FROM THE EDITOR'S DESK

Spotlight on Plagiarism

Stephan Mertens

To copy from one book is plagiarism; from two, an essay; from three, a dissertation (1). This light hearted definition illustrates plagiarism's elusiveness. The spectrum ranges from word-for-word copying of published material to the misappropriation of hypotheses and arguments. The National Association of University Presidents in Germany defines plagiarism as the unauthorized use of intellectual property under the pretense of ownership (2). The originator is concealed from the recipient, with the intention of presenting the intellectual achievements of others as one's own. Opportunities to plagiarize are widespread and include peer-reviewed articles, research proposals, conference presentations, and other publications. In the world of science, plagiarism is considered misconduct and can lead to loss of academic and professional status.

A distinction is made between self-plagiarism, in which passages and data from one's own previous works reappear, and plagiarism of others' work, in which the intellectual property of others is used without due reference. Surveys and anecdotal impression suggest that self-plagiarism is much more common than other forms of plagiarism, and is tolerated, to a degree, within the scientific world. Plagiarism of others' work, on the other hand, is unanimously condemned, although unfortunately not always punished.

Given the huge number of journals and publications, it is reasonable to assume that only a small proportion

of instances of plagiarism are detected. The Internet has increased the temptation to copy text for use in one's own publication, but also the risk of being discovered.

The extent of plagiarism is difficult to gauge

In principle it should be possible to determine the frequency of plagiarism via surveying or comparing published articles, research proposals, and books. In an anonymous survey, nearly 5% of American scientists admitted to having published identical data on multiple occasions. 1.4% presented others' ideas as their own and 1.7% used confidential information for their own research (3). These figures are based on a survey of more than 3000 scientists in 2002 with a response rate of over 50%, in which mid-career researchers were distinguished from young scientists. The former published confidential information or previously published data significantly more frequently than young researchers. Since it is known that undesirable behaviors are often underreported, it is reasonable to assume that these figures are conservative.

Scientists from the University of Texas attempted to identify the perpetrators of plagiarism using a text recognition program. The program searched through databases, identifying identical passages in different documents. Using this method, Errami et al. found that 0.04% of over 60 000 Medline abstracts were probably duplicates of those of other authors (4). 1.3% of authors from this sample copied from their own work. The duplicated passages were three times more likely to occur in journals without an impact factor, and were less frequently cited.

Around three quarters of the identified duplicates were also identified by Medline's in-house algorithm as the "most relevant article." Authors Errami and Garner used this method of cross-referencing with Medline to analyze a further 7 million abstracts, and were able to detect 70 000 very similar abstracts (5). Based on previous research (4), the researchers predicted that there would be 50 000 true copies among them. In other words, about 0.7% of the originally included abstracts are of questionable origin. Extending this to the whole database of over 17 million articles, the authors predict that Medline contains over 200 000 instances of duplication. The Figure shows the ratio of suspected plagiarized articles to other articles as an international comparison. Germany represents the fourth-largest contributor to Medline publications with a slightly greater proportion of potential instances of plagiarism.

Countries with the greatest proportion of Medline entries. Fewer potentially plagiarized articles relative to the proportion of Medline-listed articles were found in English speaking countries, and more in Japan and China. By kind permission of Nature Publishing Group, from Errami M, Garner H: A tale of two citations.

FIGURE

BOX

Procedure on suspicion of plagiarism

- Examination of allegation
- On ascertainment of plagiarism, a statement should be requested of authors.
- If no plausible explanation is given submitted articles should be rejected and reviewers, authors' supervisors and their institutions should be notified.
- Publishers should recall previously published articles and notify all relevant databases.
- The affected authors and journal should be informed.

Summary of the approved procedure as per the guidelines of the Committee on Publication Ethics (www.publicationethics.org)

However, the estimate of 200 000 duplicated articles in Medline could be too high, as a subgroup analysis of the suspected articles suggests. A manual inspection of 65% of potential duplicate articles found that just 5 were in fact duplicates, and a further 5, translations into another language (6). It may be that articles in different languages had similar abstracts that allowed them to be identified. The small number of genuine instances of plagiarism underlines that electronic searches can generate a high rate of false-positive results, and that manual checks are imperative. If these results are correct, the number of duplicated articles in Medline ought to be around 20 000, or 0.1%. Canadian researchers who searched for duplicates across all scientific disciplines in the Web of Science database present similar results. If the title of the article, the first author and the number of references were identical, the researchers assumed it was a duplicate publication. Using this framework, they discovered 5000 duplicated articles in the meta data of 18 000 articles published between 1980 and 2007 (7). Because the definition is strict, with even slightest change in the title preventing identification as a duplicated article, the authors consider this a conservative estimate. Articles from the field of clinical medicine were found to have the third most frequent occurrence of duplicates of all investigated disciplines. The article pairs tended to appear within a year of one another, which supports the assumption that they were submitted simultaneously to different journals.

Perpetrators and affected parties

Tara Long and colleagues wanted to know how affected journals and authors react when plagiarism is discovered. Using the database *Déjà vu*, in which potential duplicate publications are recorded, the researchers found 212 pairs of suspected plagiarist articles (8). On average 86.2% of the text and 73.1% of references were identical, and in 71.4% tables and figures were very similar or identical.

It is notable that only 47 perpetrators of plagiarism (22.2%) cited the original source. In order to gauge responses, Long and colleagues sent out a questionnaire along with the two articles to plagiarized and plagiarizing authors, and the affected journals, for 163 cases of plagiarism. 93% of original authors were unaware of the plagiarism. Only 37% of plagiarizers responded to the allegation. In response to evidence of plagiarism, half of editorial teams initiated investigations (*Box*), leading to the withdrawal of the plagiarized article in just 46 cases. Despite the scientific and ethical implications, half of the editorial teams took no further action and, in total, only 25% of the duplicated articles were officially corrected. Why this was left undone in 75% of cases is unclear. It may be that journals feared damage to their image. With a score of 3.87, the impact factor of the original articles was significantly higher than that of the second publication at 1.6 ($p < 0.001$). Original articles were, on average, cited 28 times, the plagiarized pieces just twice. Since plagiarized pieces are generally published later, they appear higher up in the search results of databases than the originals on account of their more recent date of publication, and are therefore possibly more likely to be cited in quick reviews of the literature.

The editorial team of *Deutsches Ärzteblatt* is also aware of cases in which authors have submitted manuscripts that to varying degrees are identical to their own or other authors' work. However, exact figures are not available to us. It is also perfectly possible that plagiarized articles have remained undetected, and been published.

Consequences

The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and international organizations such as the International Committee of Medical Journal Editors and the Committee on Publication Ethics suggest that proven instances of plagiarism result in withdrawal of the article (9–11). Correction should take place not only within the journal that published the plagiarized piece, but also in the relevant references in databases. Only in this way can plagiarism be prevented from influencing meta-analysis, and therefore distorting the scientific debate by allowing plagiarized as well as original data to be taken into account, leading to publication bias.

Conflict of interest statement

The author is an editor in the medical scientific section of *Deutsches Ärzteblatt*.

REFERENCES

1. Hochschule für Technik und Wirtschaft Berlin <http://plagiat.htw-berlin.de>
2. Deutsche Hochschulrektorenkonferenz www.hrk.de/de/be-schluesse/109_422.php
3. Martinson BC, Anderson MS, de Vries R: Scientists behaving badly. *Nature* 2005; 435: 737–8.
4. Errami M, Hicks JM, Fisher W, et al.: *Déjà vu* – a study of duplicate citations in Medline. *Bioinformatics* 2008; 24: 243–9.
5. Errami M, Garner H: A tale of two citations. *Nature* 2008; 451: 397–9.

6. Rifai N, Bossuyt PM, Bruns DE: Identifying duplicate publications: primum non nocere: Clin Chem 2008; 54: 777–8.
7. Larivière V, Gingras YY: On the prevalence and scientific impact of duplicate publications in different scientific fields (1980-2007). J Doc 2010; 66: 179–90.
8. Long TC, Errami M, George AC, Sun Z, Garner HR: Responding to possible plagiarism. Science 2009; 323: 1293–4.
9. International Committee of Medical Journal Editors. www.icmje.org
10. Committee on Publication Ethics (COPE). www.publicationethics.org
11. Office of Research Integrity. www.ori.dhhs.gov

Corresponding author

Dr. sc. nat. Stephan Mertens
Deutsches Ärzteblatt
Medizinisch-wissenschaftliche Redaktion
Ottostr. 12, 50859 Köln, Germany
mertens@aerzteblatt.de

Translated from the original German by Dr. Sandra Goldbeck-Wood.

Cite this as:

Mertens S: Spotlight on plagiarism. Dtsch Arztebl Int 2010; 107(49): 863–5.
DOI: 10.3238/arztebl.2010.0863