Codebook for Content Analysis

Directions:

1. Fill out each item below for each media story on the corresponding spreadsheet:
    a. Each row corresponds to a media story citing a particular COVID-19 preprint posted to either bioRxiv or medRxiv;
    b. Each column is a variable corresponding to one of the codes in the coding scheme presented below;
2. For each media story, identify where the preprint is mentioned.
    a. You may need to hover your mouse over all of the hyperlinks to identify the mention.
    b. If the preprint is not cited using a link, open the preprint in a new tab and look for other identifying characteristics that might be mentioned in the media story (e.g. author names or affiliations, key statistics or findings mentioned in the abstract).
    c. Note that a preprint may be mentioned multiple times. Be sure to check the entire story for any mentions before moving on.
    d. If you still cannot find the mention, mark it as an Exclusion.
3. Copy-paste the passage(s) from the media story that are most relevant to the preprint into the "mention_text" column. Bold any text that includes a hyperlink to the preprint. Keep the media story open in your browser; you will need to refer to the entire text for certain codes while performing the following steps.
4. Draw data from the mention_text, and enter data into corresponding columns using the coding scheme (e.g., 1 or 0).
5. Use the "notes" column to document interesting patterns, questions, or anything that might be important to the researchers.

Universe: English-language online news stories published by the 15 news outlets that cited the most bioRxiv or medRxiv preprints related to COVID-19 between January 1, 2020 and May 1, 2020. Preprints restricted to the 100 most cited preprints (according to Altmetric).

Unit of Analysis: Preprint mention.

Exclusions: Exclude stories that are a) not written in English, b) do not refer to the correct preprint (i.e. corresponding to the DOI pulled by the Altmetric API), or c) cannot be found using the title/URL provided by Altmetric. Mark these as exclusions with a brief explanation for why they were removed.

Media Story Qualities

For the following four codes, the unit of analysis is the media story, not just the excerpt specific to the preprint.

Code:	**scicomm**
Brief definition:	The primary focus of the story is to communicate the results and/or implications of the preprint.
Instructions:	Is the primary focus of the story to communicate the results and/or implications of the preprint? If so, enter 1. Enter 0 if the primary focus is some larger issue (i.e. the preprint is only mentioned in passing). *NOTE: The thematic focus is usually obvious from the story lede and headline, but may require further reading. Check the whole story if you're unsure.*
Examples (hyperlinks to the preprint are bolded, rationale stated in italics):	**CODE 1:** "The actions many communities are taking to slow the spread of COVID-19 may be shrinking deaths in a pattern that brings together nature, art and math, according to a new study from the University of Michigan. The findings, described in a preprint paper that has not yet been peer-reviewed, indicate that containment works and could help us snuff out the virus faster." → *this is the story lede; entire story focuses on the findings and implications of the preprint* **CODE 0:** "Many physicians noticed that people with high blood sugar, not only those with a history of diabetes but also **unexplained new diabetes, were showing up in the hospital** with the novel coronavirus. This indicated to me that something could be going on with the addition of sugar molecules to the virus, or the receptor it latches onto to infect cells, that influenced the severity of the disease." → *story focuses on the effects of adding or subtracting sugar molecules from the coronavirus's spike protein; preprint is only mentioned in passing*

Code:	**aggregate**
Brief definition:	The story was first published by another source (and is not a press release).
Instructions:	Is this an aggregated story? If so, enter 1. (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 0 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). *Note: Enter 1 if coding an original story that was later republished by another outlet (e.g. stories first published by The Conversation).*
Examples:	**CODE 1:** "Source: The Conversation – USA – By Adam M. Brufsky, Professor of Medicine, University of Pittsburgh"

	Also CODE 1: The story is accompanied by the logo of another media outlet (presumably, the original outlet). **CODE 0:** No attribution line.

Code:	**press_release**
Brief definition:	The media story is simply a published press release.
Instructions:	Is this story a published press release? If so, enter 1. (You will likely need to examine the story byline or attribution line, usually found under the headline or at the end of the story). Tips for identifying press releases include: ● 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 preprint author(s), relevant subject experts and/or a press officer ● Multiple quotes from the preprint author(s) ● Links for "More Information" (e.g. about the research institute or the preprint itself) Enter 0 if there is no mention that this story is a press release. *NOTE: 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.*
Examples:	**CODE 1:** "Provided by Howard Hughes Medical Institute" ← *This attribution line suggests the story is a press release* **Also CODE 1: "**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 0:** No attribution line.

Code:	**define_preprint**
Brief definition:	The story defines preprints in some way.
Instructions:	Does the story offer some sort of definition or explanation of what a preprint is? If so, enter 1. Enter 0 if the story includes no discussion of what preprints are or how they should be interpreted.

	Note: This explanation may or may not occur within the same paragraph that cites the preprint you are analyzing. (Be sure to check the bottom and top of the page for a disclaimer!) *Note also that explanations and definitions do not have to be correct and may not necessarily include the word "preprint".*
Examples:	**CODE 1:** "Important Notice: bioRxiv publishes preliminary scientific reports that are not peer-reviewed and, therefore, not be regarded as conclusive, guide clinical practice/health-related behavior, or treated as established information." → *This disclaimer offers a brief explanation of what a preprint is.* **Also CODE 1:** "a search on PubMed reveals that some 947 research articles have already been published (although many are commentaries and review articles); and many more have become available through preprints – the 21st century way to report data almost in real time." → *This may be an unconventional explanation of what preprints are, but it is an explanation.* **CODE 0:** "A **new preprint on the scale of US spread** estimated that, by March 1, there were already 9,484 Covid-19 cases in the US. That's about nine times the 1,034 cases reported nationally." → *Story states that study is a preprint, but does not explain what the term means.* **ALSO CODE 0: "**The study was placed on bioRxiv (pronounced "bio-Archive"), an open-access distribution service for preprints of life science research." → *Although this phrase explains what bioRxiv is, it does not explain what preprints are!*

Scientific Uncertainty Framing Devices

For the remaining codes, the unit of analysis is the preprint mention (i.e. the excerpt from the media story that relates to the preprint in question). You may need to read a few paragraphs following the first mention of the preprint to capture related content, such as commentary from independent excerpts or quotes from the researcher. NOTE: Some stories will mention multiple preprints. Code only the excerpt that is relevant to the preprint identified in the coding sheet.

Code:	**preprint**
Brief definition:	The story mentions that the study is a preprint.
Instructions:	Does the story mention that the study is a preprint? (The mention should use the word "preprint", "pre-print", or "posted to a preprint server"). If so, enter 1. Enter 0 if there is no mention that the study cited is a preprint. *NOTE: Simply stating the name of the preprint server (i.e. bioRxiv or medRxiv) without mentioning the word "preprint" is not the same as stating that the study is a preprint.*

Examples (hyperlinks to the preprint are bolded, rationale stated in italics):	**CODE 1:** "In this study, made public on **BioRxiv**, an online archive of pre-print studies, remdesivir has shown promise with a significant clinical improvement, reducing pulmonary infiltrates and a decrease in lung disease." → *It is clear this study is a preprint.* **CODE 0:** "Researchers have not drawn a conclusion on how contagious COVID-19 is before symptoms but **the latest study** from Germany shows that COVID-19 presents before symptoms and virus loads are very high during the first week of symptoms, indicating the virus is contagious during the incubation period." → *No mention that this is a preprint.* **Also CODE 0:** "Their results, published Friday in **a study** that has yet to be peer-reviewed, suggest that in Santa Clara County, in early April, the number of people who are or have been infected with COVID-19 is 50 to 85 times greater than the number of confirmed cases." → *Story explains that study has not been peer reviewed, but does not state that it is a preprint.*

Code:	**unreviewed**
Brief definition:	The story explains that the study has not been peer reviewed.
Instructions:	Does the story mention that the study has not been peer reviewed? If so, enter 1. • *Synonyms: unreviewed, currently undergoing peer review, has not yet been peer reviewed, pre-publication* Enter 0 if there is no mention that study has not been peer reviewed.
Examples:	**CODE 1:** "Because of the urgent implications of the model's results, the team has made its study available online early. The research has not been peer-reviewed." → *It is clear that the study has not been peer reviewed.* **Also CODE 1:** "But in one preprint study, meaning it is currently under peer review, researchers used a model to estimate that a person standing and speaking in a room could release up to **114 infectious doses per hour."** → *Story states preprint is undergoing peer review; hence it has not yet been reviewed* **CODE 0:** "A **new preprint on the scale of US spread** estimated that, by March 1, there were already 9,484 Covid-19 cases in the US. That's about nine times the 1,034 cases reported nationally." → *Story mentions that study is a preprint, but does not mention peer review.*

Code:	**preliminary**
Brief definition:	The story suggests that the study is preliminary.
Instructions:	Does the story present the study as initial or preliminary (e.g. by using words such as "for the first time", "in a preliminary survey", or "an initial test", "early insight")? If so, enter 1.

	• *Synonyms: incomplete, trial stage, pilot study, only few studies* Enter 0 if there is no mention that the study is preliminary. *NOTE: Simply stating that research is "new" or "recent" does not in and of itself indicate that the research is preliminary. This code is less about a study's recency and more about the tentativeness of its findings. The same is true for terms such as "seems," "appears," "estimates" or "suggests" — these terms are commonly used when communicating research results and do not in and of themselves suggest the study is preliminary.*
Examples:	**CODE 1:** "A preliminary **study** based on clinical data from researchers in several countries found "no evidence" that hydroxychloroquine is a useful coronavirus-fighting weapon. Additional research funded by the National Institutes of Health and the University of Virginia has recently posted similar findings. So much for Trump's 'game changer.'" → *Mentions study is preliminary* **CODE 0:** "A **new preprint on the scale of US spread** estimated that, by March 1, there were already 9,484 Covid-19 cases in the US. That's about nine times the 1,034 cases reported nationally." → *States that the study is "new" but not preliminary*

Code:	**verification_needed**
Brief definition:	The story suggests that the study results are inconclusive (i.e. should be replicated or verified).
Instructions:	Does the story suggest that replication or further analysis is needed (e.g, by using phrases like "more work is needed before conclusions can be drawn," or "future studies are necessary"?) If so, enter 1. • *Synonyms: approach the results with caution, the study needs to be repeated, inconclusive, should not be regarded as conclusive, we won't know the exact risk until we have the results from a large-scale analysis* Enter 0 if there is no suggestion that more research is needed. *NOTE: Be sure to check the first paragraph or two after the preprint is first mentioned; stories often include discussions of caveats after the main findings are reported.*
Examples:	**CODE 1:** "In such patients, researchers have seen considerable increase of cytokines, particularly those that control **migration and activation of macrophages**...Further clinical and immunological analysis of severe COVID-19 cases are necessary for better understanding this infection and identifying treatments." → *Story makes it clear that more research is needed* **Also CODE 1:** "...given COVID-19 is a pandemic, the sample size that they have analysed is not sufficient to draw firm conclusions." → *Story suggests there is too much uncertainty to draw conclusions, implying that more research is needed.*

	CODE 0: "Now that new COVID-19 cases are being detected in the US every day, it is too late to stop the initial wave of infections. The epidemic is likely to spread across the US. The virus appears to be about as contagious **as influenza**." → *Story does not mention that study should be replicated*

Linking Practices

REMINDER: Unit of Analysis = Mention

Code:	**research**
Brief definition:	The story refers to the preprint as scientific research.
Instructions:	Does the story state or suggest that the preprint is a research study (e.g. "A recent study found", "New research finds", "A team of scientists have found")? If so, enter 1. • *Synonyms: modelers estimated, researchers have now observed, the model shows, scientists have found, new data shows, scientific literature reveals, new review* Also Code 1 if the preprint is cited as a reference (i.e. using formal scientific citation style) at the end of the story, rather than within the main narrative. Enter 0 if there is no indication that the preprint is a research study. *NOTE: Terms like "evidence", "a preprint," or a "paper" are not enough of an indication to Code 1.*
Examples:	**CODE 1:** "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 1:** 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 0: "**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 preprint 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 preprint (i.e. is it possible to click through to the original publication)? If so, enter 1. Enter 0 if the story does not include a hyperlink for the study.

	Note: Links are often omitted in aggregate content. Also note that preprints can be updated, which means the link in the media story may differ slightly from the DOI provided by Altmetric. As long as the link in the story leads to a version of the preprint, code 1.
Examples:	**CODE 1:** "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 media 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.* **CODE 0:** "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:	**just_link**
Brief definition:	The story links to the study, but does not make it clear that it is a research study or a preprint.
Instructions:	Does the story provide a hyperlink and nothing else (i.e. no elaboration or reference to "research", "a study", or "a preprint")? If so, enter 1. Enter 0 if the story makes it clear that the link refers to a study of some sort. Also enter 0 if there is no link provided (see *has_link*). • *Other terms suggesting the mention is more than just a link: Researchers have found, the model suggests, new findings show, studies have shown, the modelers used, etc* *NOTE: Citing a preprint as a formal reference (e.g. APA style, with a DOI) suggests the preprint is research. Code 1 for "research" (see above) but 0 for "just_link".*
Examples:	**CODE 1:** "The lowered immune response of pregnancy, which is needed to stop a woman's body responding to her baby as a health threat, may actually provide extra protection with COVID-19. COVID-19 seems to be **more severe** in people with an immune system working hard dealing with other health disorders." → *The study is linked to, but it is unclear from the text that this statement is related to research of any kind.* **CODE 0:** "A **new preprint on the scale of US spread** estimated that, by March 1, there were already 9,484 Covid-19 cases in the US. That's about nine times the 1,034 cases reported nationally." → *It is clear that the link leads to a preprint.* Also CODE 0: "Other studies have drawn correlations between **cooler climes** and higher transmission rates…" → *Although studies are mentioned*

	in the sentence, it is unclear that this particular link is a study, and not something else.