Plagiarism detection in manuscripts submitted to the Journal of Surgical Sciences between 2020 and 2021: a case study

Florentina Mușat¹,², Dan Nicolae Păduraru¹,², Alexandra Bolocan¹,², Daniel Ion¹,², Alexandru Constantinescu¹,², Octavian Andronic¹,²

¹Carol Davila University of Medicine and Pharmacy, Bucharest; ²University Emergency Hospital of Bucharest, Bucharest, Romania

Abstract
The aim of this study was to share our experience with plagiarism detection in manuscripts submitted to the Journal of Surgical Sciences, a Romania-based medical journal, between 2020 and 2021. We analyzed similarity score reports from 200 articles submitted consecutively for publication between 2020 and 2021 generated by PlagScan, a software tool for plagiarism detection. The similarity score ranged from 0% to 92.4%, and 45 articles presented scores over 25.0%. According to PlagScan’s results, more than half of the submitted articles had a similarity score of more than 10% and one-third of them had a similarity score above 20%. Among submitted manuscripts with a similarity score of less than 20%, a larger proportion of the original research and review manuscripts than case reports used more than 10 sources. All articles with a similarity score below 20% were evaluated qualitatively before the final decision of rejection.

Keywords
Plagiarism; Detection; Software; Articles; Journal

Introduction

Background and rationale
Plagiarism is the act of presenting someone else’s work, ideas, or words as one’s own without giving proper credit or citation. Plagiarism is one of the most common violations of academic writing ethical principles, and it may lead to the severe sanctions to both authors and journals. Avoiding plagiarism requires a combination of careful rewriting, effective paraphrasing, diligent referencing, and meticulous editing [1]. This is a detailed and time-consuming process that demands the author’s attention to ethical writing practices and academic integrity. Several useful, well-known plagiarism detection software programs are available, including Turnitin (Turnitin LLC), iThenticate (Turnitin LLC), PlagScan (Turnitin LLC), Grammarly (Grammarly...
Inc), and Copyscape (Indigo Stream Technologies Ltd) (Suppl. 1). These tools can help prevent or identify plagiarism before publication. In the initial editorial screening of submitted articles, the similarity score is used as the first rejection criterion, but in our experience, some instances where the similarity score exceeds the accepted threshold are unintentional.

This study presents a comprehensive analysis of the plagiarism instances documented within a medical journal. We want to provide valuable insights into the prevalence of plagiarism in the context of the journal, shedding light on the extent of the issue and its potential implications. Our journal, *Journal of Surgical Sciences* (ISSN: 2360-3038, eISSN: 2457-5364, https://journalofsurgicalsciences.com/) is a double-blind peer review, open access, and Romania-based medical journal that publishes articles in the field of surgical specialties. We utilize the plagiarism detection software PlagScan to screen for duplication or plagiarism before the peer review process.

**Objectives**
The aim of this study was to analyze the records of plagiarism detection for submissions to the *Journal of Surgical Sciences* between 2020 and 2021.

**Methods**

**Ethics statement**
This study is based on submitted manuscript data; therefore, neither Institutional Review Board approval nor obtaining informed consent was required.

**Study design**
This was a case study on the editing process for a medical journal.

**Setting, data sources, and measurement**
We quantitatively evaluated the plagiarism reports from 200 articles submitted for publication between 2020 and 2021 to the *Journal of Surgical Sciences*. The reports were generated using PlagScan (Suppl. 1). The text uploaded for plagiarism analysis included the abstract, the entire content of the manuscript, and the figure and table legends. The reference list of the manuscript was not uploaded for plagiarism detection. The steps for plagiarism detection were as follows.

1. Upload the document: the user uploads the document to the PlagScan platform, either by uploading the file directly or by copying and pasting the text into the system.
2. Analysis of the document: the PlagScan software analyzes the document and creates a digital fingerprint of the text, which is used to compare it against other sources.
3. Comparison to database: the software then compares the document to a database of sources, which includes academic databases, internet sources, and other published works.
4. Report generation: PlagScan generates a detailed report that highlights any potential instances of plagiarism and provides information about the sources that were detected.
5. Review of the report: the user reviews the report and makes any necessary changes or corrections to the document to ensure that proper citation and attribution are provided.
6. Based on the collected data and on personal editorial experience, some observations were made regarding plagiarism.

**Statistical methods**
Descriptive statistics were utilized to analyze the results of plagiarism detection.

**Results**
In total, 200 manuscripts submitted between 2020 and 2021 to the *Journal of Surgical Sciences* were analyzed. Sixty-six (33.0%) were case reports, 91 (45.5%) were original articles, and 43 (21.5%) were literature reviews (Table 1). The similarity score had values between 0% and 92.4%, and 45 articles had scores over 25.0% (Table 2). The chi-square
A quantitative assessment of plagiarism

Test of independence was performed to examine the relationship between the similarity score and the type of manuscript. The chi-square statistic was 5.875, and the P-value was 0.208. The result was not significant at P < 0.05, indicating that the level of plagiarism was not dependent on manuscript type. Many of the articles contained similar fragments due to the misuse of the rules of quoting, citing and academic writing. The excessive repetition of expressions without the use of abbreviations or the adoption of clearly grounded definitions or concepts led to relatively high similarity scores without the intention of plagiarism. Although review articles showed similarity scores similar to those of original articles or case reports, the highest similarity scores (over 70%) were observed in case reports or original articles, not in literature reviews.

According to our PlagScan evaluations, more than half of the submitted articles had a similarity score of more than 10% and one-third of them had a similarity score above 20% (Fig. 1). There were no major differences in the proportion of manuscripts with a low similarity score (less than 10% or less than 20%) between the different types of research. The proportions of manuscripts with similarity scores of less than 20% were 70%, 73%, and 74% for case reports, reviews, and original research manuscripts, respectively (Fig. 2).

Among submitted manuscripts with a similarity score of less than 20%, a larger proportion of the original research and review manuscripts used more than 10 sources, compared to case report manuscripts (Fig. 3). In 69.5% of all submitted manuscripts, PlagScan identified more than 10 sources of similarity. The majority of manuscripts with high similarity scores (more than 20%) used more than 10 sources (93.0%).

The similarity score criterion used by our journal for accepting articles is less than 10%. Manuscripts with a similarity score between 10% and 50% are sent back to the author for correction of this issue. In our experience, articles with similarity scores of more than 30% usually do not reach the acceptable limit after revision.

Table 2. Classification of manuscripts based on the degree of similarity

<table>
<thead>
<tr>
<th>Similarity score</th>
<th>Total</th>
<th>Case report</th>
<th>Original research</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of manuscripts</td>
<td>Degree of freedom</td>
<td>P-value</td>
<td>No. of manuscripts</td>
</tr>
<tr>
<td>Minor (11%–24%)</td>
<td>60</td>
<td>14</td>
<td>18.86</td>
<td>1.25</td>
</tr>
<tr>
<td>Moderate (25%–50%)</td>
<td>32</td>
<td>12</td>
<td>10.06</td>
<td>0.38</td>
</tr>
<tr>
<td>Major (&gt;50%)</td>
<td>13</td>
<td>7</td>
<td>4.09</td>
<td>2.08</td>
</tr>
</tbody>
</table>

Fig. 1. Similarity score for each individual submitted manuscript.

Fig. 2. Proportion of manuscripts with low similarity scores.

Fig. 3. Correlation between similarity score and the number of sources identified by PlagScan (Turnitin LLC).
Discussion

Interpretation

The results of our study indicate the presence of considerably high similarity scores in the analyzed manuscripts, implying that authors may not have a comprehensive understanding of how to properly attribute and cite sources in their work. This study suggests some possible reasons behind the occurrence of plagiarism, such as the misuse of quoting, citing, and academic writing rules. The excessive repetition of expressions without appropriate abbreviations or clearly grounded definitions or concepts can inadvertently contribute to a relatively high similarity score without the intention of plagiarism. Furthermore, the observation that higher levels of similarity were more frequently observed in case reports and original articles than in literature reviews suggests that authors of these manuscript types may face particular challenges in maintaining originality and properly citing sources. This could be due to the nature of these manuscript types, where authors may struggle to strike a balance between presenting novel findings and providing appropriate references.

Some people might speculate that case reports may have lower similarity scores than other types of manuscripts, as they typically describe unique or rare cases and do not involve a significant amount of background research or literature review. Review articles and meta-analyses tend to summarize and synthesize existing literature from multiple sources, which can make it more challenging to properly attribute and cite all sources. Original studies, such as randomized controlled trials and observational studies, are designed to generate new data and findings and may be less likely to involve plagiarism, as the research is based on original data and analysis. However, it is important to note that plagiarism can still occur in original studies if the authors fail to properly cite and attribute sources or if they recycle previously published work without appropriate citation.

There may be a correlation between the number of sources found by PlagScan and the similarity score, but it is difficult to make a general statement as this relationship can depend on various factors [2]. In our opinion, a manuscript with a higher number of sources may be more prone to plagiarism, as it can be challenging to properly attribute and cite all sources, especially if the authors are dealing with a large amount of information from multiple sources. This can increase the risk of accidental or unintentional plagiarism and lead to a higher similarity score. However, a manuscript with a higher number of sources may also be more likely to have gone through rigorous review and editing processes, which can help to identify and correct any potential instances of plagiarism. Additionally, if the authors are diligent in properly citing and attributing all sources, the similarity score may be low, even if the manuscript has a high number of sources. Ultimately, the similarity score depends on the specific content of the manuscript and the accuracy of the citation and attribution of sources, rather than only the number of sources identified by PlagScan.

Reducing the similarity score of a manuscript can be challenging if it is initially over 30%, for several reasons. A high similarity score often indicates that a significant portion of the manuscript's content has been directly copied from other sources without proper citation or attribution. To reduce the similarity score, authors need to extensively rewrite the relevant sections to ensure that the content is original and properly cited. This process can be time-consuming and require significant effort [3]. Authors may need to identify the original sources from which the plagiarized content was taken to properly cite and attribute the information. This task can be complex and require thorough research to trace back and find the original sources. If the plagiarized content consists of verbatim or slightly modified text from other sources, authors need to engage in effective paraphrasing and rephrasing to express the ideas in their own words while maintaining the meaning and integrity of the information. This requires skill in rewriting and avoiding plagiarism [4]. Reducing the similarity score should not lead to omitting necessary references or citations. Authors need to strike a balance between expressing their ideas in original language and properly acknowledging the sources that contributed to their work [5]. This can be challenging, as authors must ensure that all relevant sources are appropriately cited while maintaining the flow and coherence of their own writing. After revising to reduce plagiarism, authors must conduct thorough reviews and edits to verify the accuracy of citations, ensure proper attribution, and check for any remaining instances of unintentional plagiarism. This iterative process may need to be repeated multiple times to achieve a satisfactory reduction in the similarity score.

Comparison with previous studies

Some journals have reported their experience in plagiarism detection. One study that presented the authors’ experience with plagiarism regarding articles submitted to American Journal of Roentgenology found that out of 110 manuscripts, the initial overall similarity index ranged from 7% to 46% [2]. Those values are much lower than those found in our study.

In another study [6], analyzing a total of 400 consecutively submitted manuscripts (357 original research articles and 43 review articles) to Genetics in Medicine, it was found that 17% of the submissions contained levels of plagiarized material considered unacceptable. The unacceptable plagiarism level referred to that study to a median score between 17% and 32%, with a minimum of 9% and a maximum of 53%. Notably,
A quantitative assessment of plagiarism

82% of these plagiarized manuscripts originated from countries where English was not designated as an official language. In our study, the median similarity score was 11.75%, with a minimum value of 0.0% and a maximum value of 92.40%.

Baždarić et al. [7] assessed the prevalence of plagiarism in manuscripts submitted for publication in the Croatian Medical Journal. Out of a total of 754 submitted manuscripts, the software flagged 105 (14%) as being potentially associated with plagiarism. Upon manual verification, it was determined that 85 manuscripts (11%) were indeed plagiarized. Specifically, 63 (8%) were classified as instances of true plagiarism, while 22 (3%) were categorized as cases of self-plagiarism. The extent of plagiarism was minor (11%–24%) in 31 out of 85 manuscripts. These results are similar to those from the present study.

Limitations and generalizability
All similar texts are not plagiarized ones. Even when references are added to the target text, PlagScan may show similarities. Therefore, the similarity score should not be interpreted as the true percentage of plagiarism. A line-by-line examination of similar text should be done to conclude whether plagiarism has taken place. Furthermore, this study reports the experience of one journal; therefore, it is difficult to generalize the results to other journals in the world.

Conclusion
All articles with a similarity score below 20% should be evaluated qualitatively before a final decision of rejection. More manuscripts had a minor level of similarity, with scores ranging from 11% to 24%, than had moderate or major levels of similarity. Overall, these findings highlight the presence of plagiarism in the analyzed manuscripts, emphasizing the need for improved adherence to academic writing guidelines and stricter enforcement of ethical standards in research publications. The results of this study suggest that authors may benefit from additional guidance on academic writing practices to avoid unintentional instances of plagiarism.

Conflict of Interest
The authors, as editors of the Journal of Surgical Sciences, pay a fee for using PlagScan. Octavian Andronic also serves as an editor of Science Editing since 2023, but has no role in the decision to publish this article. No other potential conflict of interest relevant to this article was reported.

Funding
The authors received no financial support for this article.

Reference availability
Dataset file is available from the Harvard Dataverse at https://doi.org/10.7910/DVN/CE7JPH.

Dataset 1. Raw data of the similarity score for each submission to the Journal of Surgical Sciences.

Supplementary Materials
Supplementary materials are available from https://doi.org/10.7910/DVN/CE7JPH.

Suppl. 1. Strengths and limitations of five popular plagiarism detection software programs.

References