Identifying science in the news: An assessment of the precision and recall of Altmetric.com news mention data

Codebook for Content Analysis

February 28, 2022

Alice Fleerackers, Interdisciplinary Studies, Simon Fraser University, Canada
Lise Nehring, Centre for Forest Biology, University of Victoria, Victoria, BC, Canada
Lauren A. Maggio, Department of Medicine, Uniformed Services University, USA
Laura Moorhead, Journalism, College of Liberal and Creative Arts, San Francisco State University, USA

Table of Contents

1. Key definitions
2. Coding instructions
3. Instructions for identifying research
    a. With link
    b. Without link
4. Tips for success
    a. Identifying mentions of research
    b. Coding mentions
5. Codes
    a. News story details
    b. Linking practices
    c. Mention details

Key definitions

- **Research output**: A scholarly document with a DOI or permanent identifier that is mentioned or linked to in the news story. For example, a peer-reviewed journal article, a preprint, or a dataset.
- **News story:** An online article, associated with a unique URL, that was published by one of the following six media outlets between [DATE 1] and [DATE 2], 2021: *New York Times - Science, Guardian - Science, WIRED, PopSci, News Medical, Medpage*. News stories can include “hard” news reporting, editorials, letters to the editor, blog posts, and other kinds of online texts. Do not code multimedia or visual elements, only text.
- **Mention:** A reference to a research output, either in the form of an external hyperlink or a text-based description of research. Mentions can include (but are not limited to) the following: a hyperlinked word/phrase, a short description of the research findings/authors/journal of publication, or an entire news story discussing the research output.

Coding instructions

1. Go to the “Entries” area in your Science News Story Intercoder Coding Form
    a. Note: you may need to log on to Cognito Forms first
2. Each row in the Entries area will include a URL, corresponding to a unique science News Story, along with Coder Initials (LN or AF)
    a. Select a row marked with your Coder Initials
    b. Click “Edit Entry”
    c. For the news story URL provided, read the story in full, noting any mentions of research outputs
    d. Proceed to code the news story in full, following the instructions provided in the Coding Form
3. **Only code research outputs that are identifiable** (i.e it is possible to find the specific study online, at a journal or preprint server, either by searching for it using the Search Strategy below or by clicking through the study URL).
4. You will need to focus on different parts of the news story at different times during your coding:
    a. For the ***News Story Details*** codes, draw data from the entire news story
    b. For the ***Linking Practices*** and ***Mention Details*** codes, focus on the relevant excerpt of the news story (the links, words, phrases, sentences, or paragraphs mentioning that particular research output)
    c. For the ***Study DOI*** code, you will have to draw data from the web page where the study can be found. A link to this website may be provided in the news story, or you may need to find it yourself by entering key details into a web search (e.g., Google Scholar or an alternative).
5. **Do not code embedded content**, such as YouTube videos, Tweets, Instagram posts, podcasts, advertisements, etc. Code only the text itself (including image captions, but not images).
6. **If research is mentioned in the story**:
    a. Copy-paste the relevant **excerpt** into the field marked *excerpt.* (If the research is mentioned throughout the news story, write “throughout” in this column instead.)
    b. Fill out all of the fields for that particular mention, as prompted by the Coding Form, before moving on.
    c. If there are **multiple mentions of research** in a single-story, you will need to click “Add Item” for each new research output mentioned before you can continue coding.
7. **If research is not mentioned** in the story, fill out all the *news story details* codes, but leave the rest of the spreadsheet blank. Select “No” for the *research_mentioned* field
8. For each mention, please use the ***Notes*** **feature** to document any idiosyncratic findings that may be important to the researchers (anything that is unique or stands out to you).
9. If you have any **questions**, please contact Alice Fleerackers (afeerac@sfu.ca).

Instructions for identifying research mentions

When identifying research mentions, the key question to ask yourself is: ***Would a reasonably science-literate reader be able to find the study within 5 min?***

A) With link

If the mention includes a hyperlink to a research output, click on the link to identify the DOI or scholarly identifier.

- The mention could link to research *directly* (e.g., hyperlink leads to a PDF download, a page on a preprint server, scholarly journal, or registered trial database)
- The mention could also link to research *indirectly* (e.g., hyperlink leads to a press release that cites or links to research).

Code both as identifiable research mentions (as long as you can find a DOI/scholarly identifier).

B) Without link

For research mentions that describe research but do not include a hyperlink, use the following **search strategy**:

1. Copy the phrase(s) containing key study details
2. Paste the phrases into Google Scholar's search bar
3. Remove non-essential words (e.g., prepositions, articles, adverbs, adjectives)
4. Search
5. Restrict date: If a pub date is provided, use Google Scholar's date settings to restrict your search to the year indicated; otherwise, try restricting to 2021 or 2020, depending on when the news story was published
6. Do not go past Page 1 of Google Scholar search results and do not spend longer than 5 minutes searching.

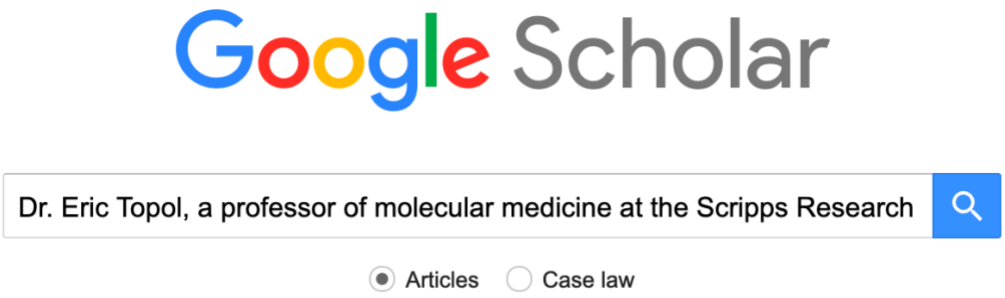
A note on removing non-essential words: Details like researcher names, journal titles, publication years, key findings (e.g., statistics/numbers), sample size or type, methodological details tend to be the most helpful for identifying research.

Example search strategy:

1. *Copy the phrase(s) containing key study details*

Dr. Eric Topol, a professor of molecular medicine at the Scripps Research Institute in San Diego, said he is starting a study to measure physiological information like heart rate, respiratory rate, temperature and markers of immune system response in people with long Covid before they receive a vaccine and weeks afterward.

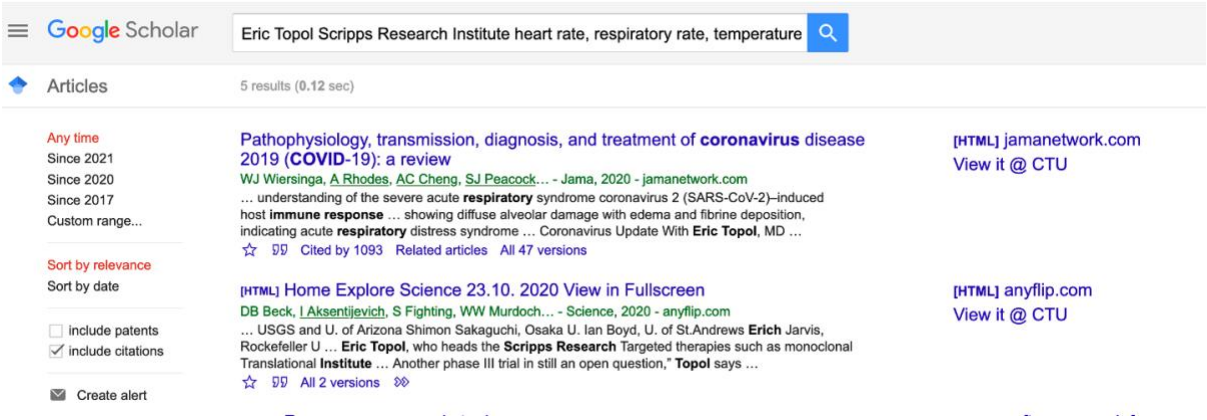
2. *Paste the phrase(s) into Google Scholar's search bar*



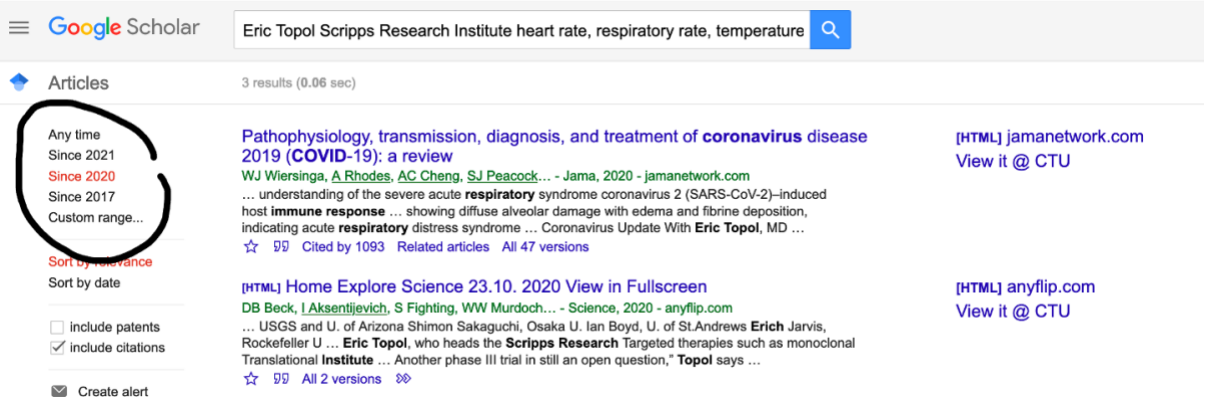
3. *Remove non-essential words*

~~Dr.~~ Eric Topol, ~~a professor of molecular medicine at the~~ Scripps Research Institute ~~in San Diego, said he is starting a study to measure physiological information like~~ heart rate, respiratory rate, temperature and markers of immune system response ~~in people with~~ long Covid ~~before they receive a~~ vaccine ~~and weeks afterward.~~

4. *Search*



5. *Restrict date*



6. Do not go past page 1 of Google Scholar search results and do not spend longer than 5 minutes searching.

Tips for success

Identifying mentions of research

- **Turn off any Library Proxies used to access research** to ensure you are not automatically redirected to newer versions of studies (e.g., for preprints). You should still be able to access the abstracts and metadata for each research mention
- **Read the entire article**, not just the areas with hyperlinks
    - As you read, pay special attention to terms such as “a study”, “a preprint”, “a new book” but also to related terms like “A researcher” or “Scientists” as well as “modelling”, “recent review”, “a commentary”, “new evidence,” “trial”, etc
- **Hover or click on each hyperlink** in the story to see if research is hyperlinked
    - I prefer Firefox’s Reader View for this, as it makes it easier to see links
    - If two links lead to the same research output, treat them as a single mention
    - Be careful: sometimes DOIs for different studies may be almost identical (i.e. you might believe they lead to the same study, when in fact they are distinct)
- **If a specific research output cannot be identified** by Google Scholar search or by following a hyperlink to the study, do not code this as a mention
    - E.g., Code No for vague statements like “According to the latest scientific evidence” or “The scientific community has found” (unless it is possible to associate a specific study with these descriptions)
- **Look out for alternative forms of research**. Scholarly books, registered clinical trials, published study protocols, editorials, letters, and commentaries in academic journals all count as research *if they have a scholarly identifier*
    - Reports do not often have a scholarly identifier, but some do. Double check.
- **Look out for mentions of multiple research outputs** in the same sentence. For example: “Dr. Miller-Jones and an international cast of colleagues reported the result in the journal Science and in a pair of companion papers in Astrophysical Journal.” In this mention, only one paper is linked, but three are discussed. Try to find DOIs for all three (reading the rest of the news story can help) and code as three separate mentions.

Coding mentions

- **Consider the broader context** when deciding what to code for each mention, not just the phrase where the study is linked but any part of the text mentioning the researchers involved, the way the study was conducted, or the results.
- If the **same research output is mentioned multiple times** in a single story, treat it all as one research mention (not separate mentions)

Codes

1. News story details
2. Linking practices
3. Mention details

1. News story details

For the following two codes, the unit of analysis is the news story, not just the excerpt specific to the research being mentioned.

Code:	**aggregated**
Brief definition:	The story was first published by another source (and is not a press release).
Instructions:	Is this an aggregated story? (i.e. Was the story first published by a different blog or news outlet?) If so, enter Yes. (You will likely need to examine the story byline or attribution line, usually found under the headline or at the end of the story). Enter No if there is no mention that this story was previously published by another source. Also Code 0 if the story was first published as a press release (see “press_release”, below). **Coding tips:** ● Enter Yes if coding an original story that was later republished by another outlet (e.g., stories first published by The Conversation). ● Enter Yes for transcripts of radio shows, podcasts, videos that were originally published by another outlet (Code No for transcripts for media published by the outlet you are coding) ● Enter No for sponsored content; code as press release instead
Examples (hyperlinks to the research are bolded, rationale stated in italics):	***CODE Yes:*** *“Source: The Conversation – USA – By Adam M. Brufsky, Professor of Medicine, University of Pittsburgh”* ***Also CODE Yes:*** *The story is accompanied by the logo of another media outlet (presumably, the original outlet).* ***CODE No:*** *No attribution line.* ***Also CODE No:*** *“This story originally appeared on WIRED UK” [for a WIRED.com story]*

Code:	**press_release**
Brief definition:	The news story is simply a published press release.
Instructions:	Is this story a published press release? If so, enter Yes. (You will likely need to examine the story byline or attribution line, usually found under the headline or at the end of the story). Enter *No* if there is no mention that this story is a press release. **Coding tips:** ● The following are hints you may be reading a press release: ○ A date and/or location at the top of the story ○ A statement that the story was provided by a research institute, group, or university ○ Contact information for the study author(s), relevant subject experts and/or a press officer ○ Multiple quotes from the study author(s) ○ Links for "More Information" (e.g., about the research institute or the study itself) ● If unsure, you can always search for the title of the story in your browser to see if it was first published by a research institute/ public relations firm. ● Code Yes for sponsored content
Examples (hyperlinks to the research are bolded, rationale stated in italics):	***CODE Yes:*** *"Provided by Howard Hughes Medical Institute" ← This attribution line suggests the story is a press release* ***Also CODE Yes: "****NEW YORK and BASEL, Switzerland, May 13, 2020 /PRNewswire/ -- Roivant Sciences today announced several updates for its adaptive, randomized, double-blind, placebo-controlled, multi-center pivotal BREATHE trial…" → This "lede" includes a date, location(s), and the name of a PR Newswire service; it is clearly a press release* ***CODE No:*** *No attribution line.*

Code:	**research_mentioned**
Brief definition:	The story mentions at least one scholarly research output.

Instructions:	Does the news story include a hyperlink to at least one identifiable scholarly output (e.g., a peer reviewed paper, a preprint, a data set that can be associated with a permanent URL or DOI)? Does the new story reference at least one identifiable scholarly output within the text itself? (E.g., by using terms like “researchers have found”, “a study shows,” “a recent preprint”, “a new assay”, “preliminary findings,” “research suggests”, or a similar phrase). If either of the above is true, enter Yes. Enter No if there is no mention of scholarly research in the story. Leave the rest of the coding form empty and move on to the next news story. **Coding tips:** • If the story includes **hyperlink to a press release** about a research study, code "Yes" for this code. • If the story includes **hyperlink to a separate news story** about a research study, but does not provide enough identifying details to find the study via the Google Scholar Search Strategy, code "No" for this code. • **Do not count reports** as research outputs, unless they are associated with a DOI or other scholarly identifier (e.g., SSRN, PubMed ID/PMID, arXiv ID, ISBN, etc). • Please **count scholarly books (including textbooks)** as research outputs. A scholarly book is a book published by a University Press (e.g., "Oxford University Press, MIT Press") or by an Academic Press (e.g., SAGE Publications), usually written by a person with a university affiliation [Check this list of Academic Book Publishers if needed]. Make a note of the fact that it was a book in the Notes box. • **Commentary articles, editorials, letters** to the editor all count as Research Outputs if they were published in an academic venue (e.g., Journal, Preprint Server, Research Repository) and are associated with a DOI/other scholarly identifier. • **Conference presentations CAN count as research, but only if it is possible to find a scholarly identifier** for the presentation. Some conferences make these available (E.g., through a published abstracts or paper), some do not
Examples (hyperlinks to the research are bolded, rationale	***CODE Yes:*** **“**While there is an ever-changing list of new SARS-CoV-2 variants, three have emerged that scientists are most concerned about — at least for now — in addition to the original virus.” → *Link points to a peer reviewed article*

stated in italics):	*in JAMA* ***CODE No:*** **"**That is what Katherine May explores in *Wintering: The Power of Rest and Retreat in Difficult Times* (*public library*) — a gorgeous book, a generous book, a layered book of uncommon sensitivity and substance, drawn from May's own experience of living through a deep and disquieting winter of life." → *Links to a book, but it is published by Penguin, a trade (not academic) publisher.* ***Also CODE No:*** "During the past year the pandemic has seen the public, and business sectors, heading back to more trusted news sources. More people than ever are paying for news from leading media organisations, according to this year's *Digital News Report* from Oxford's Reuters Institute." → *References a research report, but not one associated with a scholarly identifier*

2. Linking practices

For the following codes, the unit of analysis is the research excerpt (i.e. the part of the story that specifically refers to the research output being analyzed).

Code:	**describes_as_research**
Brief definition:	The story suggests that the research output mentioned is scientific research.
Instructions:	Is there some indication that what is mentioned involved research? (e.g., "A recent study found", "New research finds", "A team of scientists have found", "...a team of scientists", "a recent review of evidence", "a retracted paper")? If so, enter Yes. Enter No if there is no indication that research has been done. **Coding tips:** ● Look out for synonyms of research, such as *modelers estimated, researchers have now observed, the model shows, scientists have found, scientific literature reveals, new review, preprint, clinical trial, assay, patient registry* ● The following terms do not, on their own, suggest research has been done: *data, statistic(s), evidence, report, editorial* ● Code Yes if the research is cited as a reference (i.e. using formal scientific citation style) at the end of the story, rather than within the main narrative.

	• Code Yes for descriptions of “researchers” or “scientists” doing research in any way • Code Yes for quotes from researchers about the study methods/findings/etc • Code Yes if a scholarly journal/preprint server is mentioned in association with the finding • Consider the larger context within which the research is wdescribed — not just the first mention/link of the research • For each mention, ask yourself: “As an educated reader, can I tell that research has been done here?”
Examples:	**CODE Yes:** “Researchers have now observed that **some virus particles are spherical**, while others are more egg-shaped. Their sizes vary, with diameters ranging from 80 to 160 nanometers.” → *It is clear that the study is a research study, but not that it is a preprint.* **Also CODE Yes:** The study citation is written out in full at the bottom of the story → *This makes it clear that it’s a research study.* **CODE No: “**Efficient transmission in crowded, enclosed spaces also explains the high attack rates in nursing homes, food processing plants, jails and prisons and cruise ships. On the flip side, the risk of transmission does seem to be **lower** outdoors.” → *There is no mention that the research that has been linked to is a research study (or a preprint) at all. It’s just a link.*

Code:	**has_link**
Brief definition:	The story contains a link to the study it cites.

Instructions:	Does the story contain a link to the research (i.e. is it possible to click through to the original study)? If so, code Yes. Code No if the story does not include a hyperlink for the study. **Coding tips:** ● Links are often omitted in aggregated content. ● Please make sure you double check that this box is clicked ● **Code Yes if there is a link to a press release** ABOUT a research study, but no link to the study itself (but make a note in the Notes column) ● **Code No if there is a link to a separate news story** ABOUT a research study, but no link to the study itself (but make a note in the Notes column)
Examples:	**CODE Yes:** "In a new paper published on the preprint journal **bioRxiv** in April 2020, researcher T. Konishi argues that the coronavirus family is a constellation of pandemic pathogens, which includes the currently spreading severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)" → *Story includes a link to the preprint, bolded in this example.* **CODE No:** "Researchers believe that there was one infected but asymptomatic person sitting at a table in the restaurant. Because of the air currents circulating in the room due to air conditioning, people sitting at two other tables became infected, likely because of aerosols." → *This is a news story that was first published by The Conversation. In the original story, the phrase 'likely because of aerosols' included a link to a study posted on medRxiv; it has been removed in this version of the story.*

3. Mention details

For the following codes, the unit of analysis is the research excerpt (i.e. the part of the story that specifically refers to the research output being analyzed).

Code:	**journal_mentioned**
Brief definition:	The excerpt mentions the name of the journal or server in which the research can be found.

Instructions:	Does the story state the name of the journal or server in which the research can be found? (e.g., "A recent study published in *Health Communication*" or "A preprint posted to bioRxiv"? If so, enter Yes. Enter No if there is no indication of a journal or server. **Coding tips:** ● Journal titles are often noted in *italics*, but preprint servers typically are not. ● Google search the title if you are unsure whether it is a scholarly output; magazines, blogs, or newspapers mentioned can be mistaken for academic publications. ● Sometimes a research output is mentioned as a citation at the bottom of a news story (e.g., Name, A., Name,B. et al (2021), Nature…). Code YES for "Journal mentioned"
Examples:	**CODE Yes:** "'Young people may be dying because they are not getting the treatment they need,' said Brendan Saloner, an addiction researcher at Johns Hopkins Bloomberg School of Public Health who wrote an editorial published with the study Monday in *JAMA Pediatrics*." → *JAMA Pediatrics is the journal* **CODE No: "**And here's this from *Politico*, which has a slightly higher percentage of addiction: "According to a 2016 national survey conducted by the Substance Abuse and Mental Health Services Administration, 87.1 million U.S. adults used a prescription opioid—whether prescribed directly by a physician or obtained illegally." → *Although Politico is in italics, it is a news outlet, not a scholarly journal/server.*

Code:	**author_mentioned**
Brief definition:	The excerpt mentions the name of one or more authors involved in the research.

Instructions:	Does the story list at least one author's name? If so, code Yes. Code No if there are no authors mentioned. **Coding tips:** ● The story does not need to include the author's full name to be coded as Yes; a last name is sufficient. ● To receive a Yes for this code, the story can list multiple authors or just one of them. ● If the author of the story is writing about their own research, code Yes. (The author's name is in the story byline) ● Sometimes a research output is mentioned as a citation at the bottom of a news story (e.g., Name, A., Name,B. et al (2021), Nature…). Code YES for "Author mentioned"
Examples:	**CODE Yes**: "... there is a growing body of evidence suggesting the FDA's label isn't accurate. 'Essentially, it's because of the backlash against the psychedelic '60s," says Dr Thomas Kingsley Brown, an anthropologist and chemist at the University of California, San Diego. In a study spanning eight years, funded by the Multidisciplinary Association on Psychedelic Studies, or MAPS, Brown tracked outcomes for addicts who were treated with ibogaine." → *Brown is the author of the research mentioned.* **CODE No:** "Researchers believe that there was one infected but asymptomatic person sitting at a table in the restaurant. Because of the air currents circulating in the room due to air conditioning, people sitting at two other tables became infected, likely because of aerosols." → *No indication of a study author.*

Code:	**institution_mentioned**
Brief definition:	The excerpt mentions the name of the institutional affiliation of one or more authors involved in the research.
Instructions:	Does the story list at least one author's institutional affiliation? (e.g., the name of a university, college, research institute, or private research firm.) If so, code Yes. Code No if there are no institutions mentioned. Also code No if only the department or program of the author(s) is mentioned, without the institution.

	Coding tips: ● If the author of the story is writing about their own research, look at their byline. Code Yes if the byline includes their academic affiliation as well. ● Institution names are often found near the name of the author(s). However, sometimes an institutional affiliation is mentioned even when the authors are not.
Examples:	**CODE Yes**: "... A study from the University of Kentucky College of Public Health found states where Medicaid coverage was expanded under the Affordable Care Act had an increase in prescriptions for a drug to combat opioid addiction." → *University of Kentucky College of Public Health is an institution where a researcher associated with the study is affiliated.* **CODE No:** "'Emerging year-over-year figures and episodic outbreaks of fentanyl-related deaths paint a grim picture of an uncontained, plague-like contagion,' write Professors Leo Beletsky and Corey Davis in a paper on the fentanyl crisis published last year in the *International Journal of Drug Policy*." → *No indication of either author's institutional affiliation.*

Code:	**study_date_mentioned**
Brief definition:	The story mentions the year that the research was published.
Instructions:	Does the story mention the year that the research was published? If so, code Yes. Code No if the story does not provide a date for the research output in question. **Coding tips** ● Code YES if story cites a preprint or paper as a formal reference (e.g., APA style, with a date). ● If the story mentions a month but not a year, assume it was published in the same year that the news story came out. Code YES for *study_date.*

	• Code NO for mentions like “a new study” or a “recent preprint” (even though it is likely the study came out that year). Some kind of date should be explicitly mentioned.
Examples:	**CODE Yes:** “In December 2020, researchers at Georgia State University in Atlanta published a study in Nature Microbiology that suggested the potential for molnupiravir in prophylaxis. The study showed that giving molnupiravir to ferrets infected with SARS-CoV-2 prevented transmission to uninfected ferrets housed in the same cage.” **Also CODE Yes:** “A team of physicists from the University of Hamburg would beg to differ. In a new study published in Nature Communications in January, they created an unprecedentedly dense plasma out of a few thousand extremely cold atoms by hitting them with laser light for just one quadrillionth of a second.” → *News story specifies specific date (January of that year)* **CODE No:** “Recent research from around the world shows that marine protected areas increase fish stocks, ultimately allowing fishing communities to catch more fish on the edges of the reserves.” **Also CODE No:** “Public health experts have been concerned that for those receiving Moderna’s and Pfizer’s two-shot vaccine, not everyone would return for their second dose. But new CDC data shows that nearly nine out of every 10 first-shot-vaccinated Americans went back on time for their second dose. The data was based on a sample of 40.5 million Americans who were vaccinated between December 14, 2020, and February 14, 2021, and of the folks that got their second shot, nearly all of them got it on time.” → *dates indicate the data collection period, not the date of publication*

For the following code, the unit of analysis is the research output (go to the URL of the journal/server where the output can be found).

Code:	**study_doi**
Brief definition:	The digital object identifier
Instructions:	If available, enter the digital object identifier (DOI) here. The DOI is a multi-digit code in the form of 10.xxxx/xxxx that is often found in the URL of the research output and/or is provided somewhere under the title of the research. DOIs are often preceded by "https://doi.org/", "http://dx.doi.org/" or simply "DOI:" Add a URL to the study here if it is not possible to find the DOI.
Examples (hyperlinks to the research are bolded, rationale stated in italics):	*Example of DOI provided under the title of the research.* *Example of DOI in the URL:*