COMPARING QUALITY OF REPORTING BETWEEN PREPRINTS AND PEER-REVIEWED ARTICLES IN THE BIOMEDICAL LITERATURE

Clarissa F. D. Carneiro[a]*, Victor G. S. Queiroz[a], Thiago C. Moulin[a], Carlos A. M. Carvalho[b,c,d], Clarissa B. Haas[e], Danielle Rayêe[f], David E. Henshall[g], Evandro A. De-Souza[a], Felippe Espinelli[a], Flávia Z. Boos[h], Gerson D. Guercio[i], Igor R. Costa[a], Karina L. Hajdu[f], Lieve van Egmond[j], Martin Modrák[k], Pedro B. Tan[f], Richard Abdill[l], Steven J. Burgess[m], Sylvia F. S. Guerra[c,d,n], Vanessa T. Bortoluzzi[o] and Olavo B. Amaral[a]

*clarissa.franca@bioqmed.ufrj.br

Most of biomedical research becomes available after being assessed by editors and peer reviewers in a scientific journal. This system began to be established in the 1700s, but it was not until recent decades that it became ubiquitous. Prepublication peer review has relatively little empirical evidence to support it, and its effectiveness in improving the quality of the scientific literature has been questioned; nevertheless, the system remained largely unchallenged until the advent of preprint publication. Deposit of preprints has been the norm in some fields of science for more than two decades and has recently begun to rise in popularity among biomedical researchers. Still, doubts have been raised about the quality of unreviewed articles, and debates on the theme often question whether it is worth for researchers to spend time sifting through papers that have not been previously evaluated, and whether publication of non-reviewed articles might prohibitively increase the amount of noise in the literature.

With that in mind, this study aims to compare the quality of reporting between biomedical research articles that have been published in peer-reviewed journals and those deposited as preprints, in order to assess whether an objective dimension of literature quality differs between these two types of articles.

Articles were selected from bioRxiv or PubMed starting from an empty search with publication dates between January 1st and December 31st 2016. To be included, it had to be written in English and contain an original result, which had to compare statistically two groups composed of humans, other animals, cell lines or microorganisms. Articles were assessed for inclusion at random. Final sample size was determined based on a preliminary analysis containing 10 articles/group. Some articles were further excluded so that groups from both databases contained the same number of articles per category of biological model.

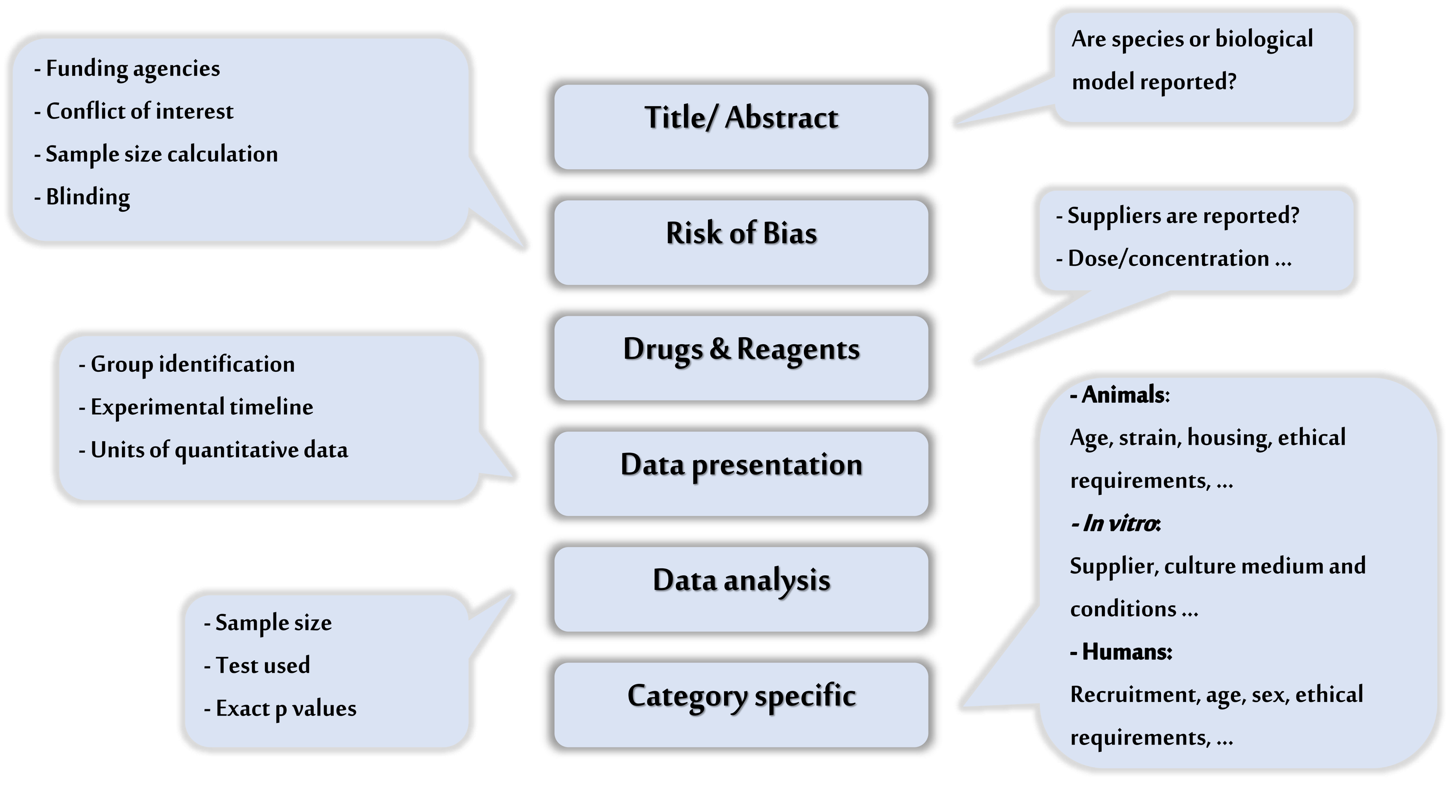

Figure 1 – Schematic representation of the questionnaire used and example items assessed.

Assessment of reporting was performed through an online form developed based on previously established guidelines and journal checklists. Each article was assessed by three evaluators and the prevalent answer was considered final. Each reported item counted as 1 point and total score was normalized to the total of applicable questions. Whenever a prevalent answer was not reached, final decision was discussed by the coordinating team.

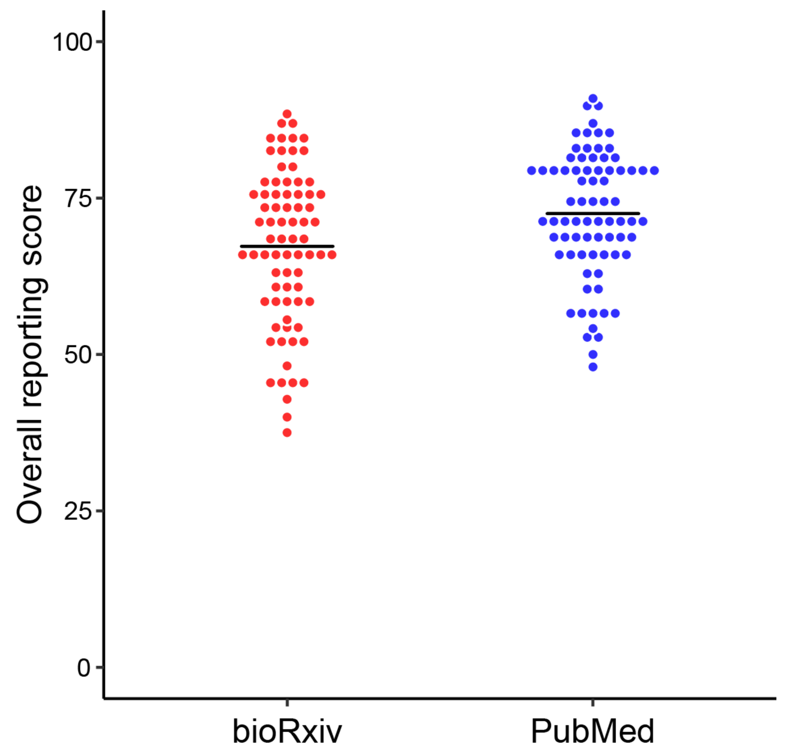

Figure 2 – Comparison of overall reporting scores between preprints and peer-reviewed articles. PubMed article's score was slightly higher than bioRxiv's (72.5±10.1 vs 67.3±12.4, respectively) and this comparison yielded p=0.005 (Student's t test, t=-2.84). 95% confidence interval for the absolute difference of 5.2 is [1.6 – 8.9]. N=76/group was determined to achieve 90% power to detect a difference of 10% between groups at α=0.05.

Score	Subset	Mean ± S.D. (bioRxiv)	Mean ± S.D. (PubMed)	t value	p value	Sample Size
Overall	*In vitro*	61.7 ± 10.1	66.1 ± 7.8	-1.46	0.15	18
	Invertebrates	77.3	80.0	-	-	1
	Vertebrates	66.9 ± 11.5	72.4 ± 9.6	-1.82	0.07	25
	Humans	70.4 ± 13.6	76.0 ± 10.3	-1.86	0.07	32
General	All	71.3 ± 12.2	74.0 ± 10.1	-1.48	0.14	76
	In vitro	65.3 ± 11.9	67.8 ± 9.3	-0.71	0.48	18
	Invertebrates	77.8	84.2	-	-	1
	Vertebrates	74.1 ± 11.2	74.7 ± 10.0	-0.21	0.83	25
	Humans	72.4 ± 12.3	76.7 ± 9.4	-1.55	0.13	32
Specific	*In vitro*	44.9 ± 29.0	56.5 ± 21.7	-1.35	0.18	18
	Invertebrates	75.0	66.7	-	-	1
	Vertebrates	52.7 ± 19.7	67.5 ± 16.1	-2.91	0.005	25
	Humans	65.2 ± 26.8	74.8 ± 16.8	-1.73	0.09	32

Table 1 – Comparison of scores between groups by category of the article and sections of the questionnaire. The difference favoring articles from PubMed remained consistent; however, considering Sidak's correction for multiple comparisons (α=0.003), we are not confident about asserting such effects.

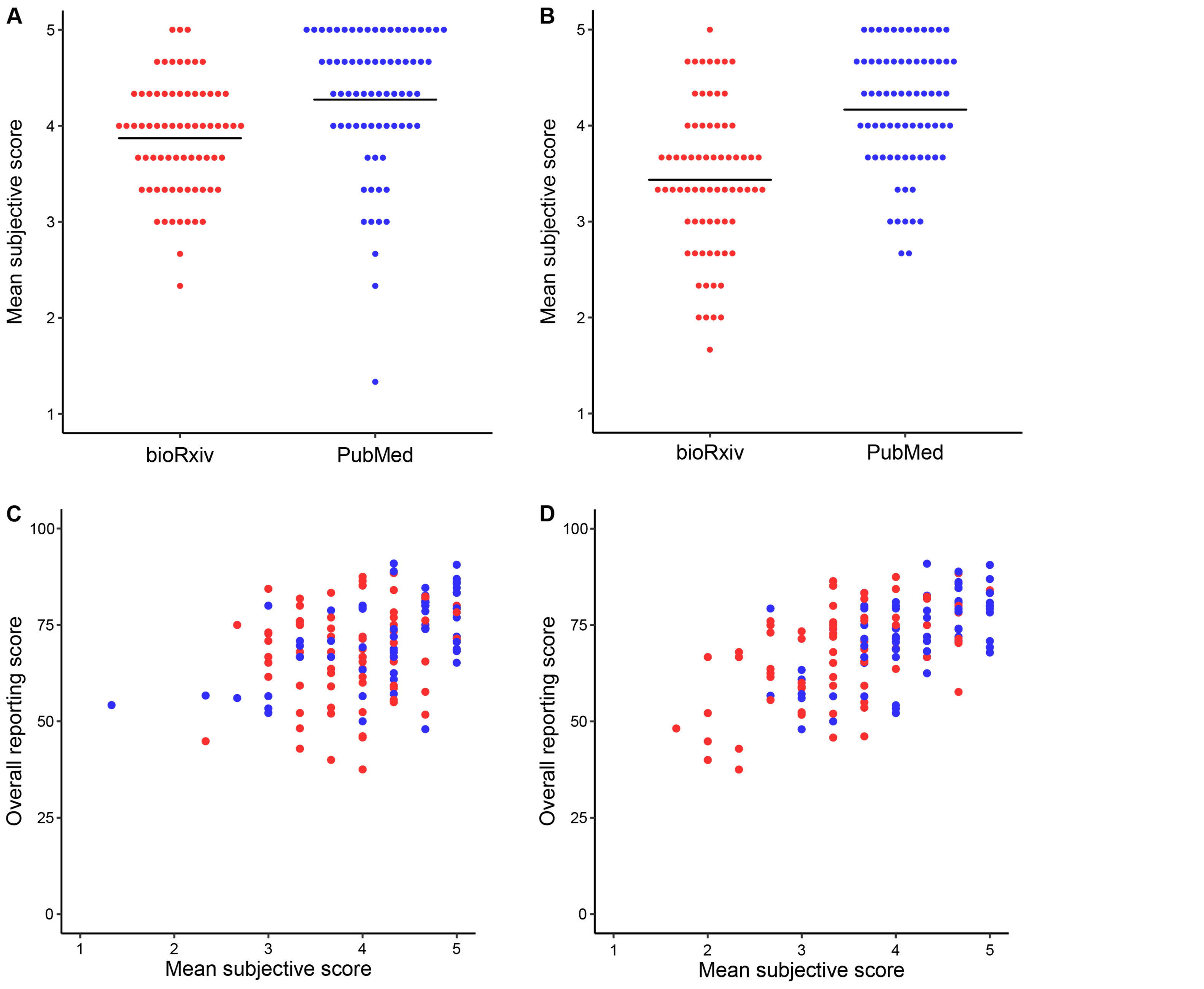

Figure 3 – Subjective assessment. (A) Scores were given as an answer to "Do the title and abstract provide a clear idea of the article's main findings?". Student's t test, t=-3.61, $p=4\times10^{-4}$. (B) Scores were given as an answer to "Was the required information easy to find and extract from the article?". Student's t test, t=-6.22, $p=5\times10^{-9}$. (C) Title and abstract clarity vs. reporting scores. r=0.39, $p=1.6\times10^{-6}$ (all articles); r=0.14, p=0.22 (bioRxiv); r=0.55, $p=5.1\times10^{-7}$ (PubMed). (D) Easiness to extract information vs. reporting scores. r=0.59, $p=9.7\times10^{-15}$ (all articles); r=0.54, $p=8\times10^{-7}$ (bioRxiv); r=0.59, $p=5.6\times10^{-8}$ (PubMed). N=72/group.

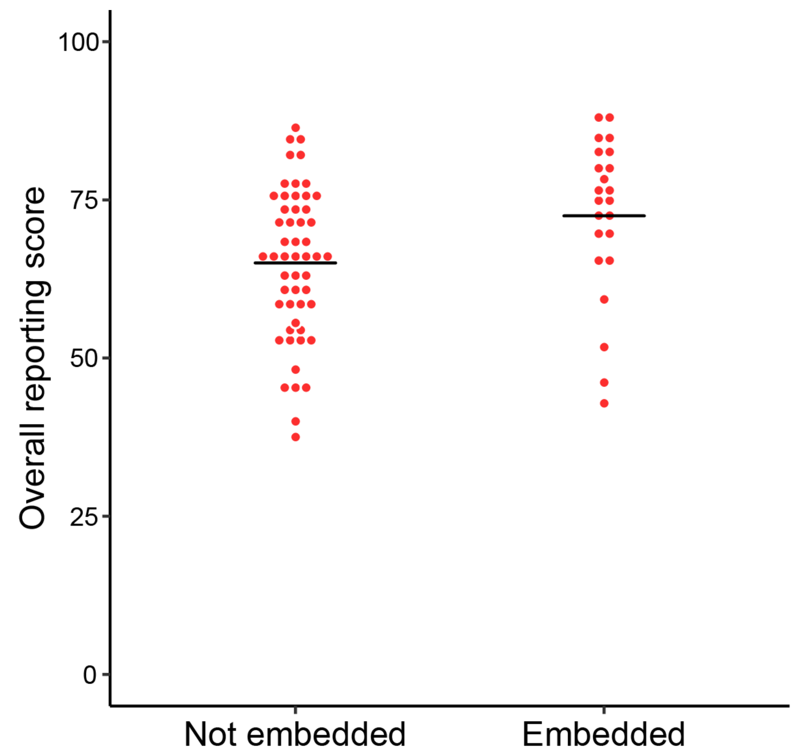

Figure 4 – Comparison between preprints with and without figures embedded in text. As an exploratory analysis, we obtained the overall reporting score for articles found in bioRxiv with figures presented throughout the manuscript text and those that presented figures by the end of the article or in separate files. We found that the mean score for the embedded group (72.5±12.6) is comparable to PubMed score (all articles had embedded figures). Meanwhile, mean score for the non-embedded group (65.0±11.8) is equivalent to the whole bioRxiv score. Comparison between groups (Student's t test, t=-2.48, p=0.01) is not significant considering Sidak's correction for multiple comparisons.

Planned secondary analyses also showed that neither impact factor (for the PubMed sample) nor region of origin had any impact on reporting scores (Spearman's correlation, r=-0.24, p=0.07; two-way ANOVA, p_{region}=0.17, p_{group}=0.007, $p_{interaction}$=0.07). We did find a negative correlation between score and number of figures in the main text (all articles, ρ=-0.30, p=0.0001; bioRxiv, ρ=-0.34, p=0.003; PubMed, ρ=-0.22, p=0.05; n=76/group).

From these results, we conclude that the average article found in bioRxiv is not of lesser quality from one found in PubMed in terms of reporting quality. Our results suggest that formatting of the article seems to play an important role in reporting assessment as we designed it. Besides, as this is an observational study of random samples of reviewed and unreviewed articles, we cannot directly attribute peer-review as the cause for the differences observed. Another limitation is that the subject areas are differently represented in both groups, so differences observed could stem from the fact that articles from both groups are different even before peer review occurs.

This work is available as a preprint at http://dx.doi.org/10.1101/581892. Currently, we are at the final stages of collecting data to directly compare the first preprint version to the peer-reviewed published version of the same article.

[a]Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
[b]Seção de Arbovirologia e Febres Hemorrágicas, Instituto Evandro Chagas, Pará, Brazil.
[c] Departamento de Morfologia e Ciências Fisiológicas, Universidade do Estado do Pará, Pará, Brazil.
[d]Centro Universitário Metropolitano da Amazônia, Instituto Euro-Americano de Educação, Ciência e Tecnologia, Pará, Brazil.
[e]Department of Neuroscience, Section Medical Physiology, University of Groningen, Groningen, The Netherlands.
[f]Biomedical Sciences Institute, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
[g]University of Edinburgh Medical School, Scotland, United Kingdom.
[h]Programa de Pós-Graduação em Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil.
[i]Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA.
[j] Departament of Neuroscience, Uppsala University, Uppsala, Sweden.
[k]Institute of Microbiology of the Czech Academy of Sciences, Czech Republic.
[l]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA.
[m]Carl R Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.
[n]Seção de Virologia, Instituto Evandro Chagas, Pará, Brazil.
[o]Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, Brazil.

Funding
CNPq
FAPERJ