

BMJ Open Plagiarism in research: a survey of African medical journals

Anke Rohwer,¹ Elizabeth Wager,^{2,3} Taryn Young,¹ Paul Garner⁴

To cite: Rohwer A, Wager E, Young T, *et al*. Plagiarism in research: a survey of African medical journals. *BMJ Open* 2018;**8**:e024777. doi:10.1136/bmjopen-2018-024777

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-024777>).

Received 13 June 2018

Revised 4 September 2018

Accepted 20 September 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Centre for Evidence-based Health Care, Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

²Sideview, Princes Risborough, UK

³School of Medicine, University of Split, Split, Croatia

⁴Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

Correspondence to

Dr Anke Rohwer;
arohwer@sun.ac.za

ABSTRACT

Objectives To examine whether regional biomedical journals in Africa had policies on plagiarism and procedures to detect it; and to measure the extent of plagiarism in their original research articles and reviews.

Design Cross sectional survey.

Setting and participants We selected journals with an editor-in-chief in Africa, a publisher based in a low or middle income country and with author guidelines in English, and systematically searched the African Journals Online database. From each of the 100 journals identified, we randomly selected five original research articles or reviews published in 2016.

Outcomes For included journals, we examined the presence of plagiarism policies and whether they referred to text matching software. We submitted articles to Turnitin and measured the extent of plagiarism (copying of someone else's work) or redundancy (copying of one's own work) against a set of criteria we had developed and piloted.

Results Of the 100 journals, 26 had a policy on plagiarism and 16 referred to text matching software. Of 495 articles, 313 (63%; 95% CI 58 to 68) had evidence of plagiarism: 17% (83) had at least four linked copied or more than six individual copied sentences; 19% (96) had three to six copied sentences; and the remainder had one or two copied sentences. Plagiarism was more common in the introduction and discussion, and uncommon in the results.

Conclusion Plagiarism is common in biomedical research articles and reviews published in Africa. While wholesale plagiarism was uncommon, moderate text plagiarism was extensive. This could rapidly be eliminated if journal editors implemented screening strategies, including text matching software.

INTRODUCTION

Plagiarism is a serious form of research misconduct when authors copy text, ideas or images from another source, and take credit for it.^{1 2} The severity varies from copying short phrases to copying of a whole paper. Besides the amount of text that is copied, assessors should consider how it was referenced, whether the deception was intentional or not, as well as whether the copied text is a commonly used or an original phrase.^{3 4} Redundant publication is an umbrella term for reusing one's own work, and ranges from reusing large parts of already published text

Strengths and limitations of this study

- This study is the first to systematically research plagiarism in African biomedical journals.
- We developed a method for reporting the extent of plagiarism beyond the overall similarity index.
- Our analysis was limited to text and excluded images and data.
- The high level of plagiarism we identified could easily be solved by screening all articles with text matching software and automatic rejection of articles showing plagiarism.
- We used an online source, the African Journals Online database, as the sampling frame for our study.

(text recycling), to publishing parts of the same study in more than one paper (salami slicing) and republishing entire papers (duplicate publication), and is also considered poor practice.^{5 6}

The availability of material on the internet facilitates mosaic writing and plagiarism, but the widespread availability of text matching software has improved detection so there is now more awareness of research integrity and research misconduct, including plagiarism. Policies are clearly available through the Committee on Publication Ethics (COPE), encouraging journal editors to screen submitted manuscripts for plagiarism.⁷ Publishing systems and standards have advanced rapidly with online publishing in a global world, and there are some cooperative programmes between the large and local players to help local players keep up with advances. An example of this is the African Journals Partnership Project (AJPP), a programme that partners African journals with mentor journals from the USA and UK.⁸

Estimates of the occurrence of plagiarism are largely based on findings from a systematic review by Pupovac and Fanelli (2014), who reported self-reported plagiarism estimates of 1.7% (95% CI 1.2 to 2.4) for participants admitting to having plagiarised, and 30% (95% CI 17 to 46) for participants knowing about others who had done so.⁹

However, none of the included studies were conducted in low or middle income countries (LMICs) although we know from our own work that Cochrane authors in Africa and other LMICs report that plagiarism is common in host institutions.¹⁰ Any self-reported estimate is probably not an accurate reflection of actual practice, mainly due to social desirability bias.¹¹ Some studies have examined plagiarism more objectively by using text matching software to screen manuscripts^{12–15} but mostly these examined manuscripts submitted to journals (before publication) and none included manuscripts submitted to or published in African journals.

We sought to examine whether regional biomedical journals in Africa had policies on plagiarism and procedures to detect it; and to measure the extent of plagiarism in their published original research articles and reviews.

METHODS

Study design and sample

We surveyed original research articles published in biomedical journals indexed on the African Journals Online database (AJOL).¹⁶ Journals were eligible if their current editor-in-chief was based in Africa, the publisher was based in a LMIC (according to the World Bank),¹⁷ if policies and author guidelines were available in English and if the journal published an issue in 2016. All eligible journals were selected. From each eligible journal, we selected articles published in 2016 as original research articles, including qualitative and quantitative primary studies, literature reviews and systematic reviews, published in English. We excluded editorials and letters. We used Microsoft Excel to generate a list of random numbers to select five articles from each eligible journal. We selected five articles per journal, as initial scoping of journals indexed on AJOL revealed substantial variation in the number of published articles per issue, as well as the number of published issues per year.

Data collection

For eligible journals, we downloaded policies and author instructions from the journal's website. We extracted data on the presence and content of policies and guidelines on plagiarism. For research articles, we downloaded the full text (PDF) of each article. We extracted data on the number of authors, country of corresponding author and type of study. One author (AR) extracted data using a prespecified, piloted data extraction form (see online supplementary file 1) and entered it into Excel.

We measured the presence and extent of plagiarism (copying of someone else's work) and redundancy (copying of one's own work) in all included research articles. We submitted the PDFs of all articles to Turnitin text matching software. Turnitin generated a similarity report containing the overall similarity index (OSI), expressed as the percentage of matching text,¹⁸ excluding quotations and references. We manually reviewed all similarity reports with the plagiarism framework (table 1). As we

Table 1 Plagiarism framework

| | No of copied sentences detected | | |
|----------------------|--|--|--|
| | Level 1 | Level 2 | Level 3 |
| Abstract | 1 to 2 | 3 to 6 | 6+; or 4+linked |
| Background | 1 to 2 | 3 to 6 | 6+; or 4+linked |
| Methods | 1 to 2 | 3 to 6 | 6+; or 4+linked |
| Results | 1 to 2 | 3 to 6 | 6+; or 4+linked |
| Discussion | 1 to 2 | 3 to 6 | 6+; or 4+linked |
| Overall score | Some plagiarism | Moderate plagiarism | Extensive plagiarism |
| Definition | One or more sections with plagiarism of 1 to 2 sentences; or level 2 plagiarism in the methods section | One or more sections with plagiarism of 3 to 6 sentences; or level 3 plagiarism in the methods section | One or more sections with plagiarism of 4 or more linked sentences, or plagiarism of more than 6 sentences |

were not able to find any existing guidance to objectively assess the extent of plagiarism, we developed a framework based on suggestions from COPE³ and Wager⁴ that propose differentiating between clear plagiarism and minor copying of someone else's (plagiarism) and one's own text (redundancy). We assessed the extent of plagiarism, stratified by which section of the article it appeared in.

We identified copied sentences from the similarity reports. Sentences had to be substantially or completely copied. When a sentence had been clearly copied but prefixed by 'However' or 'Researchers found that...', this was classed as copying, and where plagiarised strings of sentences were detected joined together with conjunctions, this was classed as copying (see online supplementary file 2). Once we identified a copied sentence, we checked the source of the original sentence, as stated in the similarity report. If the source of the original sentence contained one or more of the authors of the article under investigation, we classified it as redundancy, whereas if the source of the original sentence was from other authors, we classified it as plagiarism.

For each section of the article, we counted the number of copied sentences and assigned one of three levels, depending on the number of copied sentences (table 1). We then assigned an overall plagiarism category, using the same criteria for each section of the article, to describe the extent of plagiarism—namely, 'some', 'moderate' or 'extensive' plagiarism (table 1). As methods copying was common, and can happen when people are using standard methods, we adjusted the definition to take this into account. Therefore, copying of one to two sentences in the methods section was not regarded as plagiarism, copying of three to six sentences was regarded as some plagiarism and copying of more than six sentences or at least four linked sentences was regarded as moderate

plagiarism (table 1). Overall redundancy was scored in an equivalent way and for each article. Separate scores were given for plagiarism and redundancy.

Development of the framework was an iterative process, and piloted by AR and EW who independently assessed similarity reports of 10 articles and discussed results with the entire research team. Once the team had agreed on the framework, one author (AR) scored all similarity reports using the framework and another author (EW) independently scored a random selection of 10% of reports. Any disagreements in rating were resolved by consensus.

Data analysis

We used SPSS (V.25)¹⁹ for analysis. We report categorical data as frequencies and proportions, and continuous data as medians, means and SD, or modes and ranges. For plagiarism and redundancy, we calculated 95% CI, adjusted for clustering at the journal level using robust standard errors, with STATA (V.15).

Patient and public involvement

We did not involve patients or the public in this study.

Ethical issues

All data used in this study were available online and thus in the public domain. To ensure anonymity of authors, we did not include information identifying individual research articles in our report. We obtained an ethics exemption from the Stellenbosch University Health Research Ethics Committee (X17/08/010). Where we detected serious plagiarism in published papers, we

identified the journal editors and are currently writing to them, informing them of our findings.

RESULTS

Of the 179 biomedical journals indexed on AJOL, 100 met the eligibility criteria and were included in the study (figure 1). Detailed characteristics of journals are reported in the table of included journals (see online supplementary file 3), while excluded journals are listed in the table of excluded journals (see online supplementary file 4).

We selected five research articles published in the 2016 issue of each journal. Some had not published this number (one journal only published two research articles, and two journals published four research articles). For these we included all research articles published in 2016, giving a total of 495 research articles included (figure 1).

Plagiarism policies in included journals

Twenty six per cent of the journals had a policy on plagiarism mentioned on their website (table 2). More journals with open (35%) compared with paid (6%) access and more specialised (38%) compared with general (13%) journals mentioned a policy. Journals with a plagiarism policy included both those from non-commercial (22%) and commercial (32%) publishers. Journals with the same commercial publisher generally had similar policies. All journals published by 'AOSIS publications' or 'Health and Medical Publishing Group' had a policy and referred to text matching software. None of the journals published

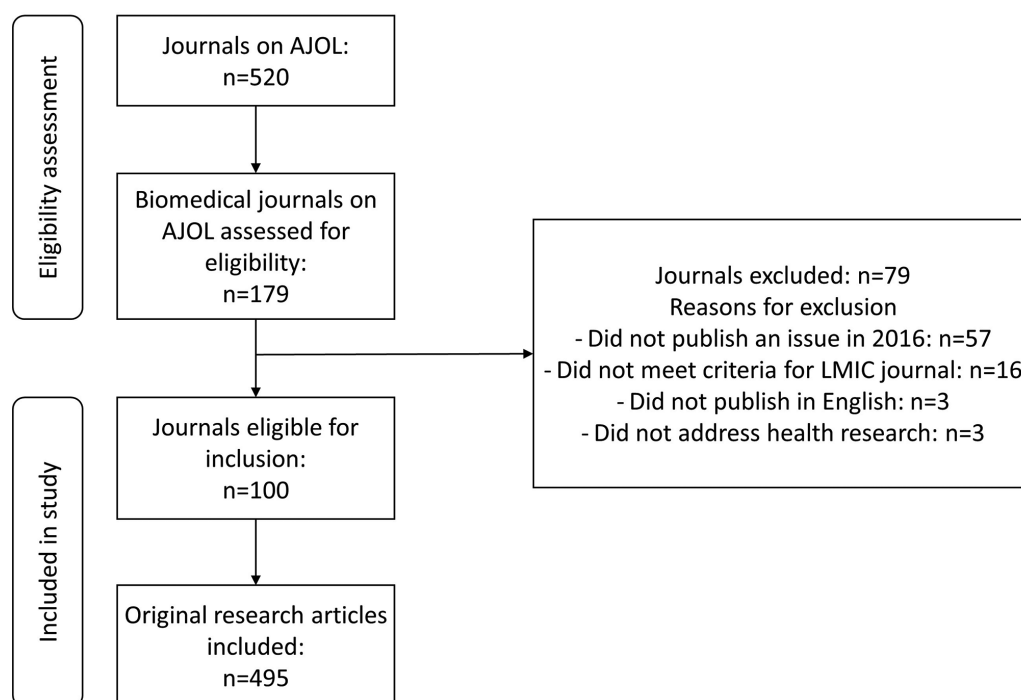


Figure 1 Flow diagram of included journals and research articles. AJOL, African Journals Online database; LMIC, low or middle income country.

Table 2 Plagiarism policies in included journals (n=100)

| | Publisher | | Access | | Scope | | |
|--|---------------------------------|-----------------------------|--------------------|--------------------|----------------|--------------------|---------------|
| | Non-commercial publisher (n=59) | Commercial publisher*(n=41) | Open access (n=69) | Paid access (n=31) | General (n=48) | Specialised (n=52) | Total (n=100) |
| Plagiarism policy available (n (%)) | 13 (22) | 13 (32) | 24 (35) | 2 (6) | 6 (13) | 20 (38) | 26 |
| Definition of plagiarism (n (%)) | 5 (8) | 9 (22) | 13 (19) | 1 (3) | 2 (4) | 12 (23) | 14 |
| Reference to text matching software (n (%)) | 6 (10) | 10 (24) | 14 (20) | 2 (6) | 3 (6) | 13 (25) | 16 |
| Consequences of plagiarism described (n (%)) | 11 (19) | 10 (24) | 20 (29) | 1 (3) | 6 (13) | 15 (29) | 21 |
| Reference to COPE flowchart (n (%)) | 2 (3) | 2 (5) | 3 (4) | 1 (3) | 0 | 4 (8) | 4 |

*Medknow Publications, based in India (19 journals); Health and Medical Publishing group (6 journals), Medpharm Publications (4 journals), AOSIS Publishing (3 journals), In House publications (2 journals) and LAM publications Ltd (1 journal), all based in South Africa; Bookbuilders Africa (1 journal), Michael Joanna Publications (1 journal), Fine Print and Manufacturers (1 journal), CME ventures (1 journal) and SAME ventures (1 journal) based in Nigeria; and AKS publications (1 journal), based in Mauritius. COPE, Committee on Publication Ethics.

by 'Medknow publications' (19 journals) and 'In House publications' (two journals) mentioned a policy. For 'Medpharm publications', three of the four journals had a plagiarism policy, but only one of these also referred to text matching software. Of the nine journals with an impact factor, three did not have a policy on plagiarism and six of the seven AJPP member journals had no policy.

Sixteen journals stated that they used text matching software to check for plagiarism, and of these, there were more journals from commercial (24%) than non-commercial publishers (10%), more journals with open (20%) than paid access (3%) and more specialised (25%) than general (6%) journals (table 2).

Characteristics of included research articles (n=495)

The characteristics of the included research articles are summarised in table 3. Most articles were published in open access journals (69%) and about half (48%) in a journal with a general scope; 41% were published in journals from a commercial publisher. Non-commercial publishers included research institutions and academic organisations that published their journals themselves.

Nine journals had an impact factor and accounted for 9% of the papers included, and seven journals were members of the AJPP (7% of research articles). Articles had a median of three authors (min 1, max 10). Overall, half of the included articles had corresponding authors based in Nigeria. Half (50%) of the included articles represented cross sectional studies.

Overall similarity index of included articles

A summary of the OSIs for all included articles is reported in table 4. The median OSI was 15%, with a minimum OSI of 0% and a maximum of 68%. Of all the included papers, 90% had an OSI of 30% or less. All five articles with an OSI of 50% or more were published in non-commercial journals.

Rates and extent of plagiarism and redundancy per section of article

The presence of plagiarism varied across different sections of the articles (see online supplementary file 5). We did not find widespread plagiarism or redundancy in the results sections of included articles. Plagiarism was mostly in the introduction of articles (47%) followed by

Table 3 Summary of the characteristics of included articles (n=495)

| Characteristic | n (%) |
|---|----------|
| Published in a journal with | |
| Impact factor | 45 (9) |
| Open access | 342 (69) |
| General scope | 239 (48) |
| Commercial publisher | 205 (41) |
| African Journals Partnership Project membership | 35 (7) |
| Country of corresponding author | |
| Nigeria | 250 (51) |
| South Africa | 83 (17) |
| Other African country | 99 (20) |
| Non-African country | 63 (13) |
| Type of study | |
| Cross sectional study | 247 (50) |
| Retrospective study | 65 (13) |
| Case report | 42 (9) |
| Trial | 36 (7) |
| Cohort study | 22 (4) |
| Review | 21 (4) |
| Case control study | 12 (2) |
| Other | 50 (10) |

Table 4 Overall similarity index of included articles (n=495)

| Overall similarity index | No of articles (%) |
|--------------------------|--------------------|
| 0–10% | 137 (28) |
| 11–20% | 202 (41) |
| 21–30% | 104 (21) |
| 31–40% | 34 (7) |
| 41–50% | 13 (3) |
| 51–60% | 2 (0.4) |
| 61–70% | 3 (0.6) |
| 71–100% | 0 |

the discussion (39%) and the methods section (30%). The extent of plagiarism also varied across sections, and plagiarism of one to two sentences occurred most commonly. Plagiarism in the introduction comprised one to two copied sentences in 23% of articles, three to six copied discrete sentences in 14% and at least four linked or more than six discrete copied sentences in 11% of articles. In the discussion section, plagiarism comprised one to two copied sentences in 18% of articles, three to six copied sentences in 13% and at least four linked or more than six discrete copied sentences in 9% of articles.

Redundancy was mostly seen in the methods section of included articles (11%), comprising one to two copied sentences in 3% of articles, three to six copied sentences in 4% and at least four linked or more than six discrete copied sentences in 3% of articles (see online supplementary file 5).

Overall plagiarism in included articles

We found plagiarism (any level) in 63% of articles, comprising some plagiarism in 27%, moderate plagiarism in 19% and extensive plagiarism in 17% of articles (table 5).

We explored the characteristics of articles with plagiarism (table 6). Articles published in journals that referred to text matching software tended to have less plagiarism than those in journals that did not refer to text matching software, with rates of 43% versus 66%, respectively, for any level of plagiarism, and 6% versus 19% for extensive plagiarism. The difference in plagiarism rates for articles published in a journal with a policy on plagiarism (54%)

compared with those published in a journal without a policy on plagiarism (66%) was smaller. Although the proportion of reviews with any plagiarism was comparable with other studies, almost half of all included reviews (48%) had extensive plagiarism.

Redundancy was less common than plagiarism. Overall, 11% of articles had any level of redundancy, comprising 4% of articles with some redundancy, 4% with moderate redundancy and 2% with extensive redundancy (table 7).

Accuracy of various OSI thresholds

We explored the accuracy of various thresholds of OSIs according to our plagiarism framework (table 8). With an OSI threshold of 5%, sensitivity was high, meaning that 97% of articles with any level of plagiarism were correctly identified as such and only 3% of articles with any level of plagiarism were missed. However, specificity was low, meaning only 17% of articles without any plagiarism were correctly identified as such and the rate of false positives was high (83%). Increasing the threshold led to decreased sensitivity and increased specificity.

DISCUSSION

Our study is the first to explore actual levels of plagiarism in biomedical journals from Africa. We proposed a framework to measure plagiarism as an OSI generated by text matching software on its own is not sufficient to describe the presence and extent of copied text. Indeed, the OSI is only an indication of the proportion (%) of copied text¹ and there is no consensus of an acceptable threshold. Furthermore, the reported sensitivity of OSI thresholds varies across studies. In our sample, the sensitivity for an OSI threshold of 10% was 84%, compared with 97% for a threshold of 5%. Taylor *et al* found an even lower sensitivity of 67% for an OSI threshold of 11.5%, excluding citations and references,²⁰ while Higgins *et al* found that an OSI threshold of 10% yielded a sensitivity of 95.5%.¹³ Zhang used text matching software to screen manuscripts submitted to a Chinese journal for plagiarism¹⁴ and found that 23% contained plagiarism or redundancy, of which 25% contained high levels of plagiarism. However, it is not clear how plagiarism was defined. A study from Pakistan that assessed plagiarism of submitted manuscripts¹⁵ found that 39% of papers contained plagiarised text,

Table 5 Overall plagiarism (n=495)

| Plagiarism score | Definition | n | % (95% CI) |
|--------------------------------|---|-----|----------------|
| Any level of plagiarism | At least one or more sections with plagiarism of 1 to 2 sentences; or level 2 plagiarism in the methods section | 313 | 63% (58 to 68) |
| Some plagiarism (level 1) | One or more sections with plagiarism of 1 to 2 sentences; or level 2 plagiarism in the methods section | 134 | 27% (23 to 32) |
| Moderate plagiarism (level 2) | One or more sections with plagiarism of 3 to 6 sentences; or level 3 plagiarism in the methods section | 96 | 19% (16 to 23) |
| Extensive plagiarism (level 3) | One or more sections with plagiarism of 4 or more linked sentences, or plagiarism of more than 6 separate sentences | 83 | 17% (13 to 21) |

Table 6 Characteristics of original articles and reviews with plagiarism (n=495)

| Characteristic | Overall plagiarism n (%) | | | |
|---|--------------------------|----------|--------------|---------------|
| | Any (%) | Some (%) | Moderate (%) | Extensive (%) |
| Impact factor | | | | |
| Yes (n=45) | 23 (51) | 9 (20) | 10 (22) | 4 (9) |
| No (n=450) | 290 (64) | 125 (28) | 86 (19) | 79 (18) |
| Open access | | | | |
| Yes (n=342) | 206 (60) | 85 (25) | 64 (19) | 57 (17) |
| No (n=153) | 107 (70) | 49 (32) | 32 (21) | 26 (17) |
| Scope general | | | | |
| Yes (n=239) | 171 (72) | 71 (30) | 58 (24) | 42 (18) |
| No (n=256) | 142 (55) | 63 (25) | 38 (15) | 41 (16) |
| Member of African Journals Partnership Project | | | | |
| Yes (n=35) | 22 (63) | 12 (34) | 6 (17) | 4 (11) |
| No (n=460) | 291 (63) | 122 (27) | 90 (20) | 79 (17) |
| Plagiarism policy available | | | | |
| Yes (n=127) | 69 (54) | 31 (24) | 21 (17) | 17 (13) |
| No (n=368) | 244 (66) | 103 (28) | 75 (20) | 66 (18) |
| Reference to text matching software | | | | |
| Yes (n=80) | 34 (43) | 18 (23) | 11 (14) | 5 (6) |
| No (n=415) | 279 (67) | 116 (28) | 85 (20) | 78 (19) |
| Commercial publisher | | | | |
| Yes (n=205) | 112 (55) | 51 (25) | 33 (16) | 28 (14) |
| No (n=290) | 201 (69) | 83 (29) | 63 (22) | 55 (19) |
| Country of corresponding author | | | | |
| Nigeria (n=250) | 175 (70) | 63 (25) | 63 (25) | 49 (20) |
| South Africa (n=83) | 32 (39) | 19 (23) | 6 (7) | 7 (8) |
| Other African country (n=99) | 67 (68) | 33 (33) | 17 (17) | 17 (17) |
| Non-African country (n=63) | 39 (62) | 19 (30) | 10 (16) | 10 (16) |
| Type of study | | | | |
| Cross sectional study (n=247) | 164 (66) | 78 (32) | 50 (20) | 36 (15) |
| Retrospective study (n=65) | 40 (62) | 22 (34) | 10 (15) | 8 (12) |
| Case report (n=42) | 27 (62) | 10 (24) | 9 (21) | 8 (19) |
| Trial (n=36) | 23 (64) | 6 (17) | 10 (28) | 7 (19) |
| Cohort study (n=22) | 9 (41) | 1 (5) | 4 (18) | 4 (18) |
| Review (n=21) | 14 (67) | 4 (19) | 0 | 10 (48) |
| Case control study (n=12) | 8 (67) | 0 | 5 (42) | 3 (25) |
| Case series (n=12) | 5 (42) | 3 (25) | 0 | 2 (17) |
| Qualitative study (n=9) | 6 (67) | 3 (33) | 3 (33) | 0 |
| Laboratory study (n=8) | 6 (75) | 2 (25) | 2 (25) | 2 (25) |
| Mixed methods (n=7) | 2 (29) | 1 (14) | 0 | 1 (14) |
| Before–after (n=7) | 3 (43) | 2 (29) | 1 (14) | 0 |
| Controlled before–after (n=7) | 6 (86) | 2 (29) | 2 (29) | 2 (29) |

using a strict definition of the presence of one or more copied sentences. They reported similar results for papers from Turkey and China. A study assessing plagiarism in manuscripts submitted to the *Croatian Medical Journal*¹²

identified plagiarism, defined as an OSI of >10% in any section of the manuscript, in 11% (85/754) of manuscripts, of which 8% (63/754) were classified as plagiarism of others, while 3% (22/754) was self-plagiarism (ie,

Table 7 Overall redundancy (n=495)

| Redundancy score | Definition | n | % (95 CI) |
|--------------------------------|---|----|---------------|
| Any level of redundancy | At least one or more sections with redundancy of 1 to 2 sentences; or level 2 redundancy in the methods section | 54 | 11% (8 to 15) |
| Some redundancy (level 1) | One or more sections with redundancy of 1 to 2 sentences; or level 2 redundancy in the methods section | 21 | 4% (3 to 7) |
| Moderate redundancy (level 2) | One or more sections with redundancy of 3 to 6 sentences; or level 3 redundancy in the methods section | 22 | 4% (3 to 7) |
| Extensive redundancy (level 3) | One or more sections with redundancy of 4 or more linked sentences, or redundancy of more than 6 sentences | 11 | 2% (1 to 4) |

redundancy). In a sample of 400 manuscripts submitted to an American specialty journal, 17% were found to have at least one copied sentence. Half of these (53%) were regarded as being self-plagiarism.¹³ Our study assessed plagiarism in published articles and found a much higher rate of plagiarism than other studies. In our sample, 72% of articles had an OSI above 10%, and 63% (95% CI 58 to 68) had any level of plagiarism, while 11% (95% CI 8 to 15) had any level of redundancy. It is possible that the rate of plagiarism in manuscripts submitted to these journals (but not published) is even higher.

In line with recommendations for best practices,^{6 7 21} increasing numbers of journal editors and publishers, especially the large publishing houses, make use of text matching software to screen submitted manuscripts for copied text.²² But software licences are expensive, and some smaller journals, especially institutional journals and those with non-commercial publishers, may not be able to afford them.²³ Indeed, we found that of the 26% of journals that had a policy of plagiarism, most were from commercial publishers. In addition, only 16% of journals in our sample mentioned the use of text matching software.

Our framework is limited in that it only measures plagiarism in terms of the number of copied sentences, although it does take into account where in the article the copied text was found. We considered plagiarism in the methods section to be less serious than plagiarism in other sections of the articles, as it is sometimes difficult to avoid repeating standard descriptions of methods.^{5 24} Our framework does not, however, consider other aspects of plagiarism, such as how the text was referenced, whether the copied text referred to a standard phrase or common knowledge and whether plagiarism was intentional or not,⁴ which are important aspects to consider when making judgements about plagiarism.

Table 8 Sensitivity and specificity of various overall similarity index thresholds

| OSI threshold | Sensitivity (%) | Specificity (%) |
|---------------|-----------------|-----------------|
| OSI >5% | 97 | 17 |
| OSI >10% | 84 | 51 |
| OSI >15% | 66 | 83 |

OSI, overall similarity index.

The framework is also limited to plagiarism of text and does not take into account plagiarism of data or images (which is also a limitation of text matching software). To test our framework, one author (AR) checked all the articles and another author (EW) checked a random sample of 10% of the articles. While our scores for overall plagiarism were mostly consistent, we found that variations depended on how we scored borderline cases in terms of what was considered a completely copied sentence. The framework therefore may lack precision in terms of inter-rater reliability and test-retest reliability, and needs further testing. However, we found that the framework was still a useful tool which facilitated assessment across articles and represented the extent of plagiarism well.

We were interested in regional journals and wanted to examine smaller and non-mainstream journals based in Africa. We considered various sampling frames, but few met our requirements. We chose AJOL to sample journals as it hosts over 500 journals, including 179 biomedical journals, from over 30 African countries. In addition, journals indexed on AJOL need to meet certain criteria linked to good publishing practices. These include, inter alia, a functioning editorial board, peer review of content and availability of content in electronic format.¹⁶ In light of the known challenges in identifying and accessing African biomedical journals,^{25 26} we thus considered AJOL to be a comprehensive and pragmatic sampling frame, although it does not represent all African biomedical journals.

The recently established African Network for Research Integrity (ANRI) has recognised the need to raise widespread awareness about research integrity among African researchers to prevent poor practices related to plagiarism, redundant publication, authorship and conflicts of interest.²⁷ Furthermore, the need to build capacity of African journals to improve the quality and visibility of African research has previously been recognised. In an attempt to address this need, the AJPP was initiated in 2004.^{8 26} In addition to building capacity of specific member journals, the project has also envisaged that the African members become 'regional leaders and share their acquired knowledge and experience with other editors and journals on the continent'.²⁶ Although only seven of our included journals were members of AJPP, the proportion of articles with any level of plagiarism was the

same for member and non-member journals. Only one of the seven AJJP member journals mentioned a policy on plagiarism. It is possible that journals had plagiarism policies but did not mention them on their online information; however, given the actual amount of plagiarism we found, we think this is unlikely, and since one purpose of such policies is to act as a deterrent we believe they should be clearly publicised by journals.

The level of plagiarism in African biomedical journals is concerning. African journals should aim to meet global expectations and follow best practices with regards to their policies and guidelines on plagiarism. This includes using text matching software to detect plagiarism and redundancy in submitted manuscripts. Not only will this help to verify the originality of submitted work, but it also has the potential to deter poor practices. However, although text matching software is a useful screening tool, editors should not rely on the OSI on its own. A high similarity score should trigger detailed assessment by a knowledgeable editor, but the possibility of false positives and false negatives should always be borne in mind and was clearly shown in our study. Our plagiarism framework provides an approach to classify the extent of plagiarism. Further testing of the tool is necessary to determine validity and reliability.

This paper has uncovered a major problem with writing and publishing in medical science in Africa needing attention both through institutional development of expectations and good practice in academic institutions, and development of journal editorial procedures to detect and respond to the problem.

Acknowledgements We would like to thank Selvan Naidoo and Traci Naidoo for their assistance in eligibility assessment of journals and Tonya Esterhuizen for statistical support. Anke Rohwer conducted this research as part of her PhD at Stellenbosch University, South Africa.

Contributors All authors contributed to the design of the study and the development of the plagiarism framework. AR collected and analysed the data from journals and articles with input from TY, PG and EW. AR reviewed all manuscripts using the plagiarism framework and EW independently reviewed 10% of the included articles. AR drafted the manuscript. TY, PG and EW critically engaged with the manuscript and provided input. All authors approved the final version of the manuscript.

Funding All authors are supported by the Effective Health Care Research Consortium. This Consortium is funded by UK aid from the UK Government for the benefit of developing countries (grant: 5242).

Disclaimer The views expressed in this publication do not necessarily reflect UK government policy.

Competing interests PG is part of Cochrane, an organisation that routinely uses Turnitin. EW has given workshops on plagiarism and spoken at a conference funded by Turnitin.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Unpublished data from the study are available upon request from AR.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Helgesson G, Eriksson S. Plagiarism in research. *Med Health Care Philos* 2015;18:91–101.
- Steneck N. ORI Introduction to the responsible conduct of research. 2007. <https://ori.hhs.gov/ori-introduction-responsible-conduct-research>
- Committee on Publication Ethics (COPE). What to do if you suspect plagiarism. 2013 <http://publicationethics.org/files/plagiarism%20B.pdf> (accessed on 12 Jun 2018).
- Wager E. Defining and responding to plagiarism. *Learned Publishing* 2014;27:33–42.
- Roig M. Avoiding plagiarism, self-plagiarism, and other questionable writing practices: a guide to ethical writing, 2006. <https://ori.hhs.gov/sites/default/files/plagiarism.pdf> (accessed 10 Oct 2017).
- World Association of Medical Editors (WAME). Recommendations on publication ethics policies for medical journals: WAME. 2017. <http://www.wame.org/about/recommendations-on-publication-ethics-police> (accessed 8 Aug 2017).
- Committee on Publication Ethics (COPE). Code of conduct and best practice guidelines for journal editors. 2011. http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf (accessed on 2 Feb 2017).
- African Journal Partnership Project (AJJP). 2016. <https://ajpp-online.org/> (accessed on 22 Sep 2017).
- Pupovac V, Fanelli D. Scientists admitting to plagiarism: a meta-analysis of surveys. *Sci Eng Ethics* 2015;21:1331–52.
- Rohwer A, Young T, Wager E, et al. Authorship, plagiarism and conflict of interest: views and practices from low/middle-income country health researchers. *BMJ Open* 2017;7:e018467.
- Fisher RJ, Katz JE. Social-desirability bias and the validity of self-reported values. *Psychol Mark* 2000;17:105–20.
- Baždarić K, Bilić-Zulle L, Brumini G, et al. Prevalence of plagiarism in recent submissions to the Croatian Medical Journal. *Sci Eng Ethics* 2012;18:223–39.
- Higgins JR, Lin FC, Evans JP. Plagiarism in submitted manuscripts: incidence, characteristics and optimization of screening-case study in a major specialty medical journal. *Res Integr Peer Rev* 2016;1:13.
- Zhang HY. CrossCheck: an effective tool for detecting plagiarism. *Learned Publishing* 2010;23:9–14.
- Jamali R, Ghazinoory S, Sadeghi M. Plagiarism and ethics of knowledge. *Journal of Information Ethics* 2014;23:101–10.
- African Journals Online (AJOL). <https://www.ajol.info/> (accessed on 9 Aug 2018).
- World Bank. World Bank list of economies. 2016 [Siteresources.worldbank.org/DATAS/STATISTICS/Resources/CLASS.XLS](http://www.worldbank.org/DATAS/STATISTICS/Resources/CLASS.XLS) (accessed on 18 Jan 2018).
- Turnitin. Turnitin Instructor QuickStart guide. https://guides.turnitin.com/01_Manuals_and_Guides/Instructor_Guides/01_Instructor_QuickStart_Guide#Evaluating_Originality_Reports (accessed 14 Feb 2017).
- IBM SPSS Statistics for Windows [program]. 25 version.* Armonk, NY, 2017.
- Taylor DB. Plagiarism in manuscripts submitted to the AJR: development of an optimal screening algorithm and management pathways. *AJR Am J Roentgenol* 2017;209:W56.
- Kleinert S, Wager E. Responsible research publication: international standards for editors. A position statement developed at the 2nd World Conference on Research Integrity, Singapore. In: Mayer T, Steneck N, eds. *Promoting research integrity in a global environment*. Singapore: Imperial College Press/World Scientific Publishing, 2010:317–28.
- Kleinert S. Checking for plagiarism, duplicate publication, and text recycling. *Lancet* 2011;377:281–2.
- Gasparyan AY, Nurmashv B, Seksenbayev B, et al. Plagiarism in the context of education and evolving detection strategies. *J Korean Med Sci* 2017;32:1220–7.
- Roig M. Plagiarism: consider the context (letter to the editor). *Science* 2009;325:813–4.
- Siegfried N, Busgeeth K, Certain E. Scope and geographical distribution of African medical journals active in 2005. *S Afr Med J* 2006;96:533–8.
- Goehl TJ, Flanagan A. Enhancing the quality and visibility of African medical and health journals. *Environ Health Perspect* 2008;116:A514–5.
- Makoni M. Network seeks to lift African research integrity: nature index. 2018 <https://www.natureindex.com/news-blog/network-seeks-to-lift-african-research-integrity> (accessed on 3 Sep 2018).