Data Dictionary

**Note:** This data dictionary corresponds to the excel file: "Unsolicited\_Journal\_Emails\_Deidentified\_2022\_v2"

**Additional Note:** Data derived from unsolicited journal emails soliciting manuscripts received by the author in their institutional inbox between July 20, 2021 and January 31, 2022.

# Identifying Information (Columns A through E)

**Record No. (Column A)**

This column includes record numbers for each of the unsolicited emails that was received by the author. These serve as identifiers for each of the emails.

**Sender ID (Column B)**

This column includes identifiers for the sender name in the unsolicited email (i.e., signatures at the end of the emails). Letters were assigned to names for the purpose of deidentification. Once all letters of the alphabet were used, letters accompanied by numbers were used (i.e., A1, B1, C1, etc.). "No Name" indicates that no signature was included in the email.

**Publisher ID (Column C)**

This column includes identifiers for the publishers name. Letters were assigned to publisher names for the purpose of deidentification. "CNV" indicates the publisher name could not be located via Google or the journal's website.

**Journal ID (Column D)**

This column includes identifiers for the journal name. Letters were assigned to journal names for the purpose of deidentification. Once all letters of the alphabet were used, letters accompanied by numbers were used (i.e., A1, B1, C1, etc.). "No Journal Name" indicates that no journal name was included in the email.

**Date Received (Column E)**

This column indicates the date upon which the unsolicited email was received. Dates are in month/day/year format.

# Characteristics Examined for Red Flags (Columns F through R)

**Note: Red shading in cells indicate the presence of a red flag.**

**Email reflects journal (Column F)**

This column indicates whether the unsolicited email's address reflected the name of the journal.

* Y: Yes (i.e., the email address does reflect the name of the journal)
* N: No (i.e., the email address does not reflect the name of the journal) -**Red flag**
* NA: No journal name provided (i.e., whether the email address reflects the name of the journal could not be verified)

**Correct name (Column G)**

This column indicates whether the unsolicited email correctly addressed the author by name in the salutation.

* Y: Yes (i.e., the email did address the author by the correct name)
* N: No (i.e., the email did not address the author by the correct name) -**Red flag**
* No Name: There was no name in the solicitation (i.e., a generic greeting was used, e.g., Dear Researcher) **-Red flag**

**Correct honorific (Column H)**

This column indicates whether the unsolicited email used the correct honorific in their salutation. For this study the correct honorific would be nothing (i.e., Dear Paije Wilson or Dear Wilson P), as the author does not have a PhD.

* Y: Yes (i.e., the email used the correct honorific)
* N: No (i.e., the email used the incorrect honorific. E.g., "Dear Dr. Wilson") -**Red flag**
* NA: Not applicable (i.e., the email used a generic or incorrect name in the salutation)

**Flattery (Column I)**

This column indicates whether the unsolicited email used flattery (e.g., "Your publication was impressive and enlightening").

* Y: Yes (i.e., the email did use flattery) -**Red flag**
* N: No (i.e., the email did not use flattery)

**Grammar/spelling errors (Column J)**

This column indicates whether the unsolicited email had spelling and/or grammatical errors.

* Y: Yes (i.e., the emails did have spelling and/or grammatical errors) -**Red flag**
* N: No (i.e., the emails didn't have any spelling or grammatical errors)

**Deadline (Column K)**

This column indicates whether unsolicited email included a deadline for submitting a manuscript.

* A date (in month/day/year format) indicates a deadline was included, with the date indicating the deadline for manuscript submission. -**Red flag**
* NA: Not applicable (i.e., no deadline was mentioned in the email)

**Scope match (Column L)**

This column indicates the degree to which the journal's scope matched the author's or the author's publication's field (as determined by the email mentioning a specific scope or deducing the scope from the journal title). The author's field was library science, with their publication representing the subjects of library science or clinical trials data sharing.

* Y: Scope matches (i.e., the scope directly matches the author's or their publication's field)
* C: Scope matches because journal is moderately broad in scope
* B: Scope matches because journal is exceptionally broad in scope -**Red flag**
* N: Scope does not match **-Red flag**
* No subject: Scope was not indicated (i.e., the degree of scope match could not be verified)

**Persuasion (Column M)**

This column indicates whether the unsolicited email used persuasion (e.g., "Here are reasons why you should publish with us").

* Y: Yes (i.e., persuasion was used in the email) -**Red flag**
* N: No (i.e., persuasion was not used in the email)

**Discount (Column N)**

This column indicates whether the unsolicited email mentioned a discount to the article processing charge (APC) if the author submitted their manuscript to them.

* Y: Yes (i.e., the email mentioned there being a discount) -**Red flag**
* N: No (i.e., the email did not mention there being a discount)

**Follow Up (Column O)**

This column indicates whether the unsolicited email was a follow up. To be categorized as a follow up emails had to specifically mention following up on a past email.

* Y: Yes (i.e., the email mentioned following up on a previous email) -**Red flag**
* N: No (i.e., the email does not mention following up on a previous email)

**Email Domain (Column P)**

This column indicates the email domain (e.g., .com, .org, etc.) of the unsolicited email. Email addresses with Gmail and Yahoo domains were specified, so that they may be distinguished from other .com email addresses. Gmail and Yahoo domains were considered to be unprofessional, and were thus marked as **red flags.**

**Email Submission (Column Q)**

This column indicates if the unsolicited email mentioned allowing for email submissions of manuscripts.

* Y: Yes (i.e., the email mentioned allowing for email submissions) -**Red flag**
* N: No (i.e., the email did not mention allowing for email submissions)

**Fake impact metric (Column R)**

This column indicates whether the emails mentioned having an impact metric known to be fake

* Y: Yes (i.e., the email mentioned an impact metric known to be fake) -**Red flag**
* N: No (i.e., the email only mentioned impact metric(s) known to be legitimate)
* NA: Not applicable (the email did not mention any impact metric)

# Total Red Flags (Column S)

**Totals (Column S)**

This column indicates the total number of red flags were identified for each unsolicited email.

# Additional Characteristics (Columns T through AA)

**Author name same as pub (Column T)**

This column indicates whether the name included in the unsolicited emails' salutation was the same as the name used in the author's past publication (i.e., Wilson P).

* Y: Yes (i.e., the email used the same name as that included in the author's past publication)
* N: No (i.e., the email used the author's first name in the salutation)
* NA: Not applicable (i.e., the email used a generic or incorrect name in the salutation)

**Pub Abstract (Column U)**

This column indicates whether the unsolicited email included the author's past publication's abstract.

* Y: Yes (i.e., the abstract was included in the email)
* N: No (i.e., the abstract was not included in the email

**Pub Title (Column V)**

This column indicates whether the unsolicited email mentioned the title of the author's past publication.

* Y: Yes (i.e., the email mentioned the title of the author's past publication)
* N: No (i.e., the email did not mention the title of the author's past publication)

**Index Mentions (Column S)**

This column indicates whether unsolicited email mentioned being indexed.

* Y: Yes (i.e., the email mentioned being indexed)
* N: No (i.e., the email did not mention being indexed)

**Name of Index (Column X)**

This column indicates, if applicable, the name of the index that was mentioned.

* Names of specific indexes mentioned in the email were included (e.g., Scopus)
* No name: Indexes were not named (i.e., the email mentioned being indexed, but didn't specify where)
* NA: Not applicable (i.e., the email didn't mention being indexed)

**Impact Factor Mentioned (Column Y)**

This column indicates if the unsolicited email mentioned having some sort of an impact factor.

* Y: Yes (i.e., the email mentioned having some sort of an impact factor)
* N: No (i.e., the email did not mention having an impact factor)

**Name of Impact Factor (Column Z)**

This column indicates, where applicable, the name of the impact factor that was mentioned. As a note, any email mentioning journal impact factor was assumed to be referring to the Journal Impact Factor metric from Clarivate's Journal Citation Reports (JCR). A note of "-NJCR" indicates the author searched for the journal in JCR, but the journal was not found.

* Specific names of impact factors mentioned in the email were included (e.g., Journal impact factor)
* No name: No name included (i.e., the email mentioned having an impact factor, but didn't specify the name of the impact factor. E.g., "This journal has an impact of X")
* NA: Not applicable (i.e., the email didn't mention having an impact factor)

**Editor in Chief/Speaking for (Column AA)**

This column indicates whether senders of the unsolicited emails claimed to be or claimed to speaking for the editor in chief of the journal.

* E: Sender claimed to be the editor in chief of the journal
* S: Sender claimed to be speaking for the editor in chief of the journal
* N: Sender did not claim to be or be speaking for the editor in chief of the journal

# Beall's List Mentions (Columns AB through AC)

**Journal in Beall's (Column AB)**

This column indicates whether the journal's name was included in [Beall's List of Potential Predatory Journals and Publishers](https://beallslist.net/).

* Y: Yes (i.e., the journal was in Beall's List)
* N: No (i.e., the journal was not in Beall's List)
* CNV: Could not verify (i.e., the journal did not provide their name in the email, so their inclusion in Beall's list could not be verified)

**Publisher in Beall's (Column AC)**

This column indicates whether the journal's publisher was included in [Beall's List of Potential Predatory Journals and Publishers](https://beallslist.net/). The publishers of the journals were verified via Google or the journal's website.

* Y: Yes (i.e., the publisher was in Beall's List)
* N: No (i.e., the publisher was not in Beall's List)
* CNV: Could not verify (i.e., the author was unable to find the name of the journal's publisher via the journal's website or Google)