Original Article

Comparison of the patterns of duplicate articles between KoreaMed and PubMed journals published from 2004 to 2009 according to the categories of duplicate publications

Soo Young Kim[1]*, Chong Woo Bae[2]*, Hye-Min Cho[3], Sun Huh[4]

[1]Department of Family Medicine, Gangdong Sacred Heart Hospital, College of Medicine, Hallym University, Seoul; [2]Department of Pediatrics, Kyung Hee University Hospital at Gangdong, Kyung Hee University School of Medicine, Seoul; [3]Infolumi Co., Seongnam; [4]Department of Parasitology and Institute of Medical Education, College of Medicine, Hallym University, Chuncheon, Korea

Abstract

This study compared the patterns of duplicate articles between KoreaMed and PubMed journals based on a division of duplicate publications into the 4 categories of 'copy,' 'salami' (fragmentation), 'imalas' (disaggregation), and 'others,' as well as in terms of the 11 subcategories suggested by Bae et al., which further elaborate on those 4 main categories. We hypothesized that these 2 groups of articles would show different patterns of duplication. Duplicate publications were identified in a random sample of 5% of the articles from the KoreaMed database published between 2004 and 2009, while all articles with the publication type of 'duplicate publication' were selected from PubMed over the same period. The selected articles were classified based on the 4 categories and 11 subcategories of duplicate publications, and the data from the 2 groups were compared. A total of 108 articles were selected from KoreaMed and 45 articles were obtained from PubMed. The category of copy was the most common in both databases. The next most frequent pattern was imalas (disaggregation). Pattern of duplicate publication between 2 databases showed no correlation (P = 0.8754). Although the 108 articles from KoreaMed were allocated to all 11 Bae et al.'s subcategories, those from PubMed were allocated to only 8. The above results showed that the articles in the 2 databases had different patterns of duplication, as defined in terms of the 11 subcategories. The use of these 11 subcategories will help journal editors to develop an appropriate framework for considering a variety of duplication types.

Keywords

Factual databases; Korea; Medical writing; Publications; PubMed

*These two authors contributed equally to this study as the first authors.

Introduction

Out of 114 randomly selected retracted articles from KoreaMed (https://koreamed.org), a database containing abstracts of the medical literature from Korea published from 1999 to 2016, the most common reason for retraction was duplicate publication (66 cases, 57.9%) [1]. The duplicate rate in medical journals published in Korea was relatively high: 5.9% in 2004, 6.0% in 2005, and 7.2% in 2006. However, it decreased to 1.2 % in 2009. Of all duplicated articles, 53.4% were classified as 'copies,' 27.8% as 'salami' (fragmentation), and 18.8% as 'imalas' (disaggregation) [2]. Duplicate publication was the cause of 149 retractions (18.1%) of the 821 retracted articles in PubMed published between 2008 and 2012 [3]. Although duplicate publication in the medical field itself is not harmful to medical practice or patient safety, it may weaken the validi-

ty of meta-analyses [4]. An increase was observed in the mean effect size and fail-safe number with duplicated data when duplicate articles were included in meta-analyses, despite the presence of only 6 duplicate publications out of the 1,194 articles that were used in meta-analyses by Korean authors [5].

To define and analyze the phenomenon of duplicate publications, a classification of duplicate publications is necessary. von Elm et al. [6] found 6 duplication patterns after comparing the study samples and outcomes of duplicates and the corresponding main articles from 141 systematic reviews on anesthesia or analgesia as follows: identical samples and identical outcomes; identical samples and identical outcomes, but several duplicates assembled; identical samples and different outcomes; increased sample and identical outcomes; decreased sample and identical outcomes; and different samples and different outcomes. In 2011, Bae et al. [7] analyzed the patterns of

Table 1. Classification of duplicate publications based on papers sampled from KoreaMed from 2004 to 2008 [7]

Category	Subcategory	Explanation
1 Copy		
1-1	Complete copy with different language	This occurred when the same content was submitted to at least 2 different journals and published without permission from both journal editors. If permission was received from both journal editors, it could be published as a secondary publication.
1-2	Complete copy with the same language	This occurred when the literature database did not include one of the journals. The authors incorrectly believed that it was not possible to trace the duplicate publication or they were not aware of the concept of a duplicate publication.
1-3	Copy with some modifications with different language	Same as 1-1 above, except adding some data and revising the Discussion section. There was no difference in the conclusions.
1-4	Copy with some modifications with the same language	Same as 1-2 above, except adding some data and revising the Discussion section. There was no difference in the conclusions.
2 Salami publication		
2-1	Salami publication with a divided sample number	This occurred when part of the data was used for one article and the other part or the whole dataset was used for the other article. The results may be the same or different based on the hypothesis.
2-2	Salami publication with a divided outcome	This occurred when the hypothesis or methods were changed after the publication of one article; therefore, the results may be the same or different.
3 Imalas publication		
3-1	Imalas publication with an extended sample number or extended study periods	This occurred when the number of subjects was increased or the observation period was prolonged.
3-2	Imalas publication with an added hypothesis	This occurred when a hypothesis was added.
3-3	Imalas publication with an extended sample number or extended study periods, and an added hypothesis	This occurred when the number of subjects was increased or the observation period was prolonged, and a hypothesis was added.
4 Others		
4-1	Reverse imalas	This occurred when the number of subjects was reduced.
4-2	Not classified above	Others not classified as above or difficult to classify.

100 pairs of duplicate publications in the KoreaMed database and some other articles that were written by Korean authors and submitted to international journals. They proposed a new classification system of duplicate publications based on the 6 criteria suggested by Mojon-Azzi et al. [8] of having a similar hypothesis, similar numbers or sample sizes, identical or nearly identical methodology, similar results, at least 1 author in common, and no or little new information made available. However, the interpretation of "similar numbers or sample sizes" was extended from the original formulation of "90% or more of the studied materials, animals, or subjects are identical" to include the duplication of a significant number of materials, animals, or human subjects. Furthermore, the possibility of secondary publication was checked in the analyzed articles. Finally, a classification of duplicate publication with 11 subcategories was suggested, as shown in Table 1 [7]. This system enabled the comprehensive classification of a variety of patterns of duplicate publications observed in KoreaMed [7]. This system was developed based on an analysis of articles in the KoreaMed database, the contents of which are mostly from Korea. Thus, we investigated how this system would apply to PubMed (https://pubmed.gov) articles.

Therefore, this study compared the patterns of duplicate publications between KoreaMed and PubMed journals based on the new classification system of duplicate publications proposed by Bae et al. [7]. We hypothesized that the 2 groups of articles would show different patterns of duplication.

Methods

Study design
This was a retrospective analysis of 2 literature databases: KoreaMed and PubMed.

Materials
Duplicate publications were identified in a random sample of 5% of the articles from the KoreaMed database published between 2004 and 2009, while all articles with the publication type of 'duplicate publication' from PubMed over the same period were selected. It is difficult to find the publication type of 'duplicate publication' from KoreaMed because there was no input of the publication type in KoreaMed; therefore, the analysis was done from randomized samples. Meanwhile, the publication type of 'duplicate publication' was already recorded in the PubMed.

Analysis
The selected articles were classified based on the category of duplicate publication, as shown in Table 2 [7], and the data from the 2 databases were compared. Classification judg-

Table 2. Comparison of patterns of duplicate publications between papers sampled from KoreaMed and articles with the publication type of 'duplicate publication' from PubMed from 2004 to 2009, based on the 11 Bae et al. [7]'s subcategories of duplicate publications

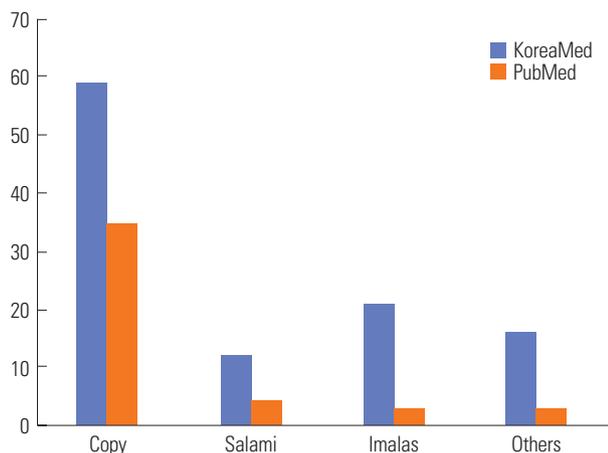| Category		Subcategory	No. of corresponding articles		
			KoreaMed	PubMed	Total
A	Copy	Complete copy with a different language	25	3	28
B		Complete copy with the same language	2	16	18
C		Copy with some modifications with a different language	26	1	27
D		Copy with some modifications with the same language	6	15	21
		Subtotal	59	35	94
E	Salami	Salami publication with a divided sample number	2	1	3
F		Salami publication with divided outcomes	10	3	13
		Subtotal	12	4	16
G	Imalas	Imalas publication with an expanded sample number or extended study period	17	3	19
H		Imalas publication with an added hypothesis	1	0	1
I		Imalas publication with an expanded sample number or extended study period, and an added hypothesis	3	0	3
		Subtotal	21	3	24
J	Other	Reverse imalas	7	3	10
K		Not classified as above	9	0	9
		Subtotal	16	3	19
		Total	108	45	153

Fig. 1. Comparison of the patterns of duplicate publications between papers sampled from KoreaMed and articles with the publication type of 'duplicate publication' from PubMed from 2004 to 2009, based on the 4 categories of duplicate publications.

ments were made by 2 pairs of authors: SYK and HMC, CWB and SH. One pair checked half of the articles from each database. If both authors in the pair agreed, an article was included in a given category. The classification was performed on February 5, 2017, after reading and discussing the relevant articles. The concordance correlation was tested to establish correlations between duplicate articles from the KoreaMed and PubMed according to subcategories. For statistical analysis, DBSTAT ver. 5.0 (DBSTAT, Chuncheon, Korea) was used; this program is available from http://dbstat.com/.

Results

A total of 108 articles were selected from KoreaMed, while 45 articles were obtained from PubMed. The results are presented in Table 2 [7] and Fig. 1. The category of 'copy' was the dominant pattern in both databases. Of the 94 copies, the predominant subcategories were 'complete copy with a different language' (28) and 'copy with some modifications with a different language' (27). The next most frequent pattern was 'imalas.' Of the 24 papers in this category, 'imalas publication with an expanded sample number or extended study period' was the predominant subcategory (19). Of the 16 'salami' papers, the subcategory of 'salami publication with divided outcomes' (13) was the most prevalent. There was no concordance correlation between the 2 databases according to the 11 Bae et al. [7]'s subcategories (P = 0.8754).

Discussion

The above results show that our hypothesis that the patterns of duplication would differ between the 2 groups of articles was accepted. There was no concordance correlation between the 2 databases according to the 11 Bae et al. [7]'s subcategories. The identification of articles belonging to more categories in the KoreaMed database may reflect the presence of more cases, as well as the smaller number of articles from PubMed that were included. Among the duplicate publications from PubMed, it was difficult to detect duplicate publications belonging to the categories of 'imalas publications with an added hypothesis' and 'imalas publications with an expanded sample number or extended study period, and an added hypothesis.' This difficulty may be a limitation due to the number of articles sampled from PubMed. If editors are appropriately vigilant in detecting imalas publications, more cases may be detected. The above results will help journal editors develop an appropriate framework for considering a variety of duplication types.

The primary limitation of this study is the small number of duplicate articles from PubMed due to the short period of publication. In this study, publication period was identical in 2 databases. If the period were to be extended, more duplicate articles would have been included in the categorization. Although 2 authors in a pair discussed and reached an agreement regarding the classification of cases of duplication, there may have been the possibility for some bias. These frameworks were applied to medical journals, so a similar analysis for the fields of agriculture, engineering, the natural sciences, the social sciences, and the arts and humanities should be done, after appropriate adaptation, to confirm the general feasibility of this approach.

In conclusion, the new Bae et al. [7]'s classification of duplicate publications, containing 11 subcategories, can be used not only for medical journals from Korea, but also for journals in PubMed. A different pattern was found in the subcategories of duplicate publications between KoreaMed and PubMed. We recommend that scholarly journal editors and librarians adopt the Bae et al. [7]'s classification of duplicate publications in order to categorize duplicate publications more precisely. More work on categorization will confirm the feasibility of this classification system.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Huh S, Kim SY, Cho HM. Characteristics of retractions from Korean medical Journals in the KoreaMed database:

a bibliometric analysis. PLoS One 2016;11:e0163588. https://doi.org/10.1371/journal.pone.0163588

2. Kim SY, Bae CW, Hahm CK, Cho HM. Duplicate publication rate decline in Korean medical journals. J Korean Med Sci 2014;29:172-5. https://doi.org/10.3346/jkms.2014.29.2.172

3. Amos KA. The ethics of scholarly publishing: exploring differences in plagiarism and duplicate publication across nations. J Med Libr Assoc 2014;102:87-91. https://doi.org/10.3163/1536-5050.102.2.005

4. Fairfield CJ, Harrison EM, Wigmore SJ. Duplicate publication bias weakens the validity of meta-analysis of immunosuppression after transplantation. World J Gastroenterol 2017;23:7198-200. https://doi.org/10.3748/wjg.v23.i39.7198

5. Choi WS, Song SW, Ock SM, et al. Duplicate publication of articles used in meta-analysis in Korea. Springerplus 2014;3:182. https://doi.org/10.1186/2193-1801-3-182

6. von Elm E, Poglia G, Walder B, Tramer MR. Different patterns of duplicate publication: an analysis of articles used in systematic reviews. JAMA 2004;291:974-80. https://doi.org/10.1001/jama.291.8.974

7. Bae CW, Kim SY, Huh S, Hahm CK. Sample cases of duplicate publication. Seoul: Korean Association of Medical Journal Editors; Chuncheon: Xmlarchive; 2011.

8. Mojon-Azzi SM, Jiang X, Wagner U, Mojon DS. Redundant publications in scientific ophthalmologic journals: the tip of the iceberg? Ophthalmology 2004;111:863-6. https://doi.org/10.1016/j.ophtha.2003.09.029