

How The Retraction Watch Database Was Constructed

By The Retraction Watch Staff

To build the [Retraction Watch database](#) (DB), we drew upon the experience of over eight years of specifically looking for and at retractions. Our initial resource was PubMed, a publicly available database maintained by the U.S. National Institutes of Health. We scoured through the publication types (e.g., “retracted publication” and “retraction of publication”) and used keywords (e.g., “Retracted:”, “Withdrawn:”, “Retraction Notice”) to identify retracted articles or retraction notices.

Using other article databases (e.g., Clarivate’s [formerly Thomson Reuters’] Web of Science, Elsevier’s Scopus) we cross-checked for retracted articles that we had not yet found, or to verify the status of articles indicated as retracted. The following were also sources from which we learned of retracted articles:

- Public access databases (e.g., JStage, Koreamed, Google Scholar)
- Publisher sites: (e.g., Springerlink, ScienceDirect, Sage, Taylor and Francis)
- Journals: (e.g., Nature, Science, The Journal of Biological Chemistry)
- Institutional reports of research misconduct investigations
- Social media sites
- Tips sent in by blog followers

Finding a retracted article was only the first step. The next step was to elicit the available metadata for each DB input. Entry fields were chosen as the information most commonly found from retracted articles and any associated notices. While the choices for some of the fields should be relatively self-explanatory, the following fields merit a bit of explanation:

- Authors: All authors on the original article were entered, generally in the order they were presented on the html site or pdf.
- Countries: All countries listed in the authors affiliations were entered. In cases where the exact affiliation is not provided, the country entered was the one most commonly associated with the author(s) for the article’s publishing timeframe.
- Subjects: The subjects chosen for entry were selected from a list originally taken from the Retraction Watch blog site. Additional subjects were added as considered necessary.
- Dates: The dates used were the earliest known publication date of the original article or any associated notices. At times, the actual date of the notice was almost impossible to discern. In such cases, the discovery date was used.
- Nature of Notice: Because of a lack of consistency in how journals and publishers title notices, we opted for three: Retraction, Expression of Concern, and Correction. Any article that was stipulated as entirely removed from the publishing site, regardless of time duration, was considered a Retraction. A Correction incorporated any amendment made

to the original text that did not entirely remove the article from the publishing platform, such as Erratum, Corrigendum and “Partial Retraction.” An Expression of Concern encompassed virtually all other notices made by publishers or editors.

- Paywalled: If a paid subscription was required to discern the article title or reason for notice, the notice was considered paywalled.
- Reason(s): Probably the most complex of all the metadata gathered. Rather than simply taking a retraction notice at face value, other resources were checked for background or clarifying information. If there was a related Retraction Watch blog post, that was reviewed; if there was an institutional misconduct finding, that was reviewed as well. Occasionally other media sites (e.g., New York Times, Times Higher Education) have background information. “Reasons” are selected based on the information available at this time of the entry into the DB, but can be (and have been) changed or updated as appropriate if more information is later received.

Maintaining the DB is not a static process. We are fortunate to have very interested and interactive readers who send in retractions they have found. To facilitate such tips, in January of 2018 we also created a “submission form” through which readers could send titles of retracted articles that were not found in the DB. Routine checks of PubMed, Web of Science, publisher and journal websites are still performed to locate new and older retractions recently published online.

When possible, we do hand searches of print journals to locate retractions notices, but those activities are very rare due to limited resources. We also must rely on our readers to bring our attention to “stealth retractions” – those retractions performed by merely removing an article from the Table of Contents or from a journal website without any notice given. We also review entries to ensure that they contain the most current and accurate information.

NB:

Not all articles shown in databases such as PubMed or the Web of Science are actual retractions, and thus will not have been entered into the Retraction Watch DB. Usually these are errata and simply have been incorrectly categorized or indexed. When found, we generally notify the database or journal, depending upon the database’s criteria for corrections to entries. Some entries are corrected immediately while others still are awaiting amendments.

Articles listed as “partially retracted,” or have a retraction notice stipulating only a portion of the article is retracted while the remainder remains intact on the journal site, are considered “corrections” and will be entered into the DB as such. Some of these have yet to be entered since, as corrections, they are not as much of a priority as retractions.