



A Practical Guide for Plagiarism Detection Through the Coordinated Use of Software and (Human) Hardware

Durga Prasanna Misra¹

¹Department of Clinical Immunology and Rheumatology, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, Uttar Pradesh, India

J Gastrointest Infect 2022;12:47–50.

Address for correspondence Durga Prasanna Misra, MD, DM, MSc (Epidemiology), FRCP (Edin), Department of Clinical Immunology and Rheumatology, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow 226014, Uttar Pradesh, India (e-mail: durgapmisra@gmail.com; dpmisra@sgpgi.ac.in).

Abstract

Duplication of content, whether it be text, figures, tables, or ideas, without duly referencing the original source of such content is called as plagiarism. Plagiarism of text can be flagged by similarity checking software. Careful curating of content before it is put into similarity checking software for screening as well as a thorough understanding of how such software works are essential to appropriately interpret such similarity check reports. Mere similarity is not plagiarism. Drawing an inference regarding plagiarism based on the output of similarity checking software requires considerable human input from editors and reviewers. Identification of plagiarism of figures, tables, and ideas almost entirely depends on the efforts of editors or reviewers rather than being detectable automatically. A thorough understanding of plagiarism is essential for prospective authors to safeguard against this academic sin.

Keywords

- ▶ iThenticate
- ▶ plagiarism
- ▶ publication ethics
- ▶ similarity

Introduction

Scientific reporting relies on the generation of original observations of natural phenomena as well as novel inferences drawn from experimental scenarios. The content as well as the language used to describe these needs to be original. Any duplication (whether intentional or not) of scientific ideas, hypotheses, content, or language is not recommended.^{1–3} Whenever such duplication is essential, it requires a clear reference to be made to the original source. Plagiarism refers to the duplication of ideas, hypotheses, content, or language without attribution to the due source (▶ **Table 1**). In its truest sense, plagiarism is an intentional act rather than accidental. Plagiarism is a moral or an ethical construct without statutes of limitation to the original source, whereas copyright is a legal construct governed by

local regulations.^{1–3} Plagiarism is distinct from similarity (which is detected by similarity checking software).^{4,5} Plagiarism has been identified for over two millennia.⁶ However, recent developments such as the advent of the World Wide Web as well as proliferation of technology along with increasing use of artificial intelligence has increased the ability to identify plagiarism over the years. ▶ **Fig. 1** reflects the results of a search on PubMed (conducted on February 27, 2022) using the search string “plagiarism.” Noticeably, the number of articles retrieved with this search considerably increased over the past three decades. With the increasing number of instances of plagiarism, the complexity of output from similarity checking software used to screen for plagiarism has also noticeably increased. Plagiarism identified during the peer review process shall invariably result in

received

February 27, 2022

first decision

March 18, 2022

accepted after revision

April 16, 2022

DOI <https://doi.org/>

10.1055/s-0042-1757486.

ISSN 2277-5862.

© 2022. Gastrointestinal Infection Society of India. All rights reserved.

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

Table 1 Key types of plagiarism and their detection

Type of plagiarism	Aids to detection
Text	Similarity checking software
Figures	Google images or other similar image repositories
Tables	Predominantly related to review articles. Manual review of other similar published articles and checking the order of references in such reviews might help
Ideas	The purported originator of the idea should provide evidence of the primacy of their idea, possibly through a prior publication

rejection. Plagiarism detected after publication of a manuscript is likely to result in retraction of the plagiarized content or at least require a considerable erratum to be published.⁷

In this article, we shall aim to develop a practical approach toward detecting plagiarism of text, graphics or tables, or ideas. This should help reviewers and editors understand the nuances of detecting plagiarism as well as provide an insight for authors to enable them avoid falling

foul of plagiarism. While the authors have published about plagiarism before, every attempt has been made to avoid overlap of content with their previously published work.^{2,5,8,9}

Plagiarism of Text

Plagiarism of text is what is most easily identifiable based on the outputs from similarity checking software.^{4,5} The most common software used in academic publishing for checking similarity is iThenticate, a product of Turnitin, LLC. What this paid software does is continually scan troves of online pages and record their content (including content from journals linked to its database via Crossref). Content available on the Internet in the past (which might have been deleted thereafter) is also recorded by iThenticate and available for comparison. Thus, plagiarized content from source material not currently available online can also be detected by iThenticate.⁵ Other paid software commonly used for similarity checking include Grammarly. Alternative free software to check similarity such as DupliChecker or Smallseotools are useful for those without institutional access to paid software.³ Journal article submission portals and institutional libraries might also help to check for similarity. A point to note is that while many journals regularly screen for similarity after manuscript submission, any detection of plagiarism at this stage might be reported by the journal to the host

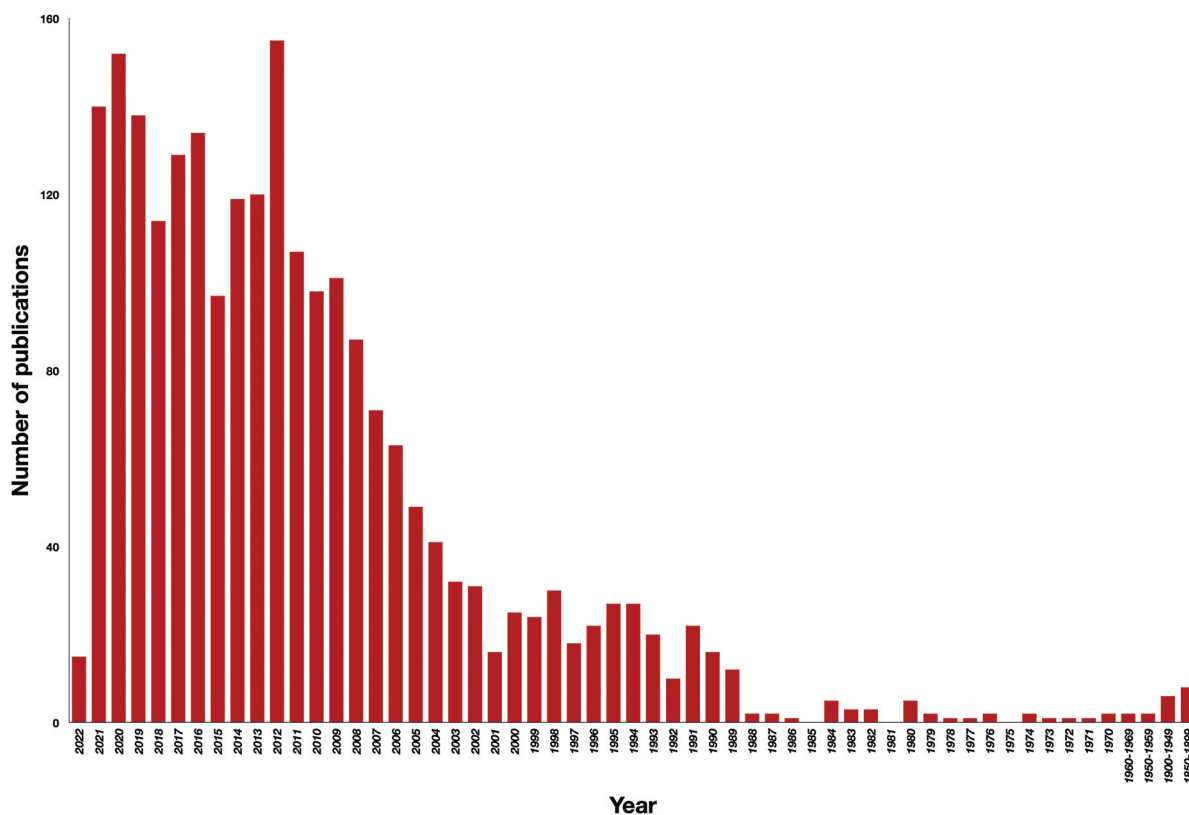


Fig. 1 Number of articles identified over time on a PubMed search. The search term used was “plagiarism” on February 27, 2022. Such a search includes articles that might have been flagged for plagiarism as well as those written on the topic of plagiarism. Overall, it reflects the relevance of the present topic over time.

institution of the submitting author and subsequently attract penalties for the authors.

As an example, iThenticate relies on the user to input content for similarity checking. Curation of content for similarity checking is essential to appropriately interpret output from iThenticate. The title page and references should not be included while checking similarity. This is because a lot of content from these sections will inherently be similar to previous literature, including author affiliations, statements regarding ethical approval and informed consent, and conflict of interest declarations. Another point to consider while feeding data for similarity checking is to set a limit for the number of consecutive words that should be similar before being flagged by iThenticate during similarity checking. Generally, such a limit is set at 8 or 10 words to avoid unnecessary flagging of commonly used phrases. Authors might copy content from elsewhere and delete or replace a few words here and there to avoid flagging of similarity by the software. Editors should suspect plagiarism when there is a sequence of content flagged to be similar with a few dissimilar words in between.^{5,10}

The original context to which similarity is flagged by the software should be sought and checked by editors and reviewers to arrive at a judgment as to whether this constitutes plagiarism. If the scientific paper has been published as a conference abstract previously, then this is likely to be picked up by iThenticate as considerably similar. However, this does not constitute plagiarism (although inexperienced editors and reviewers might be misled to think otherwise). It is a good practice for authors to declare prior conference presentations at the time of submission of the manuscript to avoid such a scenario.⁵ Increasingly, preprint publications are being used for early dissemination of results of a scientific study.¹¹ Such preprints should also be transparently declared during manuscript submission, otherwise they might be misconstrued as plagiarism after being flagged by similarity checking software.

Certain sections of the manuscript are more likely to be flagged as similar. For example, similar methods used in a previous paper might have considerable overlap of language, particularly for detailing laboratory experiments performed in a study. The reagents and machinery used for such tests will likely be identical across multiple studies. Even if such content is flagged as similar, it is of little consequence for making a judgment about plagiarism. On the contrary, minor degrees of similarity in the introduction, results, or discussion might be unacceptable.^{3,5,8}

Plagiarism of Ideas

This is probably the most difficult type of plagiarism to identify with any degree of certainty. Many a novel idea or hypothesis is based on a thorough analysis of preexisting literature in that particular area.¹² Hence, the same idea might conceivably have occurred simultaneously to two different research groups. Indeed, it is not uncommon for two or more scientific papers to publish their results related

to a similar scientific hypothesis separated by a short period of time.^{13,14}

The best way to avoid falling foul of plagiarism of ideas is to establish the primacy of one's idea by publishing it beforehand as a hypothesis or publishing a study protocol as a preprint. However, even this is not foolproof, as the idea could then be translated by a rival research group before the original group generating or publishing the idea has completed their experiments in relation to the idea. Plagiarism of ideas is difficult to detect for editors and reviewers. Generally, investigations related to alleged plagiarism of ideas come to light when a complaint is made to the journal by the person claiming primacy over the idea. The burden of proof of primacy of the idea generally rests with the complainant in such instances.

Plagiarism of Graphics or Tables

Authors should make every possible attempt to generate original figures or tables for their manuscripts. Adaptation of figures or tables from their own previously published papers is permissible with due permission from the copyright holder (which could be the authors themselves or the publisher) while duly citing the source of such adapted figures or tables.^{5,9} As an example, when we updated a review on the management of Takayasu arteritis,¹⁵ we had to seek permission of the copyright holder of the original paper (in this case, the publisher) before partially adapting a table from this source,¹⁶ even though we had conceptualized the original table ourselves. Adapting tables and figures from others' work is permissible with due attribution to the source after seeking permission of the copyright holder. In most instances, this is possible through the Web page of the article which provides a link to seek permission for the reproduction of such content. Rightslink is one such commonly used tool. Many a times, permission for academic or noncommercial reproduction of content is available at no or minimal cost.⁹ However, most journal editors would not prefer adaptation of such work which was not the original idea of the authors even if due processes are followed before adaptation.⁷

Plagiarism of tables is more frequently an issue for review articles. Whenever editors or reviewers are tasked with evaluating a review article, it is best to conduct a search for similar such review articles that have already been published, and go through their text and reference lists to assess whether any tables might have been plagiarized from them.^{5,10}

Plagiarism of figures is increasingly being recognized, with many authors preferring and journals recommending graphical abstracts to accompany original research work. Identifying duplication or similarity in figures is challenging. A starting point might be to search online repositories of scientific images such as Google Images for figures on a similar theme to that being evaluated. For review articles, a similar strategy to that proposed for identifying plagiarized tables might be useful. Advances in artificial intelligence might enable the development of tools in the future to identify more easily plagiarism of figures or tables.^{5,10}

From this discussion, it is clear that percentages of similarity cannot be a substitute for editorial or reviewer oversight to identify plagiarism. This criticism is particularly relevant for the current guidelines regarding plagiarism that have been issued by the University Grants Commission of India.¹⁷ These prescribe percentages of similarity in different sections of the manuscript as acceptable or unacceptable. Detection of plagiarism requires considerable human input supported by output from similarity checking software.⁸ Neither of these components alone can accurately judge the presence or absence of text plagiarism. Plagiarism of ideas, figures, or tables can presently be assessed only by thorough editorial or reviewer oversight. Prospective authors should carefully consider the points discussed in this article to avoid falling prey to plagiarism.

Ethical Statement

Not applicable.

Author Contributions

D.P.M. conceptualized, analyzed the relevant data, and wrote the manuscript.

Data Availability Statement

There is no data associated with this work.

Funding

None.

Conflict of Interest

None declared.

Acknowledgments

None.

References

- 1 Gasparyan AY, Nurmashev B, Seksenbayev B, Trukhachev VI, Kostyukova EI, Kitash GD. Plagiarism in the context of education and evolving detection strategies. *J Korean Med Sci* 2017;32(08):1220–1227
- 2 Misra DP, Ravindran V, Wakhlu A, Sharma A, Agarwal V, Negi VS. Plagiarism: a viewpoint from India. *J Korean Med Sci* 2017;32(11):1734–1735
- 3 Ahmed S, Anirvan P. The true meaning of plagiarism. *Indian J Rheumatol* 2020;15:155–158
- 4 Memon AR. Similarity and plagiarism in scholarly journal submissions: bringing clarity to the concept for authors, reviewers and editors. *J Korean Med Sci* 2020;35(27):e217
- 5 Misra DP, Ravindran V. Detecting and handling suspected plagiarism in submitted manuscripts. *J R Coll Physicians Edinb* 2021;51(02):115–117
- 6 Historical instances of plagiarism. Accessed February 27, 2022 at: <https://www.contentbot.ai/blog/news/15-examples-of-plagiarism-throughout-history/>
- 7 Misra DP, Agarwal V. Integrity of clinical research conduct, reporting, publishing, and post-publication promotion in rheumatology. *Clin Rheumatol* 2020;39(04):1049–1060
- 8 Misra D, Ravindran V, Agarwal V. Plagiarism: software-based detection and the importance of (Human) hardware. *Indian J Rheumatol* 2017;12(04):188–189
- 9 Misra DP, Ravindran V. Publication misconducts related to copyright: tread carefully to avoid falling. *J R Coll Physicians Edinb* 2020;50(01):3–5
- 10 Zimba O, Gasparyan AY. Plagiarism detection and prevention: a primer for researchers. *Reumatologia* 2021;59(03):132–137
- 11 Misra DP, Ravindran V. Preprint publications: waste in haste or pragmatic progress? *J R Coll Physicians Edinb* 2021;51(04):324–326
- 12 Misra DP, Gasparyan AY, Zimba O, Yessirkepov M, Agarwal V, Kitash GD. Formulating hypotheses for different study designs. *J Korean Med Sci* 2021;36(50):e338
- 13 Saadoun D, Garrido M, Comarmond C, et al. Th1 and Th17 cytokines drive inflammation in Takayasu arteritis. *Arthritis Rheumatol* 2015;67(05):1353–1360
- 14 Misra DP, Chaurasia S, Misra R. Increased circulating Th17 cells, serum IL-17A, and IL-23 in Takayasu arteritis. *Autoimmune Dis* 2016;2016:7841718
- 15 Misra DP, Wakhlu A, Agarwal V, Danda D. Recent advances in the management of Takayasu arteritis. *Int J Rheum Dis* 2019;22(Suppl 1):60–68
- 16 Misra DP, Sharma A, Kadiravan T, Negi VS. A scoping review of the use of non-biologic disease modifying anti-rheumatic drugs in the management of large vessel vasculitis. *Autoimmun Rev* 2017;16(02):179–191
- 17 University Grants Commission Academic Integrity Regulations Accessed February 28, 2022 at: https://www.ugc.ac.in/pdfnews/7771545_academic-integrity-Regulation2018.pdf