Data Dictionary

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

**Record number (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 name (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.

**Journal (Column C)**

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.

**Email reflects journal (Column D)**

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)
* NA: No journal name provided (i.e., whether the email address reflects the name of the journal could not be verified)

**Correct name (Column E)**

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)
* NA: Not applicable (i.e., a generic greeting was used, e.g., Dear Researcher)

**Name same as pub (Column F)**

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)

**Correct honorific (Column G)**

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")
* NA: Not applicable (i.e., the email used a generic or incorrect name in the salutation)

**Flattery (Column H)**

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)
* N: No (i.e., the email did not use flattery)

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

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)
* N: No (i.e., the emails didn't have any spelling or grammatical errors)

**Date (Column J)**

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

**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.
* 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
* N: Scope does not match
* NA: Scope was not indicated (the degree of scope match could not be verified)

**Abstract (Column M)**

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

**Persuasion (Column N)**

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)
* N: No (i.e., persuasion was not used in the email)

**Publication title (Column O)**

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)

**Discount (Column P)**

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

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

**Follow Up (Column Q)**

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)
* N: No (i.e., the email does not mention following up on a previous email)

**Email Domain (Column R)**

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.

**Index (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)

**Impact (Column T)**

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)

**Email Submission (Column U)**

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)
* N: No (i.e., the email did not mention allowing for email submissions)

**Index Mentioned (Column V)**

This column provides additional information for Column S. If an unsolicited email has "Y" in Column S (i.e., they mentioned being indexed), Column V indicates what index, if any, was mentioned. NA means that the email mentioned being indexed, but didn't specify the index.

**Impact Metric (Column W)**

This column provides additional information for Column T. If an unsolicited email has "Y" in Column T (i.e., they mentioned having any sort of impact factor), Column W indicates the source of the impact factor, if any, that was mentioned. NA means that the email mentioned having an impact factor, but didn't specify the source. JCR-NCJR means that the email mentioned having a journal impact factor, but, when the author checked, the journal was not listed in Journal Citation Reports (JCR).