DZHW

Deutsches Zentrum für
Hochschul- und Wissenschaftsforschung

An in-depth Analysis to the Annotation of Retractions in PubMed and Web of Science

Amsterdam, May 2017
Marion Schmidt

5th World Conference on Research Integrity

Contents

Introduction

Definitions and Conceptual Clarifications

Methods

Empirical Results

Conclusions

Introduction

Functions of Retractions:

- Correcting the publication record
- (Evidently) having a penalizing and dissuasive function
- Empirical basis for research on scientific misconduct

Comparability of bibliometric studies on (e.g., prevalence and effects of) retractions seems to be impeded to a certain extent due to:

- No real standards in corpus delineation (partly insufficient description), different sample strategies
- reliability of different databases?

Research Question:

How transparent and reliable is the annotation of retractions in literature databases?

Concepts

Retraction (COPE):

Journal editors should consider retracting a publication if:

- ✓ they have clear evidence that the findings are unreliable, either as a result of misconduct (e.g. data fabrication) or honest error (e.g. miscalculation or experimental error)
- ✓ the findings have previously been published elsewhere without proper cross-referencing, permission or justification (i.e. cases of redundant publication)
- ✓ it constitutes plagiarism
- ✓ it reports unethical research

Withdrawal (Elsevier):

Only used for ***Articles in Press*** which represent early versions of articles and sometimes contain errors, or may have been accidentally submitted twice. Occasionally, but less frequently, the articles may represent infringements of professional ethical codes, such as multiple submission, bogus claims of authorship, plagiarism, fraudulent use of data.

State of Research: PubMed-based studies

PubMed: The majority of studies is (or has long been) PubMed-based. Regularly PubMed data are taken as they are:

<<Data used in the analysis were acquired from the PubMed database (United States National Library of Medicine, Bethesda, MD). The search was 'limited' to

"Retracted Publication" under the "Type of Article".>>

<<We used PubMed in order to identify retractions from the biomedical literature, which we refer to as "retracting articles",>>

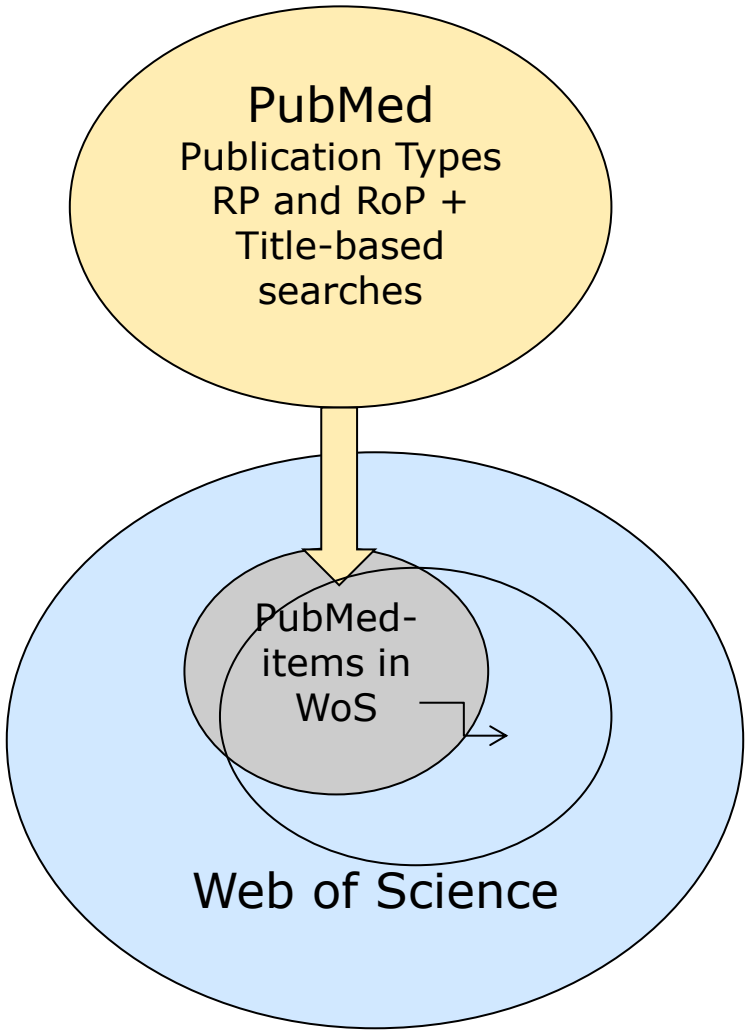
<<All 788 English language research papers retracted from the PubMed database between 2000 and 2010 were evaluated.>>

Focus only on biomedicine.

State of Research: Varying search strategies have been proposed for retractions and retracted publications in WoS

Retraction of publication	Retracted publication	Time window	Study
TI = retraction, DT = Correction or DT = Correction, Addition		1901/1980-2012	Fanelli, 2013
Preselection of highly cited journals, TS =retraction, limit to retraction		1960-2012	Bilbrey, Creamer &O'Dell, 2014
	Preselection of highly cited journals; TI = retract* OR withdraw* OR fraud* OR scientific misconduct OR deception manual screening of results, name searches for authors implicated in falsification	1980-2006	Trikalinos, Evangelou, & Ioannidis, 2008
TI = retract* [Retractions/Retracted Publications] TI = (withdraw* AND (article* or paper* or publication* or manuscript*)) NOT (TI=retract* OR SO="Cochrane Database of Systematic Reviews" [for Withdrawals]	TI = retract* [Retractions/Retracted Publications] TI = (withdraw* AND (article* or paper* or publication* or manuscript*)) NOT (TI=retract* OR SO="Cochrane Database of Systematic Reviews" [for Withdrawals]	1980-2011	(Grieneisen & Zhang, 2012)
TI = "retraction of"	TI = "retracted article", ti = "retraction article"	2000-2009	Lu, Jin, Uzzi, & Jones, 2013
TI = retraction AND vol	TI = retracted AND article Restricted to doctypes Article, Letter, Review	2001-2010	He, 2013
TI="(retraction of vol"	TI = "(retracted article"	2001-2011	Chen, Hu, Milbank & Schultz, 2013
	TI = "retracted article"	[1900?]-2013	Van Leeuwen & Luwel, 2014

Methods (time period 1980-2013)



1. Title-bases searches within and outside the publication type-delineated PubMed corpus, manual validation

2. Matching algorithm

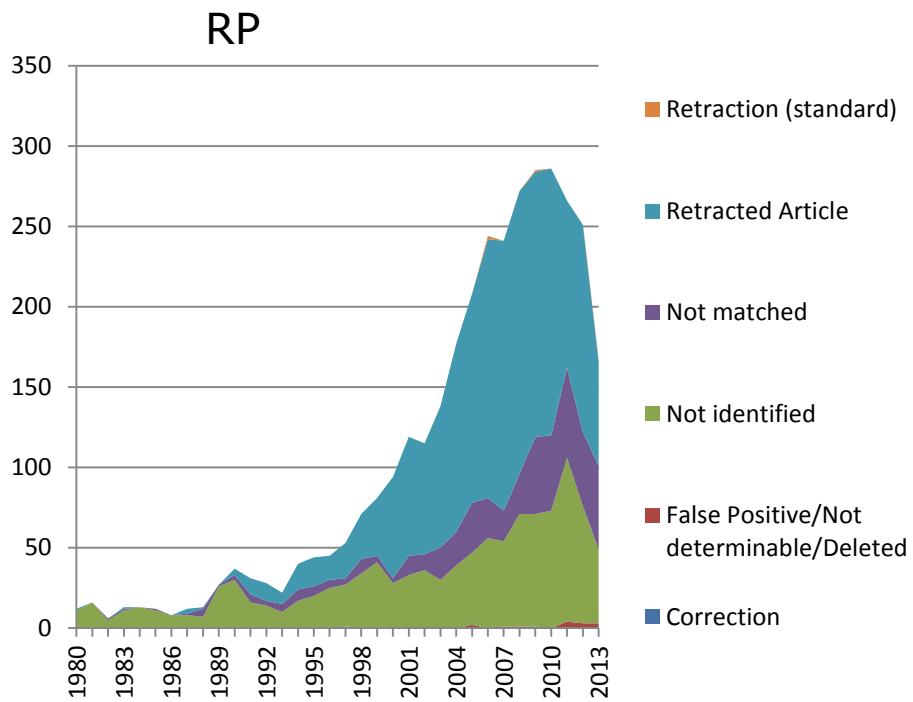
3.
Parsing of WoS titles of PubMed items that have been matched to WoS Generation of Regular Expressions.

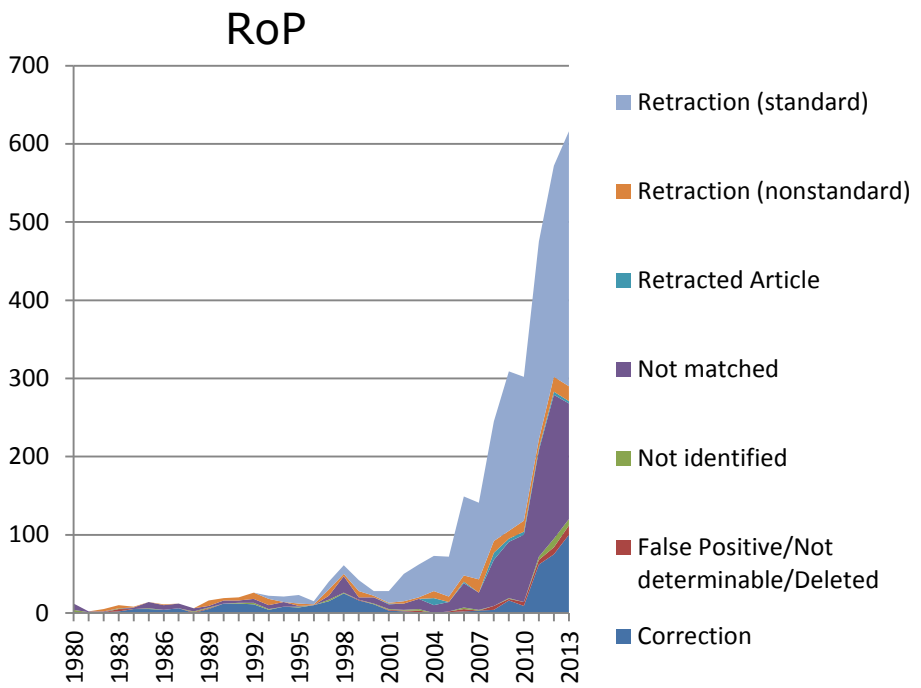
Results (time period 1980-2013)

PubMed, search with Publication Types, 1980-2013
Retraction of Publication: 3226 Items (3505 retraction incidents)
Retracted Publication: 3446 Items
Amount of retracted publications/retractions without publication type: 2-3%
Prevalence of related publication types in PubMed:
Withdrawal:

Search Strategy	Number
(("withdrawn"[Title] OR "withdrawal"[Title]) NOT "Retracted publication"[Publication Type]) NOT "Retraction of publication"[Publication Type];	888
("withdrawn"[Title] OR "withdrawal"[Title]) AND "Retraction of publication"[Publication Type]	70
("withdrawn"[Title] OR "withdrawal"[Title]) AND "Retracted publication"[Publication Type]	12

Results: WoS-Annotation of PubMed Retracted Publications and Retractions of Publications matched to WoS





Summary

- Fuzzy delineation of retractions from related publication types, such as withdrawals, but also from corrections.
- *Massive number of withdrawals*, which are not indexed by PubType in PubMed: A corpus of articles which amounts to 26% of the pubtype-delineated RoP is identifiable as withdrawals.
- *Smaller portions of false negatives* of the RP and RoP publication types.
- *A considerable amount (32%)* of PubMed Retracted Publications matched to WoS are not annotated in WoS.
- *A considerable amount (13%)* of PubMed's Retractions of Publication matched to WoS are only annotated as corrections in WoS.
- Research strategies proposed in the literature for retrieving retractions in WoS differ in the amount of false negatives.

Discussion, open questions

- The formats proposed and used by publishers do not seem to translate loss-free into different databases, therefore some incompleteness especially in WoS should be acknowledged.
- To which degree can we assume that citing authors aware of a retraction when (apart from other resources like social network tools) even Web of Science and PubMed may differ between annotation and non-annotation or divergent annotation?

Thank you for your attention!