE and CONSORT Declaration* According to the degree of fitness of a journal to both these guides they were assigned the following scores: 0 (not mentioned), 1 (recommended) and 2 (required).
- *Journal scope (Scope)* Depending on the nature of the journal it was rated as: 0 (generalist) or 1 (specialized).

In a first step all the factors and co variables were included as possible explicative variables in the model; then, in light of their significance, a second stepwise estimation phase selected the most influential variables to be considered.

All the statistical calculations were performed by using R software, version 3.4.4, for x86_64-pc-windows-gnu (www.R-project.org).

Results

Having introducing all the covariables and factors in the model the estimation procedure concludes that neither adhesion to CONSORT Declaration, journal scope, nor the constant term exert a significant influence on the response variable, so that after re-estimating by means of a stepwise algorithm, we obtain the significant variables given in Table 2.

Therefore, the final multiple regression model to estimate the JIF in 2016 is given by:

$$\text{JIF16} = 0.212\ \text{ICMJE} + 8.052\ \text{Slope} + 0.834\ \text{Intercept} + 0.613\ \text{AnsPRev} \tag{1}$$

Table 3 contains the real and estimated JIF of the 30 journals included in the training sample, as well as the corresponding 95% level confidence interval (95% CI) for each by applying the above model. Moreover, Fig. 1 shows for each journal the JIF punctual estimation and the 95% CI.

In Table 4 ten new journals not previously considered in the estimation phase are listed with the values of the significant variables as well as their true and forecasted JIF values, together with the predictive 95% CI. The corresponding graphics are shown in Fig. 2. As can be seen, in seven cases the true value of the JIF was within the 95% CI associated with the forecast value.

Table 3 Real JIF of the sampled journals in 2016, JIF estimated through the model and 95% CI

Journal	JIF 2016	Estim. JIF	95% CI
J. Dental Research	4.755	4.477	4.032–4.922
Dental Materials	4.070	3.955	3.502–4.408
Int. J. Oral Sciences	3.930	4.050	3.550–4.550
Clinical Oral Implants Research	3.624	3.651	3.212–4.090
J. Clinical Periodontology	3.477	3.194	2.750–3.637
J. Dentistry	3.456	3.458	3.032–3.885
Int. Endodontic J.	3.015	2.956	2.530–3.383
J. Endodontics	2.807	3.191	2.745–3.638
J. Oral Facial Pain Headache	2.760	2.421	1.982–2.860
J. Evidence-Based Dental Practice	2.477	2.406	1.943–2.869
Clinical Oral Investigations	2.308	2.327	1.900–2.753
Int. J. Oral & Maxillofacial Implants	2.263	1.943	1.513–2.373
J. Oral Maxillofacial Surgery	1.918	2.093	1.659–2.527
Gerodontology	1.681	1.821	1.380–2.262
European J. Orthodont.	1.622	1.511	1.069–1.953
Odontology	1.602	1.790	1.356–2.223
J. Cranio-Maxillofacial Surgery	1.583	1.552	1.113–1.992
Dental Traumatology	1.413	1.542	1.108–1.978
Int. J. Prosthodontics	1.386	1.543	1.123–1.963
J. Public Health Dentistry	1.378	1.593	1.178–2.008
Head & Face Medicine	1.370	1.442	1.012–1.872
Int. Dental J.	1.362	1.620	1.187–2.053
Int. J. Dental Hygiene	1.358	1.502	1.053–1.951
J. Esthetic & Restorative Dentistry	1.273	1.085	0.664–1.507
Int. J. Periodontics & Restorative Dentistry	1.113	1.386	0.955–1.817
J. Advanced Prosthodont.	1.027	1.119	0.696–1.541
British Dental J.	1.009	0.972	0.546–1.398
Quintessence International	0.995	1.058	0.597–1.518
J. Oral Sciences	0.876	0.984	0.563–1.404
Australian Endodontic J.	0.838	0.572	0.117–1.027

Discussion

In this work we expound a mixed cross-sectional and longitudinal statistical model used to estimate the JIF in the field of *Dentistry, Oral Surgery and Medicine* in the year 2016. The initial variables were the percentage of review papers included in all the issues of a journal in 1 year, the compliance to ICMJE and CONSORT Declaration as the guide for authors, and the scope of the journal, divided into generalist or specialized. In the longitudinal part of the model we included the slope and intercept of the straight line adjusted from 2007 to 2015 for each journal, even though some JIF records began after the starting year.

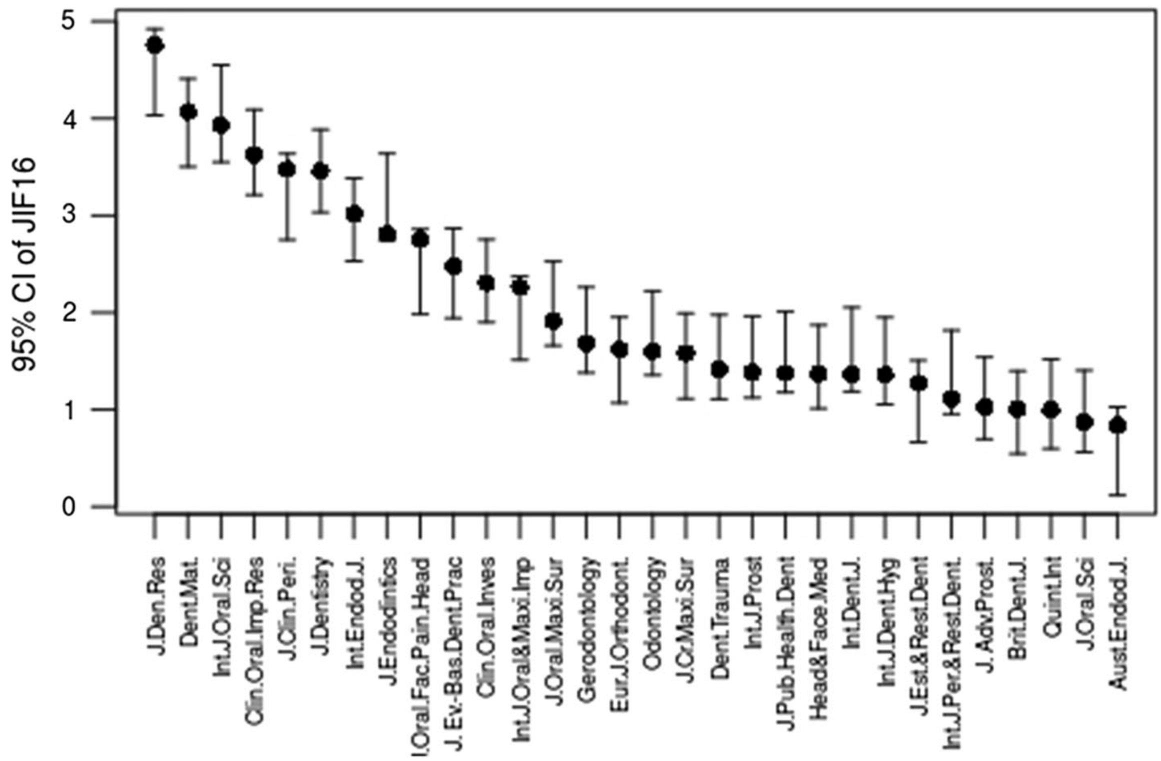


Fig. 1 JIF punctual estimation and 95% CI for the journals of the training sample

Table 4 Slope, intercept and linear determination coefficient (R^2) for each journal in the forecasting sample, percentage of review papers (% Rev) in 2016, adhesion degree to ICMJE, JIF forecast and 95% CI

Journal	Slope	Intercept	R^2 adj	% Rev	ICMJE	JIF16	Pred	95% CI
J. Periodontology	0.111	1.866	0.855	0.096	2	3.030	3.067	2.641–3.494
Mol. Oral Microbiology	0.425	0.000	0.749	0.122	2	2.908	4.065	3.576–4.554
Operative Dentistry	0.181	0.000	0.480	0.027	2	2.893	1.982	1.537–2.429
J. Adhesive Dentistry	0.000	1.492	0.000	0.016	1	2.008	1.534	1.113–1.954
Med. Oral. Patol. Oral y Cir. Bucal	0.144	0.000	0.626	0.296	0	1.156	1.513	1.044–1.981
J. Dental Education	0.000	1.112	0.326	0.027	2	0.927	1.453	1.018–1.888
Cranio	0.000	0.477	0.293	0.099	1	0.877	0.806	0.378–1.235
J. Clinical Pediatric Dentistry	0.000	0.000	0.159	0.047	0	0.775	0.134	0.000–0.557
Swedish Dental J.	− 0.116	1.547	0.431	0.000	0	0.581	0.356	0.000–0.789
Australian Orthodontic J.	0.000	0.000	0.000	0.034	2	0.423	0.538	0.085–0.991

The estimation of the model through a stepwise procedure concluded that neither adhesion to the CONSORT Declaration nor the scope of the journal had a significant influence upon the explanation of the JIF, whereas the remaining variables must be taken into account. The reason for this would be that CONSORT is a guideline for researchers focused on clinical essays, yet not all of the total journals considered have this aim. On the other hand, the character, whether generalist or specialized, of a journal could bear some

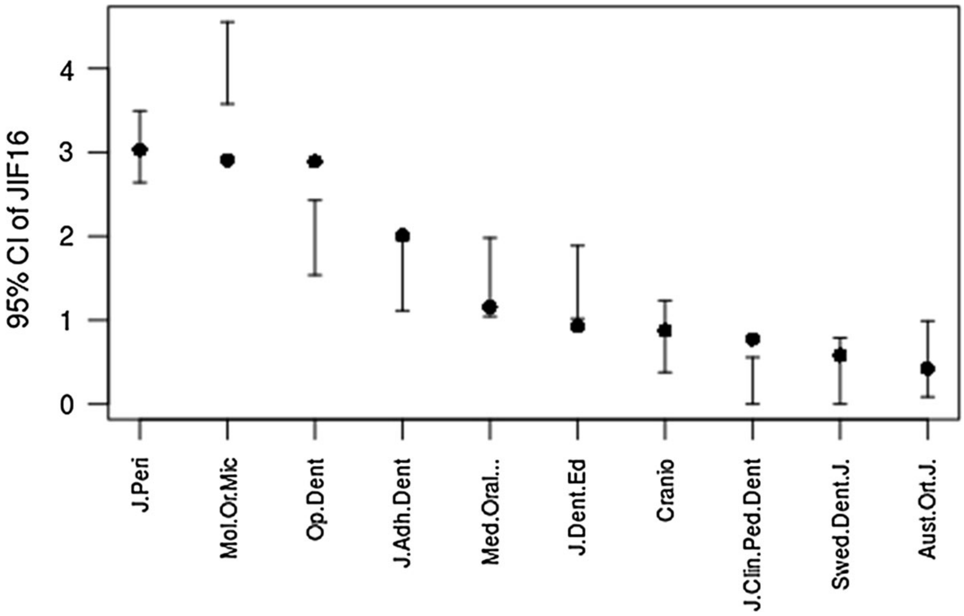


Fig. 2 JIF punctual estimation and 95% CI for the journals of the forecasting sample

influence in isolated fashion regarding the JIF estimation but, together other factors and co variables, its effect is absorbed by them.

One of the most innovative aspects of this work is the incorporation of the time evolution component of the JIF stochastic process by means of the slope and intercept of an adjusted straight line to the sampled journals. Although a dependence on the past is surmised, the results of this research quantitatively confirm the influence of both parameters in the explicative model.

Likewise, a second interesting co variable whose contribution to the model's performance has been analyzed is the percentage of review papers as opposed to the total of published articles in a year; indeed the greater this percentage, the higher the JIF of the journal. This result is consistent with the one reported by Chew et al. (2007) that recorded views of editors of seven outstanding medical journals and concluded that, among possible reasons given for rises in citation counts, it can be included active recruitment of high-impact articles by relevant researchers as well as publishing more review papers. Moreover, on the basis of a previous paper by Falagas and Zouglakis (2006), a longitudinal analysis carried out between 1999 and 2008 by Chen et al. (Chen et al. 2011), focused in the field of *Rheumatology*, reinforced the idea that the review journals have more rapid increase in JIF than those publishing original papers.

It must be taken into account that the current work does not discriminate among the different types of reviews (literature reviews, systematic reviews, meta-analysis reviews, etc.), instead including all of them in the same group. It would no doubt prove interesting to analyze some of them separately, most importantly the systematic reviews.

Finally, because the ICMJE provides a list of recommendations that clinical journals must include in their instructions to authors, the degree of fit to this list on a nominal scale of 0 (not included), 1 (recommended) and 2 (required) was also considered in the current paper as a significant factor in the model.

Equation (1) should not be viewed as a magic formula to estimate the JIF in terms of the mentioned independent variables, because it is estimated in a concrete time, the year 2016, among a sample set of journals from the total included in Journal Citation Reports. In fact the study is focused only in Dentistry, that is a specific field of Medicine. The objective rather is to reveal the main independent inputs that allow one to estimate the JIF.

Conclusions

Among the possible influential variables in the estimation of JIF, the only ones showing a significant effect have been classified in two groups: cross-sectional and longitudinal variables. The first group includes adhesion of the journal to the recommendations of the Committee of Medical Journal Editors (ICMJE) and the percentage of review-papers, while the second one contains the parameters of the least-squared fitted straight line to JIF evolution on the last 8 years. Neither adhesion to CONSORT Declaration guideline nor scope of the journal had significant effect on the JIF estimation.

References

Anscombe, F. J. (1948). The transformation of Poisson. Binomial and negative-binomial data. *Biometrika, 35,* 246–254.

Ayaz, S., Masood, N., & Islam, M. A. (2018). Predicting scientific impact based on *h*-index. *Scientometrics, 114*(3), 993–1010. https://doi.org/10.1007/s11192-017-2618-1.

Chen, M., Zhao, M. H., & Kallenberg, C. G. (2011). The impact factor of rheumatology journals: An analysis of 2008 and the recent 10 years. *Rheumatology International, 31*(12), 1611–1615. https://doi.org/10.1007/s00296-010-1541-z.

Cheng, K. L., Dodson, T. H., Egbert, M. A., & Susarla, S. M. (2017). Which factors affect citation Rates in the oral and maxillofacial surgery literature? *Journal of Oral and Maxillofacial Surgery*, 75, 1313–1318. https://www.sciencedirect.com/science/article/pii/S0278239117303610?via%3Dihub.

Chew, M., Villanueva, E. V. & Van Der Weyden, M. B. (2007). Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994–2005) and their Editors' views. *Journal of the Royal Society of Medicine*, 100(3), 142–150. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1809163/.

Contreras, C., Edwards, G., & Mizala, A. (2006). The current impact factor and the long-term impact of scientific journals by discipline: A logistic diffusion model estimation. *Scientometrics, 69*(3), 689–696. https://doi.org/10.1007/s11192-006-0165-2.

Falagas, M. E. & Zouglakis, G. M. (2006). Trends in the impact factor of scientific journals. *Mayo Clinic Proceedings*, 81(10), 1401–1402. https://search.proquest.com/docview/216868735/12959745A4AB4D7CPQ/19?accountid=14542.

Li, X., Wu, Q., & Liu, Y. (2017). A quantitative analysis of researcher citation personal display considering disciplinary differences and influence factors. *Scientometrics, 113*(2), 1093–1112. https://doi.org/10.1007/s11192-017-2501-0.

López-Abente, G., & Muñoz-Tinoco, C. (2005). Time trends in the impact factor of public health journals. *BMC Public Health, 5,* 24. https://doi.org/10.1186/1471-2458-5-24.

Low, W. Y., Ng, K. H., Kabir, M. A., Koh, A. P., & Sinnasamy, J. (2014). Trend and impact of international collaboration in clinical medicine papers published in Malaysia. *Scientometrics, 98*(2), 1521–1533. https://doi.org/10.1007/s11192-013-1121-6.

Park, S. (2015). The R&D logic model: Does it really work? An empirical verification using successive binary logistic regression models. *Scientometrics, 105*(3), 1399–1439. https://doi.org/10.1007/s11192-015-1764-6.

Qian, Y., Rong, W., Jiang, N., Tang, J., & Xiong, Z. (2017). Citation regression analysis of computer science publications in different ranking categories and subfields. *Scientometrics, 110*(3), 1351–1374. https://doi.org/10.1007/s11192-016-2235-4.

Smith, D. R. (2008). Citation analysis and impact factor trends of 5 core journals in occupational medicine, 1985–2006. *Archives of Environmental & Occupational Health*, 63(3), 114–122. https://search.proquest.com/docview/201150096/6D830AD0522B4C17PQ/3?accountid=14542.

Valderrama, P., Escabias, M., Jiménez-Contreras, E., Rodríguez-Archilla, A., & Valderrama, M. J. (2018). Proposal of a stochastic model to determine the bibliometric variables influencing the quality of a journal: Application to the field of dentistry. *Scientometrics, 115*(2), 1087–1095. https://doi.org/10.1007/s11192-018-2707-9.

Valderrama, P., Escabias, M., Jiménez-Contreras, E., Valderrama, M. J., & Baca, P. (2017). Bibliometric variables determining the quality of a dentistry journal. In Skiadas, C. H. (ed.) *Proceedings of the 17th Conference of the Applied Stochastic Models and Data Analysis International Society* (pp. 825–831).

Yu, L., & Yu, H. (2016). Does the average JIF percentile make a difference? *Scientometrics, 109*(3), 1979–1987. https://doi.org/10.1007/s11192-016-2156-2.

Yu, T., Yu, G., Li, P. Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics, 101*(2), 1233–1252. https://doi.org/10.1007/s11192-014-1279-6.