RetractoBot: a protocol for a randomised controlled trial to assess the impact of notifying authors that they have cited a retracted paper

Karolina Wartolowska, Francis Irving, Nicholas DeVito, Benjamin Feakins, Helen Curtis, Seb Bacon, Liam Smeeth, Carl Heneghan, Ben Goldacre

Structured summary

Background: There is a problem with the visibility of retractions, and many retracted papers continue to be cited as if they were still valid. It has been suggested that authors citing retracted papers should be contacted about it, but this has been deemed too challenging and has never been attempted.

Design: This is a randomised controlled trial.

Methods: All eligible retracted papers will be randomised either to the intervention (an email notifying authors of the citing papers about the fact that they have cited a retracted publication) or the control group (no notification email).

Outcomes: The primary outcome will be the rate of citation of retracted papers during 12 months follow-up. The secondary outcomes will be a 24-month follow-up and the qualitative analysis of feedback from the authors in response to the intervention email.

Conclusions: Our trial will investigate whether the number of citations of retracted papers can be reduced by notifying researchers that they cited a retracted paper.

Background and rationale

Retractions are intended to warn readers of the scientific literature about serious errors or misconduct and to reduce future citations of unreliable papers [1,2]. The number of retractions is increasing [2,3] faster than the number of publications per year [4]. This is most likely not because the quality of research is getting worse, but rather because of more journals issue retractions [56].

Although new citations tend to decline after a paper has been retracted [7] they persist and many retracted papers continue to be cited [8]. Some citing papers specifically refer to the fact of the retraction - for example, to analyse study flaws - and these "negative" citations may help to limit accidental "positive" citations of retracted papers [9,10], i.e. as if the paper were still valid. However, many citing papers still refer to the retracted papers as if their results were valid, propagating these results even further [11,8]. This is, partly because journals do not sufficiently notify the readers[11] and, at partly, because many researchers are not aware of retractions and often cite papers saved in their personal library [1]. It has been suggested that authors citing retracted papers should be contacted about it [12,13] but this was deemed too challenging and has never been attempted.

We have now developed a system that can automatically send an email to all contactable authors of papers that cite retracted papers to inform them that they have done so. We are now implementing this service in the context of a randomised controlled trial. Please see also a related blog by Retraction Watch advising researchers on what to do if they have cited a retracted paper [14]. The primary aim of the trial is to investigate whether the number of new citations of retracted papers is reduced when existing citing authors are sent a notification email; in comparison with a non-intervention group of retracted papers where no email notices are sent to the authors citing them. The secondary aims are to characterise the responses of the academic community to the email alerts; to report the feasibility of this service, and to describe the barriers we faced.

Methods

Trial design and settings

This will be a randomised controlled trial with a 1:1 randomisation ratio at the retracted paper level. The trial settings will be electronic publication records of the biomedical literature. The trial-related data will include details of retracted papers, details of papers that cited these retracted papers, and contact details of authors of the citing papers. The data will be collected using a multi-stage process of identifying retraction notices, finding the retracted paper to which they refer, finding articles which have cited them and email addresses of the citing papers' authors.

Data collection

We will search the PubMed and the Scopus database from their beginning to the date of the initial data scrape. We will use the PubMed ESearch Application Programming Interface (API) to identify all retraction notices by using the term "*Publication Type*"="*Retraction of Publication*". For each notice, we will use PubMed EFetch API to find the retracted papers listed in the *CommentsCorrections* element with the attribute *RefType=RetractionOf*. The above process will include withdrawn papers as PubMed does not differentiate between a retraction and a withdrawal [15]. The list of tools and abbreviations in in Appendix 5.

For each retracted paper and retraction notice, we will use the PubMed EFetch API to collect the following data from the PubMed: the PubMed identifier (PMID), the Digital Object Identifier (DOI), paper title, publication type, journal ISO abbreviation, journal title, the "article date" (date of electronic publication, *Article/ArticleDate),* and, because the article date is

sometimes missing we will also collect the "journal date" (of journal publication, *JournalIssue/PubDate).* As some publication dates give only a "year" or "month and year", we will also generate a derived variable describing the granularity of the journal publication date.

As there is no complete and openly accessible database on citation data, we will find all papers which have cited each retracted paper using the commercial Scopus APIs from Elsevier. First, we will use the Scopus Search API to find the Scopus identifier of the retracted paper, searching initially by the PMID and, if that fails, by the DOI. To get a full list of citing paper Scopus IDs, we will then perform a REFEID search for each retracted paper using the Scopus Search API.

For each citing paper, we will use the Scopus Abstract Retrieval API to collect its PMID, DOI, title, journal ISSN, journal title, the "prism cover date" (publication date of the citing paper, *prism:coverDate*) and details of all authors (given name, surname, and email address), where these are given.

For citing papers with no PMID given in Scopus, we will attempt to find its PMID by searching for DOI using the PubMed ESearch API. Then, for all citing papers with a PMID, we will use the PubMed EFetch API to retrieve its publication type, the date of electronic publication, the date of journal publication, and the granularity of the journal publication date (generated from the date field, describing whether it gives a year, year-month, or year-month-day).

For retracted and citing paper, if their publication date is given as a quarter or as "Spring" or "Winter", the date will be logged as the 1st January of the year the paper was published.

Eligibility criteria

We will include all eligible retracted papers mentioned in retraction notices listed in *PubMed* on the date of the initial database search. For each retracted paper, we will include all citing papers found in Scopus, provided that they contain the authors' email address, so that the authors can be contacted.

Retracted papers will be excluded if:
- their retraction notice is not in the PubMed database
- their retraction notice does not have the correct attribute "Retraction of Publication" in the PubMed database

3

- their retraction notice is not linked to a retracted paper in the PubMed database
- their retraction notice refers to multiple retracted papers (because these are too complex for our software to parse within budget)
- their retraction notice doesn't have a structured date (either *article date* or *journal date)*
- the retracted paper is not listed in the Scopus database or is not listed with a correct PMID or DOI (and therefore citation data cannot be determined by our software)
- the retracted paper does not have a structured date (either *journal date or article date)*
- the retracted papers are not associated with any citing papers in Scopus at the trial commencement date
- the retracted paper has no citations other than retraction notices, i.e., those having the PubMed article type = "Retraction of Publication" or "Published Erratum", or the title on Scopus commences with "retract*" (and is therefore highly likely to be a retraction notice)
- the retracted paper has no citations other the citing papers that always cause the Scopus API to "time out" due to a technical error in the Scopus system (this happens to 6 out of over 100,000 specific citing papers in our software piloting); therefore, no citing papers can be retrieved for this retracted publication
- the retracted paper has no citations with email addresses in Scopus for any of its authors. If none of the citing papers have any email addresses for any of its authors and the only citing publication is the retraction notice – such retracted paper will be excluded from the analysis.
- any retracted paper is excluded if all its citing papers are excluded at the trial commencement date.

Intervention

All eligible retracted papers will be randomised into two groups. Retracted papers in the intervention group will have a single standardised email sent to all contactable authors (with an email in the Scopus API) listed on any citing paper. The email will inform the authors of the fact that they have cited a retracted paper. Each notification for each author will be sent to every email we have obtained from the Scopus Abstract Retrieval API for that Scopus Author Identifier. Notifications for multiple papers by a single author who has cited a single retracted paper will be compiled into a single email and the author will be asked about the most recent citing paper (see Appendices 3 and 4).

The text of the email has been designed collaboratively and tested on colleagues within our research group. The text of the email will be as per the example in Box 1; full unpopulated text and permutations (if an author cited a retracted paper after it had been retracted, cited several retracted papers or cited a retracted paper several times) are presented in the appendices. If an author cited several retracted papers, they will receive a separate notification email for each retracted paper.

When comparing papers to see which is the most recent, for example when a researcher cited a retracted paper in multiple publications, the article date will be used, or if it is missing, the journal date will be used, and if these dates are missing for citing papers, the prism cover date will be used. If a retracted paper has multiple notices of its retraction, the date from the one with the lowest PMID in the subject heading will be used. For display purposes in the email the year from the journal date for retracted papers or retraction notices will be used, or if it is missing, the year from the article date will be used. For citing papers, the year from the journal date, or if it is missing from the prism cover date, is used.

Box 1. An example notification email text if the citation was before the retraction

Subject: You cited a retracted paper in your *Journal of Interesting Research* paper published in *2000.*

Dear *John Smith*,

We're writing to let you know that your paper *"Ontogenesis as a recapitulation of phylogenesis"* (*Journal of Interesting Research*, *2000*) cites a paper which has since been retracted. You cited "*Meta-analysis of previous important research"* *(Annals of Science, 1990*), which was retracted in *2010* [link to retraction notice].

Was this information useful?

We're sending this mail as part of the RetractoBot research project. If you have a moment, we'd appreciate you making a single click to let us know if you found our email useful. Your click is taken as consent for your response to be included in our analysis.

I **didn't know** that paper was retracted, thanks!
I **already knew** this paper was retracted, thanks!
PubMed records are incorrect. **This paper has not been retracted**.

If you have any other comments, please reply to this email; your reply is taken as consent for your comments to be used in our qualitative analysis of the project.

Many thanks for your time.

Yours sincerely,

Dr Ben Goldacre and team

EBM DataLab, Centre for Evidence-Based Medicine
Nuffield Department of Primary Care Health Sciences, University of Oxford
Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

In accordance with the European Union General Data Protection Regulation 2016, we would like to inform you of the following information. We are using publicly accessible bibliographic information from the PubMed and Scopus databases. We are processing only your name and email address

associated with your Scopus Author ID, which we obtained from Scopus only to send you this message. If you would like to stop receiving emails from RetractoBot at this email address, choose the 'unsubscribe' link below. If you would like to correct your data on PubMed or Scopus, please contact those organisations directly.

1
2

7

We will send intervention emails in daily batches, to allow iteration and improvement of the text if required, depending on responses received. We will send emails to 100 papers on the first day, quickly review any replies, and then rapidly increase the number sent each day, aiming to have all emails sent within a 2-week period. We will use an automated service (Mailgun), which will automatically send emails, and track the first time the emails were opened and the last time the emails were interacted with (one of the options clicked).

Control

Retracted papers randomised to the control group will have no notification emails sent out to their citing papers' authors.

Outcomes

Primary outcome

The primary outcome will be the rate of citation of retracted papers during a 12-month follow-up. We have not included a "washout period", because it would mean missing all the changes happening after the start of the trial.

A 12-month follow-up period was chosen for three main reasons:
- it has to be specified in whole years, because many journals publishing annually or quarterly, leading to citations data being unevenly distributed over months within the year.
- a longer follow-up period may increase the risks of changes to the tools used by researchers in such a way that the citation of retracted papers is substantially reduced. For example: electronic journals or their databases, including PubMed and Scopus, may modify their contents or presentation of information about retracted papers; or additional features might be added to reference manager software to warn users that they are citing retracted papers.
- it is not feasible for us to formally commit to a study with a very long follow-up due to funding limitations.

Secondary outcomes

The secondary outcomes will be a 24-month follow-up and the qualitative analysis of feedback from the authors(in the intervention group only) in response to the intervention email. We will analyse authors' responses to the intervention emails including: the proportion of emails sent that are returned as undeliverable (if we have multiple email addresses for a single author, the email is sent to all email addresses, and we count that communication as

"undeliverable" if all emails to all multiple email addresses are returned as "undeliverable"); the proportion of authors opening any email (if we have multiple email addresses for a single author, and the email is sent to all email addresses, we count that communication as "opened" if any of the multiple emails is "opened"); the proportion of authors clicking each option in response to our "is this useful did you know" questions; and the preceding broken down by citing article's publication year. We will do qualitative research extracting themes in free text replies if any.

Power calculation

The power calculation is based on a preliminary collection of data on retracted and citing papers performed in January 2018. The scrape ran on the data from the beginning of the PubMed and Scopus databases to January 2018 and it returned 4,045 unique retracted papers and 122,372 citing papers. We calculated that in 2016 and 2017 the median number of citations per year for a retracted paper was 7 (IQR 3-16) and 7 (IQR 1-17) respectively. We, therefore, estimate, using a normal approximation to Poisson rates, that 2,000 retracted papers (1,000 per randomisation group) will give better than 80% power to detect an absolute change of 0.4 citations per year during 12 months of follow-up time.

Potential limitations

The key limitation is contamination. Our objective is to raise awareness of the fact that a paper has been retracted. Individual authors may cite retracted papers from both the intervention and control groups: those authors receiving a notification about a retracted paper in our intervention group may also become interested in whether they have cited any further retracted papers; they may research the matter independently; they may discover that they have also cited retracted papers on which they have not received a notice from us; and so the retracted papers in the control group may also see a reduction in subsequent citations. Any such contamination will reduce the effect size, nonetheless, this study should be sufficiently powered to detect an effect of the intervention.

Follow-up

The citation rate will be assessed over a 12-month period after the last batch of emails is sent. The secondary outcomes will be a 24-month follow-up.

Methods used to generate and implement the random allocation sequence

A computer-generated randomisation sequence created at the retracted papers' level will be applied to the list of authors citing retracted papers included in this trial. Emails will be sent only to authors allocated to the intervention group. As this process will be automated and

performed using pre-specified scripts written before trial initiation; therefore, there will be no risk of allocation bias.

Data extraction and management

Data on retracted papers and citing papers will be gathered by repeating the process described in the "Data collection" section above. The date the email was sent will be recorded in a PostgreSQL database. We will use the service Mailgun to send emails and will call its API to get dates for delivery and opening of each mail.

A Django web application will record the authors' responses to indicate whether the author cited the retracted paper knowing it was retracted or not: this will be saved in a PostgreSQL database. We will also record free text responses if the contacted authors will provide any. All author-level data will be anonymised.

Statistical methods used to compare groups for primary and secondary outcomes

The primary outcome will be analysed using adjusted and unadjusted regression models for count/rate data (Figure 1). We will appraise the significance of the publication date of the retracted paper and its citation rate prior to the trial start date in count rate-based models; these will be included in the final model for the primary outcome if they are significant.

Figure 1. Control flow diagram for model selection
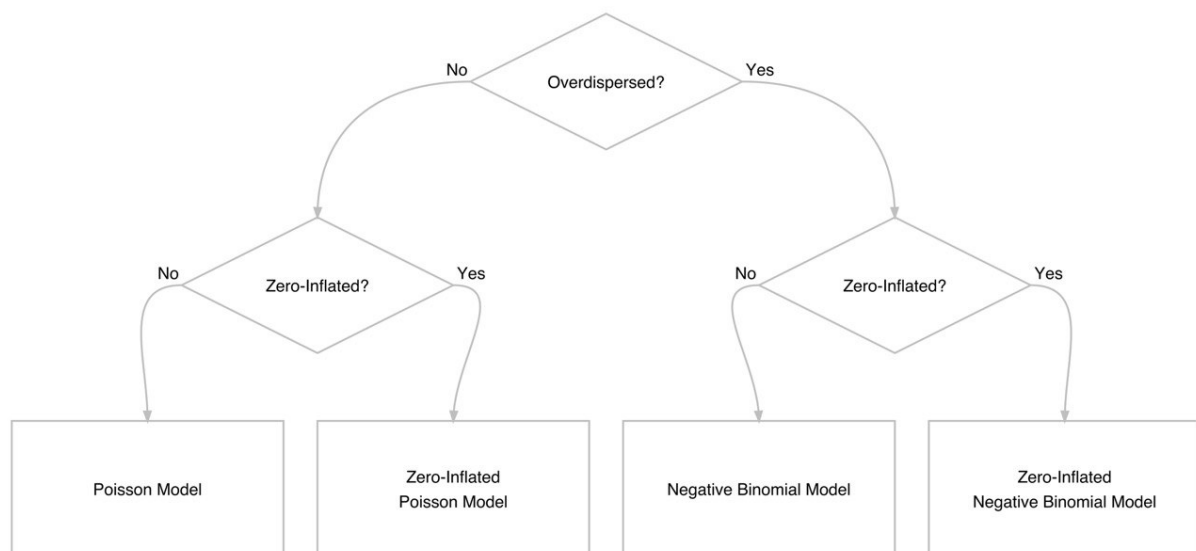
Sensitivity analyses

We do not know, before trial commencement, what proportion of emails will be undeliverable. We will devise an appropriate sensitivity analysis to work around uncontactable authors when this information becomes available after the emails are sent, but before the follow-up period has passed.

One issue for the trial is a potential contamination, especially in situations where multiple papers by the same author or team are retracted, as citing paper authors may receive notifications about some but not all of that team's retracted work, and may, in turn, be triggered to explore and discover retractions of papers that we did not notify them about due to their being in our control group of retracted papers. This contamination would lead to an attenuation of any apparent impact of our service. We have considered various possible mechanisms to eradicate this source of contamination through study design but found no adequate method that does not also dramatically reduce the number of retracted and citing papers leaving only an unrepresentative sample. However, we will conduct a sensitivity analysis, in which we analyse only citation rates for papers where the situation never arose that a citing author received emails about less than all the retracted papers they cited.

Qualitative and quantitative analysis of replies to emails.

Responses to the emails will be tabulated and analysed. This will include calculating the proportion of authors opening the email, the proportion of authors clicking each option in response to our "is this useful? did you know?" question and qualitative research extracting themes in free text replies to our emails (if any). Mailgun will provide data on emails sent, emails delivered (i.e. not bounced), emails opened and the time they were first opened, links clicked and the time they were last clicked. We will give basic descriptive statistics on n% emails sent, delivered, opened, replied to.

An open reply email address will be provided in our emails (retractions@ebmdatalab.net), which forwards to a Doorbell.io account containing four team members (BG, FI, KW, ND).

Replies will be monitored for error reports requiring action and queries requiring a response. They will be also used for a qualitative analysis of responses.

All reply emails will be coded for tone and content. The tone of the emails will be categorised as "Positive," "Negative," "Neutral," or "Other." After initial categorisation, further sub-themes

1 can be investigated if needed. Coding for content will reflect more specific feedback and

2 comments beyond the tone of the responses. Indications, why the respondent felt

3 particularly positive or negative regarding the intervention, will arise and categories will be

4 formed as needed from the data as it is received.

5

6 All content coding will be done initially by a single coder, including creating the initial

7 categorisation. This coder will then finalise the codebook when the trial is completed.

8 Depending on the volume of emails received, a second coder will then use the codebook

9 developed by the first coder to dual-code either all or a sub-sample of the emails for tone

10 and content. The extent of the dual-coding and subsequent assessment of inter-rater

11 reliability will depend on the volume and character of free-text responses received, if any.

12 Qualitative coding of replies will continue until saturation of themes or for a maximum of

13 three months from the first email sent.

14 RetractoBot service

15 Alongside this trial, we will develop an ongoing RetractoBot service. This application will

16 check PubMed for newly retracted papers each week and then search on Scopus for

17 publications citing these retracted papers as described in the "Data collection" above. It will

18 then email authors of any citing papers as described in the "Intervention" section. This will

19 happen on an ongoing basis, with the intervention applied to all authors of any new citing

20 papers as they appear. This service will start at the same time as the trial, but only the

21 retracted papers identified during the original scrape at the beginning of the trials will be

22 included in the follow-up analyses; any new retracted paper and its citing papers will be

23 included in the RetractoBot service. The service will send emails to authors citing newly

24 retracted papers and new citing papers for previously retracted papers. Retracted papers in

25 the control arm of the trial will be excluded from the service, and no notification emails will be

26 sent regarding these retracted papers to the citing papers' authors. The code for the service

27 will be available from https://github.com/ebmdatalab/retractobot.

28 Ethics

29 We are using only publicly accessible data to contact authors of citing papers regarding

30 these papers. We contacted the University of Oxford Medical Sciences Interdivisional

31 Research Ethics Committee with a summary of the notifications service and the randomised

32 controlled trial and were advised that no ethics approval was required [link].

33

34 Trial steering committee

A Trial Steering Committee will be convened to provide overall supervision of the trial and ensure it is in accordance with the principles of good clinical practice and the relevant regulations.

Data monitoring committee

We did not consider a separate data monitoring would have been useful as this study has no substantial risks and no early termination rules. Any data monitoring issues will be incorporated into the terms of reference of the Trial Steering Committee.

Dissemination

The randomised controlled trial will be published in a peer-reviewed journal. A website will be launched at retracted.net describing the project, linking to the randomised controlled trial protocol, and giving basic summary statistics on the number of emails sent and the number of papers notified on. The site will also report the results once the trial is completed. Depending on resources available, we will also aim to produce a service where any user can give their email or author details and be told about retracted papers they have cited.

Conclusions

This trial will investigate whether notifying authors about retracted papers they have previously cited reduces subsequent citations. Its findings will inform the development of the RetractoBot service.

Acknowledgements

Contributorship

BG conceived the study. BG and KW designed the methods with input from FI, ND, SB, BF. KW drafted the manuscript with input from BG, FI, BF, HC, SB, ND, CH, and LS. All authors contributed to and approved the final manuscript. FI was lead engineer on the software with input from BG and KW. BG supervised the project and is a guarantor.

Conflicts of Interest

Funding

References

1. Davis, P. M. The persistence of error: a study of retracted articles on the Internet and in personal libraries. *J. Med. Libr. Assoc.* **100,** 184–189 (2012).

2. Wager, E. & Williams, P. Why and how do journals retract articles? An analysis of Medline retractions 1988-2008. *J. Med. Ethics* **37,** 567–570 (2011).

3. Grieneisen, M. L. & Zhang, M. A comprehensive survey of retracted articles from the scholarly literature. *PLoS One* **7,** e44118 (2012).

4. He, T. Retraction of global scientific publications from 2001 to 2010. *Scientometrics* **96,** 555–561 (2012).

5. Fanelli, D. Why growing retractions are (mostly) a good sign. *PLoS Medicine* **10,** e1001563 (2013).

6. Steen, R. G. Retractions in the scientific literature: is the incidence of research fraud increasing? *J. Med. Ethics* **37,** 249–253 (2011).

7. Furman, J. L., Jensen, K. & Murray, F. Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine. *Res. Policy* **41,** 276–290

1   (2012).

2   8.   Budd, J. M., Sievert, M., Schultz, T. R. & Scoville, C. Effects of article retraction on

3   citation and practice in medicine. *Bull. Med. Libr. Assoc.* **87,** 437–443 (1999).

4   9.   Catalini, C., Lacetera, N. & Oettl, A. The incidence and role of negative citations in

5   science. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 13823–13826 (2015).

6   10. Korpela, K. M. How long does it take for the scientific literature to purge itself of

7   fraudulent material?: the Breuning case revisited. *Curr. Med. Res. Opin.* **26,** 843–847

8   (2010).

9   11. van der Vet, P. E. & Nijveen, H. Propagation of errors in citation networks: a study

10   involving the entire citation network of a widely cited paper published in, and later

11   retracted from, the journal Nature. *Research Integrity and Peer Review* **1,** 3 (2016).

12   12. Sox, H. C. & Rennie, D. Research misconduct, retraction, and cleansing the medical

13   literature: lessons from the Poehlman case. *Ann. Intern. Med.* **144,** 609–613 (2006).

14   13. da Silva, J. A. T. & Bornemann-Cimenti, H. Why do some retracted papers continue to

15   be cited? *Scientometrics* **110,** 365–370 (2017).

16   14. McCook, A. What should you do if a paper you've cited is later retracted? *Retraction*

17   *Watch* (2016). Available at: http://retractionwatch.com/2016/11/01/what-should-you-do-

18   if-a-paper-youve-cited-is-later-retracted/. (Accessed: 26th October 2018)

19   15. Fact Sheet: Errata, Retractions, and Other Linked Citations in PubMed. (2002).

20   Available at: https://www.nlm.nih.gov/pubs/factsheets/errata.html. (Accessed: 16th

21   January 2018)

22

23

Appendix 1. Notification email text if the citation was before the retraction

Subject: You cited a retracted paper in your <citing paper's journal> paper published in < citing paper's publication year>.

Dear <first name> <surname>,

We're writing to let you know that your paper <citing paper's title> (<citing paper's journal>, <citing paper's publication year>) cites a paper which has since been retracted. You cited <retracted paper's title> (<retracted paper's journal>, <retracted paper's publication year>), which was retracted in <retraction notice's date> [link to a retraction notice].

Was this information useful?
We're sending this mail as part of the RetractoBot research project. If you have a moment, we'd appreciate you making a single click to let us know if you found our email useful. Your click is taken as consent for your response to be included in our analysis.

I **didn't know** that paper was retracted, thanks!
I **already knew** this paper was retracted, thanks!
PubMed records are incorrect. **This paper has not been retracted**.

If you have any other comments, please reply to this email; your reply is taken as consent for your comments to be used in our qualitative analysis of the project.

Many thanks for your time.

Yours sincerely,
Dr Ben Goldacre and team

EBM DataLab, Centre for Evidence-Based Medicine
Nuffield Department of Primary Care Health Sciences, University of Oxford
Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

In accordance with the European Union General Data Protection Regulation 2016, we would like to inform you of the following information. We are using publicly accessible bibliographic information from the PubMed and Scopus databases. We are processing only your name and email address associated with your Scopus Author ID, which we obtained from Scopus only to send you this

1

2

message. *If you would like to stop receiving emails from RetractoBot at this email address, choose the 'unsubscribe' link below. If you would like to correct your data on PubMed or Scopus, please contact those organisations directly.*

17

Appendix 2. Notification email text if the citation was after the retraction

Subject: You cited a retracted paper in your <citing paper's journal> paper published in < citing paper's publication year>.

Dear <first name> <surname>,

We're writing to let you know that your paper <citing paper's title> (<citing paper's journal>, <citing paper's publication year>) cites a paper which had previously been retracted. You cited <retracted paper's title> (<retracted paper's journal>, <retracted paper's publication year>), which was retracted in <retraction notice's date> [link to retraction notice].

Was this information useful?

We're sending this mail as part of the RetractoBot research project. If you have a moment, we'd appreciate you making a single click to let us know if you found our email useful. Your click is taken as consent for your response to be included in our analysis.

I **didn't know** that paper was retracted, thanks!

I **already knew** this paper was retracted before you emailed me, but **I didn't know** when I cited it.

I **already knew** this paper was retracted, and I **explicitly described this as a retracted paper** when I cited it, thanks!

PubMed records are incorrect. **This paper has not been retracted**.

If you have any other comments, please reply to this email; your reply is taken as consent for your comments to be used in our qualitative analysis of the project.

Many thanks for your time.

Yours sincerely,

Dr Ben Goldacre and team

EBM DataLab, Centre for Evidence-Based Medicine

Nuffield Department of Primary Care Health Sciences, University of Oxford

1
2
3

19

Appendix 3. Notification email text for multiple citing papers published before the retraction.

Subject: You cited a retracted paper in your <most recent citing paper's journal> paper published in <most recent citing paper's publication year>

Dear <first name> <surname>,

We're writing to let you know that you have cited a paper which has since been retracted in the following of your own publications:

- <citing paper 1 title> (<citing paper 1 journal>, <citing paper 1 publication year>);
- <citing paper 2 title> (<citing paper 2 journal>, <citing paper 2 publication year>);
- <citing paper 3 title> (<citing paper 3 journal>, <citing paper 3 publication year>)

These papers cited <retracted paper's title> (<retracted paper's journal>, <retracted paper's publication year>), which <u>was retracted in <retraction notice's date> [link to retraction notice]</u>.

Was this information useful?
We're sending this mail as part of the RetractoBot research project. If you have a moment, we'd appreciate you making a single click to let us know if you found our email useful. Your click is taken as consent for your response to be included in our analysis. We are interested in your most recent citation of this retracted paper:

<u>I **didn't know** that paper was retracted, thanks!</u>
<u>I **already knew** this paper was retracted, thanks!</u>
<u>PubMed records are incorrect. **This paper has not been retracted**.</u>

If you have any other comments, please reply to this email; your reply is taken as consent for your comments to be used in our qualitative analysis of the project.

Many thanks for your time.

Yours sincerely,

Dr Ben Goldacre and team

EBM DataLab, Centre for Evidence-Based Medicine
Nuffield Department of Primary Care Health Sciences, University of Oxford
Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

In accordance with the European Union General Data Protection Regulation 2016, we would like to inform you of the following information. We are using publicly accessible bibliographic information from the PubMed and Scopus databases. We are processing only your name and email address associated with your Scopus Author ID, which we obtained from Scopus only to send you this message. *If you would like to stop receiving emails from RetractoBot at this email address, choose the 'unsubscribe' link below. If you would like to correct your data on PubMed or Scopus, please contact those organisations directly.*

1
2
3

21

Appendix 4. Notification email text for multiple citing papers where one or more of them was published after the retraction.

Subject: You cited a retracted paper in your <most recent citing paper's journal> paper published in <most recent citing paper's publication year>.

Dear <first name> <surname>,

We're writing to let you know that you have cited a paper which has been retracted in the following of your own publications:

- <citing paper 1 title> (<citing paper 1 journal>, <citing paper 1 publication year>);
- <citing paper 2 title> (<citing paper 2 journal>, <citing paper 2 publication year>);
- <citing paper 3 title> (<citing paper 3 journal>, <citing paper 3 publication year>)

You cited <retracted paper's title> (<retracted paper's journal>, <retracted paper's publication year>), which was retracted in <retraction notice's date> [link to retraction notice].

Was this information useful?
We're sending this mail as part of the RetractoBot research project. If you have a moment, we'd appreciate you making a single click to let us know if you found our email useful. Your click is taken as consent for your response to be included in our analysis. We are interested in your most recent citation of this retracted paper:

I **didn't know** that paper was retracted, thanks!
I **already knew** this paper was retracted before you emailed me, but **I didn't know** when I cited it.
I **already knew** this paper was retracted, and I **explicitly described this as a retracted paper** when I cited it, thanks!
PubMed records are incorrect. **This paper has not been retracted**.

If you have any other comments, please reply to this email; your reply is taken as consent for your comments to be used in our qualitative analysis of the project.

Many thanks for your time.

Yours sincerely,
Dr Ben Goldacre and team

EBM DataLab, Centre for Evidence-Based Medicine
Nuffield Department of Primary Care Health Sciences, University of Oxford
Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG

In accordance with the European Union General Data Protection Regulation 2016, we would like to inform you of the following information. We are using publicly accessible bibliographic information from the PubMed and Scopus databases. We are processing only your name and email address associated with your Scopus Author ID, which we obtained from Scopus only to send you this message. If you would like to stop receiving emails from RetractoBot at this email address, choose the 'unsubscribe' link below. If you would like to correct your data on PubMed or Scopus, please contact those organisations directly.

1
2
3

23

Appendix 5 Explanation of abbreviations and tools used in the RetractoBot project

API - An application programming interface (API) is a particular set of rules and specifications that software programs can follow to communicate with each other. It serves as an interface between different software programs and facilitates their interaction, similar to the way the user interface facilitates interaction between humans and computers.

Django - Django is a high-level Python framework for web development (https://www.djangoproject.com/)

DOI - Digital Object Identifier (https://www.doi.org/)

Doorbell.io - is an online application to collect feedback from users (https://doorbell.io/)

IQR - Interquartile Range is a measure of statistical dispersion.

ISO - International Organization for Standardisation is an international standard-setting body composed of representatives from various national standards organizations.

ISSN - International Standard Serial Number (ISSN) is an eight-digit serial number used to uniquely identify a serial publication

PMID - PubMed identifier or PubMed unique identifier is a unique number given to each PubMed record

PostgreSQL - is an object-relational database management system.

PubMed - PubMed is a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics

PubMed EFetch API and PubMed ESearch API are applications consisting of a series of calls to retrieve records from the database searched by PubMed (https://www.ncbi.nlm.nih.gov/books/NBK25499/#chapter4.ESearch)

24

1   RCT - randomised controlled trial

2   REFEID - reference identificator

3

4

5

6

7

8

9

10