Toward Better Communication of Uncertainty in Science Journalism

Priyanka Nanayakkara
Jessica Hullman
priyankan@u.northwestern.edu
jhullman@northwestern.edu
Northwestern University
Evanston, Illinois

ABSTRACT

Science journalists must negotiate, assess, and ideally communicate scientific reliability throughout the reporting process. Using interviews with science journalism professionals and an analysis of existing resources for science journalism, we describe current practices used in the reporting process from identifying a story to writing an article. We also conduct a preliminary investigation into how uncertainty terms used in science news writing may differ across quality of articles.

KEYWORDS

uncertainty communication, science journalism, natural language processing

1 INTRODUCTION

Science journalists play a vital role in shaping public understanding of science by brokering information between the scientific community and the general public [48]. Thus, they are indispensable intermediaries whose practiced judgment and expertise enables the public to meaningfully engage with new scientific data. The ways in which journalists frame their findings influences public opinion [22], and can have significant impacts on how the public accepts or rejects new scientific information.

At each step of the reporting process, journalists must filter some information, and distill it such that it is digestible and interpretable to a general audience. They must not only identify relevant and newsworthy scientific studies, but also effectively communicate their findings such that the public understands both the results and limitations. Additionally, journalists are required to interpret and contextualize statistical methods and analyses though they may not have an advanced scientific or mathematical background [32, 48].

Uncertainty deriving from data collection (e.g., random assignment), data transformations, modeling, or missing information affects the results of any experimental scientific study. The relative amount of quantified uncertainty describing a result informs the extent to which it should be accepted as true, if at all. Additionally, the potential for unquantified uncertainty stemming from a lack of information, such as whether the assumptions of a chosen model or measure hold, may mean that quantified estimates of uncertainty are not conservative enough. In reporting on scientific experiments, journalists must first assess and negotiate scientific reliability for their own understanding before articulating their assessment to a general audience.

As evidence of how difficult it can be even for scientists themselves to assess uncertainty in experiment results, consider the replication crisis, wherein large bodies of experimental evidence in psychology have been found to be unreproducible [50]. The crisis has been attributed to a number of factors, including "the frequent use of overly small sample sizes; the widespread reliance on inappropriate (or just plain wrong) statistical procedures; incomplete reporting of experiments; questionable research practices and even (rarely!) outright fraud; publication bias in favor of 'positive' results" [21], and overinterpretation of the significance of single studies [31]. A science journalist's reporting process is made significantly more difficult by the known presence of studies that overestimate effects. In addition, even scientists struggle to properly interpret conventional statistical presentations of experiment results like bar charts with error bars showing confidence intervals [8, 30], and the relationship between study design and the correct interpretation of effects continues to be a topic of scientific debate [6, 17, 38, 47], the quality of the evidence aside. As the scientific community grapples with how to solve the replication crisis [5, 26, 46] and properly interpret statistical evidence, journalists must contend with the severe risk of spreading false scientific information to a wide audience.

While some science journalists may have scientific and statistical training, and consequently sophisticated techniques of contending with uncertainty, we aim to assist science journalists who lack deep scientific or statistical background, the number of which are likely on the rise with cuts to science journalism funding [9], in supporting uncertainty-aware journalism. We draw on literature in how journalists contend with uncertainty and research in uncertainty communication and interpretation by broad audiences to inform our research agenda.

We present observations from four interviews we conducted with professionals in journalism, synthesize existing resources for science journalism, and conduct a preliminary analysis of how expressions of uncertainty may differ between typical and award-winning science articles in a corpus of science news writing.

2 RELATED WORK

Journalists and uncertainty. In a study of TV clips on molecular medicine, Ruhrmann and Milde find that journalists use scientific uncertainty as one of four major ways of framing science news

Computation+Journalism, March 2020, Boston, MA, USA

narratives [44], implying that uncertainty is treated as a topical subject area as opposed to a factor of consideration across all stories. Lehmukhl et al. find that "[o]nly if the major peg on which to hook the audience's interest is a scientific truth claim, journalists deem it relevant to investigate its fragility or ambiguity" [39].

Research describes how when journalists *do* portray uncertainty, they do so for various reasons and in different ways. In some cases, they communicate uncertainty to increase the journalistic quality of their own articles, and to encourage their audiences to think more critically about reported scientific findings [28]. Lekhmul and Peters characterize journalistic uncertainty communication strategies in four ways, two of which are *omission* ("choosing not to publish any descriptions of a scientific truth claim that are perceived as ambiguous or fragile") and *structure and language* (using linguistic tools such as the conditional tense to portray uncertainty) [39].

Understanding uncertainty in experimental studies requires the ability to interpret and contextualize statistical methods and analyses. Yet, researchers have noted how journalists may not always have a scientific or mathematical background [32, 48]. In fact, Dunwoody and Griffin find that journalism administrators largely support the notion that students should receive statistical training, yet many programs don't "require most or all of their students to take a statistical reasoning course within journalism" [20].

Communicating uncertainty. Scientific discourse, which lays the groundwork for evidence-based decision making, is largely a discussion of uncertainty. Common conventions for presenting uncertainty in scientific experiment results include reporting confidence intervals (e.g., Frequentist 95% confidence intervals or Bayesian credible intervals) or other ranges (e.g., plotting a standard deviation or standard error range around a reported effect), reporting standardized measures of effect size like Cohen's D [14, 15], or reporting on the results of Null Hypothesis Significance tests by presenting p-values and test statistics. The level of detail in the scientific literature may however be deemed inappropriate for a decision-maker or lay reader, either because it is too minute to easily decipher or fails to state assumptions that are important but assumed known among scientists [25].

When it comes to whether communicating uncertainty is consistently beneficial to consumers of data-driven estimates, empirical research is mixed. A recent survey of empirical evidence concluded that findings are mixed on the broad question of how communicating quantified, reducible uncertainty impacts audiences [49]. For example, Johnson and Slovic [33, 34] find that presenting numerical ranges signals transparency for a sizable proportion of people, Dieckmann et al. [19] find that decision-makers reported lower levels of blame and rated source credibility higher for forecasts that reported ranges, and Joslyn and colleagues have shown how communicating uncertainty can increase trust in a forecast in a weather domain (e.g., [36]). On the other hand, other studies find no effect or that expressions, such as in the IPCC report on climate change [13], increased interpretation errors [41].

Naturally, the format of uncertainty expression has some influence on its effects. Though verbal expressions of uncertainty may be the most common form in journalism, most research on the effectiveness of verbal expression of uncertainty assumes a relatively constrained scenario in which people's spontaneous impressions of probability from phrases meant to convey probability (e.g., "very likely", "certain") are evaluated. A clear high-level finding from such studies is that people's interpretations of verbal expressions of probability can vary greatly from person to person [10, 11], making verbal expressions less precise than numerical ones.

When it comes to detecting uncertainty language beyond simple probability terms, research in hedge detection aims to automatically detect linguistic hedging, which includes statements that qualify a claim or express uncertainty. As part of The CoNLL-2010 Shared Task, researchers were asked to develop methods of automatically detecting uncertainty cues in both biological publications and texts on Wikipedia [23]. Some teams relied on a sentence classification approach using bag-of-words representations while others attempted to classify tokens as cue phrases and make predictions about sentences based on the occurrence of cue phrases. Similar to the problem of hedge detection in specific cases such as medical literature, NLP methods might be employed to understand the diversity and prevalence of specific expressions of uncertainty used by journalists. (While our goal is not to develop a new hedge detection method, but rather to conduct a preliminary investigation into employed uncertainty expressions, we see hedge detection specifically for a journalism context as future work.)

One instance of text mining of linguistic uncertainty communication corpora, which might provide a valuable "case study" in designing approaches for science news, is the analysis by the Federal Reserve and other central banks of economic-themed news articles to gauge perceived economic risk. For example, the EPU is an index for capturing uncertainty over policy outcomes defined by the co-occurrence of forms of "economic" and "uncertainty" with a policy suggestive term from a small vocabulary (e.g., "regulation") [7]; various other similar text frequency approaches also exist [18].

One insight from economists' attempts to characterize linguistic uncertainty is that dictionaries must be tailored to the context in which they are being applied, as a term's sentiment may depend on context. For example, the word "contained" has a particular importance in conveying low uncertainty in a financial stability context ("risk is contained") but very different meanings outside of this context. Approaches used in economics tend to use standard dictionary creation techniques to identify and track small sets of vocabulary across a corpus, applying sentiment analysis to gauge the relative amount of uncertainty implied at a given time point [16]. Recently, Fast et al. created Empath by using a deep learning approach to generate categories of related words given some seed words [24]. We demonstrate the use of Empath for growing dictionaries of uncertainty language manually identified in science news. Prior research has also successfully used word embeddings to simplify complex scientific terminology [37]. We demonstrate an initial analysis of differences in the prevalence of classes of uncertainty terms in typical versus great examples of science news, aided by using word embeddings to grow categories of terms.

A recent study by van der Bles et al. [49] takes a slightly different approach to understand the impact of text versus numeric communication of uncertainty, comparing point estimates to numerical ranges and verbal expressions of the possibility of uncertainty only (e.g., "there is some uncertainty around this figure"), in a mock news article context. They find that while uncertainty communication in any form leads to greater perceived uncertainty in an estimate, verbal communication had a much greater effect than numerical. While further studies are needed, their results suggest that what a journalist "says" is interpreted as more intentional than numerical expressions, in contrast to common attitudes among data analysts or scientists that more precise, numerical information is preferable.

3 INTERVIEWS & CONTENT ANALYSIS OF GUIDELINES

3.1 Methods

Interviews. We conducted four semi-structured interviews with professionals in journalism. We compiled interview guides in advance, which included questions covering the reporting process from story identification to publication, and general questions pertaining to training and resources for science journalists. We used the interview guide to structure our conversations, but deviated where it seemed fit to allow for discovery of what the journalists found most important or compelling related to our prompts.

Reviewing guidelines. We systematically compiled and reviewed science journalism guidelines from online sources and books looking for salient themes across resources. We identified sources by conducting general web searches for science journalism handbooks and noting which other sources handbooks cited. Additionally, we spoke with a journalism professor at a leading journalism school in the U.S. to get a general sense of modern pedagogy in science journalism.

3.2 Results

Story identification. Multiple interviewees described choosing stories based on their own interests or what their readers would find interesting. Relatively little was said about using the level of uncertainty in a study as an initial criterion for story identification. Guidelines similarly recommend choosing story topics based on how interesting, relevant, and important the work is to a given readership [1, 2, 45]. In addition to interest level, one interviewee noted the significance of a given scientific study as a factor in choosing whether to report on it. Interviewees mentioned using platforms such as EurekAlert!, arXiV, PubMed, and NBER to find studies.

Understanding the science/assessing uncertainty. When reporting on a research study, multiple guides recommend reaching out to experts not affiliated with the study [2, 29, 43]. As one interviewee described, there are instances when an outside expert questions the validity of a study's findings, leaving the journalist to contend with conflicting scientific viewpoints. This tension, however, allows for a more complete representation of the story [2] and can highlight a study's limitations [29]. Multiple interviewees also noted reaching out to the researchers of the study. While they did not all use researcher interviews in place of examining the written study, one interviewee described the practicality of asking the researcher for an explanation of the work: "There's no point in me taking hours to sort through it if I can call up the first author or someone on the paper and have them summarize what they did and what they found."

Additionally, multiple guides recommend considering the study in the context of prior work [4, 12, 45]. One interviewee described tracing a scientific argument by examining a study's citations. This process helps determine the current state of the field. The guides, however, generally do not provide explicit instructions for how to search for and investigate prior work. Rather, they focus on simple directives–"Science builds on science. Know the previous studies that matter so you can paint a fuller picture" [45].

Multiple guides warn against small sample sizes [29, 45], but do not define what constitutes a small sample, either generally or in the context of a specific field or type of study. One interviewee explicitly mentioned checking whether a study's sample size is "high enough." Guides focus not only on investigating sample size, but also on paying attention to the study population. They make a distinction between population versus individual risk [42], and note that results found on non-human populations may not extend to human populations [45]. In practice, one source said they categorically do not report on studies on mice, considering the research to be too early stage.

The Science Writer's Handbook provides the most detailed recommendations in terms of statistical assessment. It draws a distinction between correlation and causation, asks journalists to consider several probing questions related to uncertainty ("Determine the sources of uncertainty in the fields you cover. Do they arise from the tools themselves? What do the experts worry about? What do they criticize each other for?"), be critical of findings with large confidence intervals, and to verify their interpretations with a statistician [29]. Other guides are more vague, recommending that journalists "make sure [they] understand what [the statistics] actually mean, and how certain they are" [2] and make general checks of statistical significance [45]. The lack of statistical resources for science journalists exacerbates the limited statistical reasoning focus in journalism programs, which the science journalism professor we interviewed pointed out.

Writing about scientific findings. While guides recommend using simple, understandable language [4, 29], they also advise against using a tone that is patronizing [42, 43, 45], suggesting a balance between general interpretability and scientific preciseness. To aid in interpretability, some guides suggest carefully using metaphors and analogies [2, 43] and rounding numerical figures [2, 4]. When describing their writing process, one interviewee said they first write drafts with an excess of scientific detail for their own understanding, and then "translate it into less jargon and more of a prose." Another interviewee highlighted the challenge of explaining to an audience the meaning of a word the audience *thinks* they know, but that actually has a different meaning in a certain scientific context ("But when science uses a common word in a way that most people will misunderstand and get exactly the wrong idea, that's where you need to be most careful").

On a broader level, guides recommend logical flow of sentences [27] and paragraphs [2]. In a similar vein, they recommend keeping an article's "stories" in mind [3], a sentiment echoed by one interviewee, who described considering an article's narrative arch throughout the research and reporting process.

Guidance specifically related to uncertainty communication is limited. *The Science Writer's Handbook* recommends basing the extent of uncertainty communication on article length, the audience, and the implications of a study's uncertainty [29]. When verbally incorporating uncertainty, one interviewee mentioned employing the conditional tense and stating the study's venue or peer-review status when introducing its findings.

4 UNCERTAINTY TERMS IN SCIENCE NEWS

4.1 Methods

We collected 30 science news articles from seven well-respected news organizations: *The Guardian*, *The New York Times*, *The Washington Post*, *Vox*, *Gizmodo*, *Science News*, and *Scientific American*. We intentionally chose articles that we believed would use language to convey uncertainty, such as recent articles on scientific studies. Both authors independently tagged the first 10 articles for words and phrases that appeared to be chosen intentionally by the journalist to communicate uncertainty in the study's findings or the scientific process. For example, words such as "suggest" and "could" seem clearly intended to convey uncertainty in summarizing a study's claims or implications. However, some words associated with uncertainty may refer to a finding rather than uncertainty about the finding itself. For instance, the word "likely" in the phrase "Mice lacking REST were also more likely to have busier brains" [35] refers to a study's findings rather than the idea that the findings are tentative, which is expressed through the word "suggests" in the phrase "the study suggests people should prioritize sleep" [35]. Through discussion we reached consensus regarding which tags explicitly communicate uncertainty. The first author then tagged 20 more articles based on the agreed-upon guidelines for selecting terms. We collectively categorized the terms into broad themes such as *association* (e.g., associate, link, related), *evidence* (e.g., sample, measurement, data), and *speculate* (e.g., seems, could, might) (see Table 1). Because our "training corpus" was small, we expanded these categories by feeding small sets of the words to a deep learning system trained on news articles for generating comprehensive dictionaries of similar words [24]. We then found matches for the terms in the typical and great writing categories of a corpus of science journalism articles published in *The New York Times* [40].

We created a score for the prevalence of a particular term in an article based on number of occurrences and total word count, and compared these scores across term categories. Specifically, the score we used was the number of times the term appeared in an article divided by the article's word count divided by 10^4. We used the score primarily to compare prevalence, but the scores are not directly interpretable on their own.

Category	Examples
association	associate, link, related
limitations	doesn't prove, not credible, imperfect
evidence	sample, measurement, data
phrase	if not a reason, shouldn't be taken to mean
publication	under review, not published
replication	further validation, more fieldwork, important step
speculate	seems, could, might, maybe
unclear	spurious, ambiguous, dubious

Table 1: Uncertainty term categories

	Great		Typical	
Category	**mean**	**sd**	**mean**	**sd**
association	1.21	0.809	1.89	1.92
evidence	0.808	1.36	0.803	1.32
limitations	0.481	1.05	0.562	1.32
phrase	0	0	0	0
publication	0	0	0.0594	0.0431
replication	0.058	0.128	0.0612	0.0843
speculate	6.03	7.05	4.06	4.48
unclear	1.03	2.02	0.779	1.24

Table 2: Mean scores across categories

4.2 Results

We found eight major categories of uncertainty terms and phrases (see Table 1) after tagging 30 science news articles. For both the great writing and typical writing articles, the terms from the speculate category were most prevalent. *Could*, *may*, *might*, *seem* (all from speculate) were the most-used words for both writing quality categories.

Discussion. We found there was not much of a difference in categories of uncertainty terms used based on article quality. Based on mean scores per category (see Table 2), it may appear that more associative words are used in lower quality science writing and that better science writing may use words that more directly imply limitations (like words in the speculate and unclear categories), however the standard deviations for these mean values are too large to make confident speculations.

One complication to an analysis like ours is that some words are more common than others irrespective of whether they are being used to intentionally convey uncertainty. For example, words like "could" and "suggest" are likely to be much more prevalent in a corpus unrelated to science writing than words like "published" or "reviewed." Future work might use an external, non-journalism corpus to measure the relative frequency of terms irrespective of their uncertainty status, and use this information to normalize the occurrence of words in a journalism context in order to identify which categories of terms are more unique to a certain quality level of science news articles. Future work might also include further investigating the speculate category and breaking it up into further categories. For example, the category currently includes more generally used words such as "might" and "may" but it also includes less general words such as "hypothesis" and "theorize."

Limitations. Our goal was to do a preliminary investigation into how science journalists are expressing uncertainty and to investigate differences in uncertainty language by article quality. We used a small training corpus, and didn't use part-of-speech or other linguistic cues employed in some work on hedge detection, which may have boosted accuracy in detecting uncertainty expressions. In addition, we did not validate the categories for precision or recall, and see this as valuable future work. Finally, while the training set of articles included recently published articles (within the past couple years), the articles we analyzed were published between 1999 and 2007. It is possible that writing styles have changed; with the visibility of the replication crisis in recent years, it's possible that certain categories of terms that we observed in our training corpus of 30 more recent articles would be better represented in a more recent corpus of great and typical science writing.

5 CONCLUSION

Science journalists face the challenge of informing the public about complex scientific ideas. They must understand such ideas as well as describe them in accessible ways to a general audience. A large part of this process centers on assessing and communicating scientific reliability. An analysis of current guidelines in conjunction with further investigation into the types of uncertainty phrases employed by journalists can provide a starting point for providing systematic assistance to journalists as they contend with a study's reliability, and write about their assessment.

REFERENCES

[1] 2008. How Do Science Writers Get Their Stories? https://casw.org/casw/how-do-science-writers-get-their-stories
[2] 2008. Planning and writing a science story. https://www.scidev.net/global/communication/practical-guide/planning-and-writing-a-science-story.html
[3] 2014. Brown University Science Center's Quick Guide to Science Communication. https://www.brown.edu/academics/science-center/sites/brown.edu.academics.science-center/files/uploads/Quick_Guide_to_Science_Communication_0.pdf
[4] 2019. Chapter 32: Writing about science technology. http://thenewsmanual.net/Manuals%20Volume%202/volume2_32.htm
[5] Samantha F Anderson and Scott E Maxwell. 2017. Addressing the "replication crisis": Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research* 52, 3 (2017), 305–324.
[6] Thom Baguley. 2009. Standardized or simple effect size: What should be reported? *British journal of psychology* 100, 3 (2009), 603–617.
[7] Scott R Baker, Nicholas Bloom, and Steven J Davis. 2016. Measuring economic policy uncertainty. *The quarterly journal of economics* 131, 4 (2016), 1593–1636.
[8] Sarah Belia, Fiona Fidler, Jennifer Williams, and Geoff Cumming. 2005. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods* 10, 4 (2005), 389.
[9] Curtis Brainard. 2019. Science Journalism's Hope and Despair. https://archives.cjr.org/the_observatory/science_journalisms_hope_and_d.php
[10] David V Budescu, Stephen Broomell, and Han-Hui Por. 2009. Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological science* 20, 3 (2009), 299–308.
[11] David V Budescu and Thomas S Wallsten. 1985. Consistency in interpretation of probabilistic phrases. *Organizational behavior and human decision processes* 36, 3 (1985), 391–405.
[12] Khalil Cassimally. 2013. Translating Science: Does It Still Have A Place In Science Communication? https://blogs.scientificamerican.com/incubator/translating-science-does-it-still-have-a-place-in-science-communication/
[13] IPCC Climate Change. 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 2013. *There is no corresponding record for this reference.[Google Scholar]* (2013), 33–118.
[14] Robert Coe. 2002. It's the effect size, stupid: What effect size is and why it is important. (2002).
[15] Jacob Cohen. 1969. *Statistical Power Analysis for the Behavioral Sciences.*
[16] Ricardo Correa, Keshav Garud, Juan M Londono, and Nathan Mislang. 2017. Sentiment in Central Banks' Financial Stability Reports. *Available at SSRN 3091943* (2017).
[17] Peter Cummings. 2011. Arguments for and against standardized mean differences (effect sizes). *Archives of pediatrics & adolescent medicine* 165, 7 (2011), 592–596.
[18] Deepa Datta, Juan M Londono, Bo Sun, Daniel O Beltran, Thiago RT Ferreira, Matteo M Iacoviello, Mohammad R Jahan-Parvar, Canlin Li, Marius Rodriguez, and John H Rogers. 2017. Taxonomy of Global Risk, Uncertainty, and Volatility Measures. *Matteo M. and Jahan-Parvar, Mohammad R. and Li, Canlin and Rodriguez, Marius and Rogers, John H., Taxonomy of Global Risk, Uncertainty, and Volatility Measures (2017-11-21). FRB International Finance Discussion Paper* 1216 (2017).
[19] Nathan F Dieckmann, Robert Mauro, and Paul Slovic. 2010. The effects of presenting imprecise probabilities in intelligence forecasts. *Risk Analysis: An International Journal* 30, 6 (2010), 987–1001.
[20] Sharon Dunwoody and Robert J. Griffin. 2013. Statistical Reasoning in Journalism Education. *Science Communication* 35, 4 (2013), 528–538. https://doi.org/10.1177/1075547012475227
[21] Brian D Earp and Jim AC Everett. 2015. How to fix psychology's replication crisis. *The Chronicle of Higher Education* 62, 9 (2015), B14.
[22] Robert M Entman. 1991. Framing US coverage of international news: Contrasts in narratives of the KAL and Iran Air incidents. *Journal of communication* 41, 4 (1991), 6–27.
[23] Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task.* Association for Computational Linguistics, 1–12.
[24] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 4647–4657.
[25] Baruch Fischhoff and Alex L Davis. 2014. Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences* 111, Supplement 4 (2014), 13664–13671.
[26] Andrew Gelman and John Carlin. 2014. Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science* 9, 6 (2014), 641–651.
[27] George Gopen and Judith Swan. 2018. The Science of Scientific Writing. https://www.americanscientist.org/blog/the-long-view/the-science-of-scientific-writing
[28] Lars Guenther, Klara Froehlich, and Georg Ruhrmann. 2015. Science Communication (Un)Certainty in the News: Journalists' Decisions on Communicating the Scientific Evidence of Nanotechnology. *Journalism & Mass Communication Quarterly* 92, 1 (2015), 199–220. https://doi.org/10.1177/1077699014559500
[29] Thomas Hayden and Michelle Nijhuis. 2013. *The science writers handbook: everything you need to know to pitch, publish, and prosper in the digital age.* Da Capo Lifelong Books.
[30] Rink Hoekstra, Richard D Morey, Jeffrey N Rouder, and Eric-Jan Wagenmakers. 2014. Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review* 21, 5 (2014), 1157–1164.
[31] John PA Ioannidis. 2005. Why most published research findings are false. *PLoS medicine* 2, 8 (2005), e124.
[32] Arnold H Ismach and Everette E Dennis. 1978. A profile of newspaper and television reporters in a metropolitan setting. *Journalism Quarterly* 55, 4 (1978), 739–898.
[33] Branden B Johnson. 2003. Further notes on public response to uncertainty in risks and science. *Risk Analysis: An International Journal* 23, 4 (2003), 781–789.
[34] Branden B Johnson and Paul Slovic. 1995. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk analysis* 15, 4 (1995), 485–494.
[35] Carolyn Y. Johnson. 2019. Excessive brain activity linked to a shorter life. Retrieved Feb 19, 2020 from https://www.washingtonpost.com/science/2019/10/16/excessive-brain-activity-linked-shorter-life/
[36] Susan L Joslyn and Jared E LeClerc. 2012. Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of experimental psychology: applied* 18, 1 (2012), 126.
[37] Yea Seul Kim, Jessica Hullman, Matthew Burgess, and Eytan Adar. 2016. Simple-science: Lexical simplification of scientific terminology. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing.* 1066–1071.
[38] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* 4 (2013), 863.
[39] Markus Lehmkuhl and Hans Peter Peters Forschungszentrum Jülich. 2016. Constructing (un-)certainty: An exploration of journalistic decision-making in the reporting of neuroscience. *Public Understanding of Science* 25, 8 (2016), 909–926. https://doi.org/10.1177/0963662516646047
[40] Annie Louis and Ani Nenkova. 2013. A corpus of science journalism for analyzing writing quality. *Dialogue & Discourse* 4, 2 (2013), 87–117.
[41] Anthony Patt and Suraje Dessai. 2005. Communicating uncertainty: lessons learned and suggestions for climate change assessment. *Comptes Rendus Geoscience* 337, 4 (2005), 425–441.
[42] Andrew Pleasant. 2008. Communicating statistics and risk. https://www.scidev.net/global/disease/practical-guide/communicating-statistics-and-risk.html
[43] Akshat Rathi. 2014. How to avoid common mistakes in science writing. https://www.theguardian.com/science/2014/apr/24/how-to-avoid-common-mistakes-in-science-writing
[44] Georg Ruhrmann, Lars Guenther, Sabrina Heike Kessler, and Jutta Milde. 2015. P U S Frames of scientific evidence: How journalists represent the (un)certainty of molecular medicine in science television programs. *Public Understanding of Science* 24, 6 (2015), 681–696. https://doi.org/10.1177/0963662513510643
[45] Ian Sample. 2014. How to write a science news story based on a research paper. https://www.scidev.net/global/communication/practical-guide/planning-and-writing-a-science-story.html
[46] Jonathan W Schooler. 2014. Metascience could rescue the 'replication crisis'. *Nature News* 515, 7525 (2014), 9.
[47] Uri Simonsohn. 2014. We cannot afford to study effect size in the lab. Retrieved Dec 12, 2019 from http://datacolada.org/20
[48] Debbie Treise and Michael F Weigold. 2002. Advancing science communication: A survey of science communicators. *Science Communication* 23, 3 (2002), 310–322.
[49] Anne Marthe van der Bles, Sander van der Linden, Alexandra Freeman, and David J Spiegelhalter. 2019. The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences of the United States of America* (2019).
[50] Ed Yong. 2012. Replication studies: Bad copy. *Nature News* 485, 7398 (2012), 298.