

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Confidence Levels in Scientific Writing:Automated Mining of Primary Literature and Press Releases

Permalink

<https://escholarship.org/uc/item/4mw6k6qg>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Fox, Will

Donoghue, Thomas

Publication Date

2018

Confidence Levels in Scientific Writing: Automated Mining of Primary Literature and Press Releases

Will Fox (wdfx@mit.edu)

Department of Mathematics, Massachusetts Institute of Technology, USA

Thomas Donoghue (tdonoghue@ucsd.edu)

Department of Cognitive Science, University of California San Diego, USA

Abstract

Scientific communication includes primary scientific literature written by and for scientists, as well as press releases written about these scientific articles that are used to inform the popular press. By the time new scientific findings are reported by the press, the reporting can often reflect 'spin', or reporting that minimizes uncertainties and exaggerates impact, as compared to the original study. In this work, we examine the role that the press release may play in communicative change, in particular with respect to differences in portrayed confidence between abstracts of scientific articles and press releases. We examine a large corpus of over 15,000 documents collected from online databases covering a range of scientific topics, leveraging automated analysis tools from natural language processing to examine how the readability, sentiment, subjectivity, and portrayed confidence varies between primary literature and press releases. We find that press releases are often easier to read, portray more positive sentiment, use language that implies greater objectivity, and demonstrate higher confidence in the findings. Future work should focus on examining if these differences between press releases and primary articles do indeed engender different perceptions in readers.

Keywords: text-mining; science communication; natural language processing; web-scraping; automated analysis

Introduction

The communication of results is a key component of the scientific process, both within the scientific community and to the general public. Scientists primarily share their findings through scientific articles, with a target audience of academic peers. Press releases are often written about these new scientific publications by institutions aiming to share the results with a broader audience. Based on these press releases and/or the primary literature, science journalists may then write articles about the research in the press.

Although science journalists need not source their material primarily from press releases, it has been shown that press releases heavily influence the way science journalism proceeds. Press releases are typically produced for studies that are expected to be popular in the media (Styker, 2002). When press releases are written, they often have a significant impact on the information reported in the media. For example, a recent case study analysis showed that the press release was one of three sources that accounted for 85% of the content written about a particular article, with many sources using almost verbatim text from the press release (Taylor et al., 2015). Despite their importance, press releases have been shown to vary in quality and often neglect to mention important limitations in a way that may

warp the perception of the importance of the findings (Woloshin & Schwartz, 2002).

Press releases may therefore have a large impact on the public's perception and knowledge of science by influencing news coverage framing the discussion and interpretation of primary articles. Notably, the mass media is the primary source of information by which the public stays up to date in science. Public perception, knowledge, and interest in science is correlated with support for science funding (Besley, 2016). Additionally, the degree to which a study is covered by the popular press also influences how much other researchers hear about the study, at least in the medical community (Phillips et al., 1991). Having research appropriately covered in the popular press is therefore important for propelling future research.

Given the importance of press releases in influencing how novel scientific findings are reported in the media, the key question of interest here is to examine how the framing and discussion of science differs between the primary literature and press releases. In particular, we investigate whether press releases emphasize the confidence in and importance of novel findings in a way that may ultimately influence public perceptions of science. Specifically, the hypothesis is that when writing about scientific studies, press releases may inflate the confidence in and importance of the findings. This bias in press releases could then propagate to the news media that covers the study and ultimately influence public perceptions of science. Recent work has indeed shown that press releases that contained 'spin', defined as using reporting practices that emphasize the beneficial aspect of the research, were associated with news releases also containing 'spin' (Yavchitz et al., 2012).

Here the focus is on bias or 'spin' that is introduced after the primary scientific article is written. It is a separate question of whether scientists themselves are over-interpreting their results, as has been shown in some cases (Menachemi et al. 2013; Lazarus et al., 2016). The assumption here is that the peer review system for primary literature publication helps to reduce spin by scientists, and there is some empirical support for this (Lazarus et al., 2016). In comparison, press releases are written by press officers with a wide range of editorial oversight, including often having no oversight (Woloshin & Schwartz, 2002). The approach here is therefore to consider primary literature as 'ground truth', and investigate how communication changes after initial publication in ways that potentially influence subsequent interpretation of the science.

Recent analyses have found that there is exaggeration in press releases, as compared to primary articles, and that this

is associated with subsequent exaggeration in news sources (Sumner et al., 2014, 2016). These prior investigations were, however, restricted to subsets of medical literature and utilize labor-intensive methods including manually coding specific components of how press releases report on journal articles. Although this is a high fidelity method for characterizing these documents, the practical implications of using such laborious approaches severely limit the number of articles that can be investigated.

In the present work, we therefore seek to extend this line of inquiry using automated, computational methods developed in the fields of text-mining and natural language processing (NLP). Although coarser and noisier, the benefit of such approaches is that they can be applied across a much larger set of scientific literature. Here we seek to establish a proof-of-concept report of applying NLP methodology across scientific literature and press releases. These methods are scalable, and can be used to examine patterns across topics, institutions, and/or journals at a scale that would be prohibitive using manual coding methods.

NLP techniques have been utilized for many years as a way of automatically parsing and characterizing text data. Broadly, most of these techniques take as input a text document and calculate weighted sums of specific words and/or other document features, in order to output a score reflecting the metric of interest. For example, readability scores often consist of a weighted sum of average word and sentence length. These existing metrics have been validated with respect to the properties that they measure. In this study, we use measures for readability, sentiment and subjectivity which are already well-studied.

In addition to the aforementioned existing metrics, a key goal is to measure something like 'spin', for which there are no existing automated tools. To approximate this, we developed an NLP method to measure the portrayed confidence of documents, and then compare between literatures, to see how confidence levels change from journal paper abstracts to press releases. The method adapted to assess confidence in this study is called Linguistic Inquiry and Word Count (LIWC). LIWC counts word occurrences based on predetermined dictionaries of terms and computes results based on frequency of words from particular categories (Pennebaker, Francis & Booth, 2001). Variants of this method can be used to measure different characteristics of writing, for example 'clout', which refers to status, confidence or leadership that people portray through writing (Kacwicz et al., 2013). To apply this method to measure portrayed confidence, we curated a novel corpus of high confidence words, indicating conclusiveness, and low confidence words, indicating hedging or uncertainty, which allows for automated analysis of the confidence or lack thereof in a writing sample.

There are of course expected differences between press releases based on the distinct audience, context and purpose of the writing. The key question, therefore, is not whether there are differences, but specifically whether there are differences in press releases that could influence

interpretation in a way that could potentially go on to generate systematic distortions of the public's understanding of science. In terms of the employed measures of readability, sentiment, and subjectivity, we hypothesize that primary literature abstracts will have lower scores, reflecting more complexity, less affective tone, and lower subjectivity as compared to press releases. With regards to whether press releases and paper abstracts engender differential levels of confidence in the results that are being described, we hypothesize that as compared with paper abstracts, where writing is tuned to a knowledgeable scientist audience and the final text is moderated by the peer review process, press releases will over-emphasize the confidence in and importance of the research at hand.

Methods

Data Collection

Data, consisting of abstracts from peer-reviewed primary scientific articles and press releases from press offices of research institutions, were collected from open-access databases. To select documents from these sources, a selection of search terms were chosen to cover topics from across cognitive science, neuroscience, and the biomedical sciences, including clinically focused topics which may have more salience for the general public. We launched searches on these databases specifying a publication date within the calendar year of 2017. All documents returned by these searches were collected and used for this investigation. Search terms, including the number of documents collected, are described in Table 1.

The dataset of primary scientific literature was collected from PubMed, an open database of primary biomedical literature. We launched searches through Pubmed's EUtils¹, a public application programming interface (API) for the Pubmed database, which allows for programmatic querying and extraction of information from the database. For each term, approximately a thousand documents published in the year 2017 were gathered. For each result, collected data included the title, authors, journal, digital document identifier (DOI), pubmed ID, and the complete abstract text.

Data for press releases were collected with a similar procedure using EurekAlert², a free online database for scientific press releases. The same terms and time period were used to search EurekAlert. Searches were limited to include only research-based press releases. For each search result, the data gathered includes the title of the press release, the institution that published it, and the full text of the press release. Where available, data was also collected on the journal, conference, funder, and region for the research, as well as links to the original posting of the press release, and to the referenced article.

In total 5659 press releases, and 9913 abstracts were collected, across the 11 search terms, as described in Table 1. Note that this procedure does not directly match papers and press releases written about the same individual study.

¹ <https://eutils.ncbi.nlm.nih.gov/>

² <https://www.eurekalert.org>

Analysis reported here are collapsed across all topics, though the patterns of results reported are also broadly maintained within each topic specifically.

Table 1: Terms & Number of Documents

Term	# of Press Releases	# of Abstracts
Autism	321	956
Cognition	346	966
Consciousness	81	954
Dementia	391	927
Diabetes	1694	870
Epilepsy	201	877
Hypertension	337	906
Optogenetics	53	606
Perception	701	980
Stroke	1078	910
Vaccines	456	960

Data Analysis

All documents were analyzed across four dimensions: readability, sentiment, subjectivity, and portrayed confidence. Data analysis was done using the Python programming language, using existing open source tools, including the natural language toolkit (Loper & Bird, 2002), as well as custom analysis scripts.

Readability gives a measure of the how difficult it is for a reader to understand a writing sample, where a lower score indicates a sample is harder to read. Readability metrics are calculated using weighted sums over measures including average syllables per word, average letters per word, and/or average words per sentence. We focused on the Flesch-Kincaid reading ease measure, a weighted score of word and sentence length developed for technical documents (Kincaid et al., 2014). We also computed other related metrics, including the SMOG Index and the Linsear-Write formula, which showed the same pattern (results not reported here).

Sentiment analysis refers to the procedure of automatically scoring the affective state portrayed by a text document, analyzing them in terms of the opinions, evaluations, attitudes, and emotions conveyed (Liu, 2012). Common methods, which we apply here, score texts based on word use, as defined by pre-existing dictionaries of words that have been validated as reflecting positive, neutral, or negative states. For this analysis, the VADER lexicon was used (Gilbert, 2014), with results also corroborated when calculated using the Liu-Hi Lexicon (Liu, 2012).

Subjectivity analysis is a subset of sentiment analysis, distinct from the positive/negative classification discussed above. Rather, subjectivity refers to the extent to which a text appears to reflect a subjective state or opinion, as compared to presenting objective information. For this analysis, a classifier was trained on an existing available database that contains labeled examples of subjective and objective sentences (Pang & Lee, 2004) using a Support Vector Machine. The input features to the classifier were tokenized words for each labelled sentence. On a held out test set of the labelled corpus, this classifier achieved 89% accuracy (chance level: 50%). For our analysis, each

sentence of each document was classified using this pre-trained classifier, with the overall document score consisting of the number of subjective sentences minus objective sentences, such that a negative score reflects a document with an overall objective stance, whereas a positive score reflects a document with a more subjective stance.

A custom procedure was developed to measure portrayed confidence using the linguistic inquiry and word count (LIWC) method (Pennebaker, Francis & Booth, 2001). To apply LIWC for confidence ratings, a custom corpus of terms was created, with terms scored as reflecting either high or low confidence with respect to a claim being made. This corpus was initially compiled and scored by the authors, and then qualitatively evaluated by an external group for consistency to determine the inclusion and score of individual terms. In total, this dictionary included 100 terms, scored for portrayed confidence. Table 2 includes examples of scored terms. Using this dictionary, the LIWC method was used to count occurrences of these confidence-related words across each document, assigning a score to each document.

Table 2: Examples of High & Low Confidence Words

High Confidence Words	Low Confidence Words
conclusively	appears
confirms	could
definitively	may
demonstrates	possibility
establishes	seems
proves	suggests

Preprocessing procedures included removing any documents that received outlier scores, defined as being outside of four standard deviations of the distribution of measures within their group for a given metric. This procedure resulted in 156 documents being dropped from further analysis. Re-running the analyses with these data points included did not significantly change the pattern of results. Paper abstracts and press releases also tended to differ in terms of their length. To ensure this difference was not driving the results, all analyses were also performed while explicitly controlling for document length, and it was verified that there were no significant changes in the results.

Results

To investigate whether the distinct modes of scientific communication (primary literature versus press releases) engender different levels of confidence in the conclusions of a given study, metrics covering readability, sentiment, objectivity and portrayed confidence were calculated across collected documents. We hypothesized that, when compared to primary scientific abstracts, press releases would be found to be easier to read, show higher levels of positive sentiment, include more subjective language, and demonstrate higher confidence in the conclusions discussed.

Results presented reflect comparisons between document types, collapsed across all topics, though findings were also broadly consistent for every measure within every topic.

The four measures employed display significant but small magnitude correlations with each other, as described in table 3, suggesting that each captures unique variance in the data. For each measure, averages, as the mean value, and standard deviation (std) were calculated. Bartlett's test for equal variances was used to assess differences in variances between document type. Since all measures had significantly different variances, the Welch's independent samples t-test was used to assess for significant differences in the means, as it does not assume equal population variances. Effect sizes were calculated with Cohen's-d.

Table 3: Pearson Correlations Between Metrics

	Readability	Sentiment	Subjectivity	Confidence
Readability	-	0.022*	-0.063**	0.093**
Sentiment	0.022*	-	0.032**	0.110**
Subjectivity	-0.063**	0.032**	-	-0.109**
Confidence	0.093**	0.110**	-0.109**	-

* $p < 0.01$; ** $p < 0.001$

In terms of readability, the Flesch-Kincaid Reading Ease average for paper abstracts across all terms was 29.77 (std: 14.79), reflecting a reading grade level of a college graduate. The average readability score for press releases was 39.74 (std: 9.65), reflecting a reading grade level of a college student. These readability measures between documents types were statistically significantly different ($t(9915)=50.05, p<0.001$), with an effect size of 0.79.

The average sentiment score for paper abstracts across all terms was 0.33 (std: 1.64), where positive scores indicate positive affect, suggesting that the primary literature tends to be relatively neutral, including positive and negative language and style in approximately equal amounts. Comparatively, the average sentiment score for press releases across all terms was 1.17 (std: 3.30), indicating a trend towards a more positive affect in these documents. The relatively high variance reflects in part that different topics tended to have relatively different sentiment scores, though it was consistent across each topic that press releases were more polar (further from a sentiment score of 0) than paper abstracts. Across all topics, these group differences between document type were significantly different ($t(9915)=17.60, p<0.001$), with an effect size of 0.32.

The average subjectivity score for paper abstracts across all terms was -1.01 (std: 3.58), wherein more negative scores reflect more objective sentences (and positive scores reflect more subjective sentences) indicating that primary literature tends to be relatively neutral, including subjective and objective language and style in approximately equal amounts. The average subjectivity score for press releases across all terms was -3.54 (std: 5.56), indicating that this type of writing tends towards using more objective language as compared to abstracts. These group differences between document type were significantly different ($t(9915)=30.40, p<0.001$), with an effect size of 0.54.

The average confidence score for paper abstracts across all terms was 2.45 (std: 3.78), whereby more positive results

indicate more high confidence terms used. For press releases, the average was 4.39 (std: 6.67), indicating that this writing tends to be more confident and conclusive, as compared to the primary literature that appears to convey uncertainty more frequently. These group differences between document type were significantly different ($t(9915)=20.11, p<0.001$), with an effect size of 0.36.

We also examined some simple descriptive analyses of the press releases, finding that only 45.4% of press releases included a link to their primary article, and only 36.0% contained a link back to the original press release from the press office of the institution who wrote and published it.

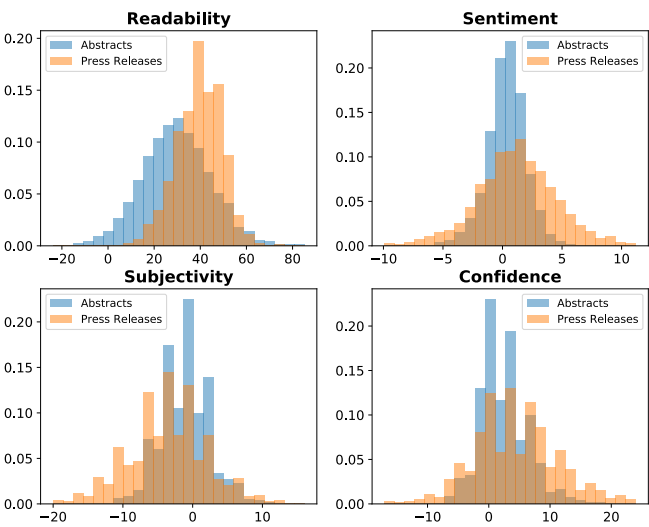


Figure 1: Distributions of metric results, across all topics, separated by document type (abstracts vs. press releases).

Discussion

In the above computational analysis of writing style, we found systematic differences in the features of paper abstracts and press releases, showing press releases to have a lower reading level, more positive sentiment, more objectivity, and higher confidence in their findings.

One primary difference between abstracts and press releases is audience, and this distinction is perhaps best demonstrated by the readability measure. Press releases are designed for consumption by scientific journalists rather than technically-trained scientists themselves, and the analysis found that press releases had a somewhat higher readability score, supporting this hypothesis. This difference in readability may also indicate that press releases may reduce the complexity and perhaps nuance which is conveyed by the scientific paper. Notably, press releases do still have a high reading level, though this is consistent with them serving as a middle ground, before further translation for the news media.

Another important distinction is that press releases are often intended to convey not only the results, but also the significance, impact, and valence of a scientific study. Because of this, they can be hypothesized to be likely to take a more polar tone in order to communicate the potential

impact of a study. The analyses here support this hypothesis. Whereas a paper abstract is designed to describe the specific details and outcomes of a study, press releases are more likely to provide context and extrapolate, taking on more of an affective tone, reflected by the sentiment scores.

A somewhat surprising outcome resulted from the subjectivity analysis. In this case, the data refuted the hypothesis, suggesting press releases contain a higher proportion of objective sentences to subjective sentences as compared to paper abstracts. This is consistent with both document types being overall objectively framed, though press releases seem to be even more so than abstracts. This may relate to the information contained in the different documents, as much of the typically longer press releases could be dedicated to providing context to journalists who are not presumed to have field-specific training. This context may be largely factual and objective, with perhaps a relatively small amount of the press release containing more subjective content. Other possible explanations for this phenomenon include that writing about one's own work (in abstracts) may include subjective descriptors, and/or that increased subjective components in abstracts could reflect hedging or uncertainty of the author. The current analysis did not offer the resolution to parse out these more nuanced details, making this an avenue for future work.

Finally, the confidence analysis found press releases to be notably more confident in their conclusions than primary scientific literature, supporting the original hypothesis. As the confidence corpus reflects, these press releases tend to contain fewer words that would indicate hedging, uncertainty, or inconclusiveness, instead using words that highlight the conclusiveness and importance of the findings.

Taken together, these results align closely with the hypotheses of this experiment. The two types of writing give rise to distinct features and styles due to their expected audience and function, though it is important to note that both serve primarily to communicate scientific findings. The measured differences, however, demonstrate a more critical contrast between the two - press releases on average and across fields tend to show more confidence and conclusiveness in the findings as compared to their primary literature counterparts over a number of linguistic features. The analyses described above indicate that press releases may not convey as much uncertainty, while taking on a more positive tone and approach to the literature at hand. This is consistent with press releases containing more 'spin' (Sumner et al., 2014, 2016), which may lead to misconceptions by journalists and the public as to the confidence, and uncertainty of scientific results.

This finding about confidence is important as it suggests that the known influence of press releases on the news media (Yavchitz et al., 2012) may serve to give a biased view of science to the lay public, who may be more likely to view science as the collection of well described, reliable facts, as opposed to a constantly evolving process. If the public has a view of science as being more like a collection of facts, this could create cognitive dissonance with the frequent reporting of incompatible results between different scientific studies. Scientists may be more aware of science as an evolving process, with a body of knowledge that is

constantly being refined, and at times even re-written, based on new information. The public, by comparison, may often build a conception of science using news media which over-emphasizes the confidence and importance of findings. When inconsistent findings are reported, this could potentially lead to confusion and ultimately a loss of confidence in the scientific process itself.

A consistent finding across all employed measures is that the variance for each measure was significantly larger for press releases than for paper abstracts. The increased variance in press releases may reflect the relative lack of editorial oversight in press releases (Woloshin & Schwartz, 2002), allowing for more idiosyncratic writing, as compared to the peer-review process for primary literature.

We also examined some simple descriptors of how often press releases link to the primary literature that they are describing and found a disappointingly low number, with less than half of press releases doing so. This is a problem in of itself, as press releases and news media that become disconnected from the scientific studies upon which they are reporting make it harder for readers to refer back to the primary literature.

Limitations

There are several limitations to the study. In the data collection procedure employed, primary articles and press releases are not directly linked to each in a 1-1 manner, confining analyses to group comparisons between topics. The search procedures also do not ensure that selected documents are primarily about the searched-for topic, and may merely mention them. All of this serves to add some 'noise' to the analysis. Similarly, we have only coarse measures by virtue of the relatively simplified methods employed here. These methods focus primarily on word distributions, and as such are rough measures of complex concepts. Nevertheless, this approach is automated and scalable allowing for examination of patterns across much larger datasets in a manner complementary to hand-coding.

Though press releases do have a large impact on the popular press (Taylor et al., 2015), they are still a step removed from material that the public actually reads, and so further work is needed to ascertain both how the reported findings in press releases influence the popular press and what impacts this may have on the public. In particular, further work must investigate what effects the language used in science reporting actually has on readers. Though we might expect that the communication of confidence and/or uncertainties in science may influence readers beliefs, some work has shown that communication of uncertainties does not necessarily change beliefs about the nature of science (Retzbach & Maier, 2015).

Papers and press releases are also difficult to systematically compare on a specific properties such as confidence because of their distinct purposes and writing styles. Each fulfills a different role in the process of scientific communication causing the two to take on slightly different forms, though both are primarily informative. Press releases are often longer than paper abstracts and written to inform a less scientific audience about both the outcome and significance of a study. Although we controlled for

differences such as word length, differences in the document type and purpose are a potential confound to keep in mind, in terms of ensuring that the employed measures faithfully reflect the desired aspects being investigated, and don't simply reflect trivial differences related to document format.

Future Work

This work sought to demonstrate that automated procedures such as those employed can be used at scale to further investigations in a research area that has traditionally relied on more manual procedures. Further work is still required to further validate these approaches. In particular, further work should be done to validate the custom dictionaries curated here, perhaps using norming studies that validate the portrayed confidence of the chosen words, and by an analysis that compares the present methods directly to a manually scored subset, to investigate the degree to which these methods are consistent, and/or how they may diverge.

There are also numerous other investigations that can be run on this data, including using other analysis metrics from natural language processing, extending the analyses to include more topics, as well as a potentially fruitful avenue that would seek to specifically connect individual press releases to the primary articles they are written about, as well as to news articles to examine specific instances, rather than relying on the group approach used here.

Conclusion

The present study sought to examine differences between the primary scientific literature and press releases, specifically in terms of how differences in writing and style may relate to portraying the confidence in and importance of the work being communicated. We tested the hypothesis that institutional press releases exaggerate the confidence in and importance of a given study compared to primary scientific writing. Using automated analyses across a large dataset, we found that press releases score higher than paper abstracts in readability, sentiment, and confidence, and lower in subjectivity. This suggests that press releases hedge less frequently and reduce uncertainties which may be conveyed in primary scientific writing. This potentially influences subsequent reporting by journalists which could provide the lay public with misconceptions regarding the scientific process. Future work should expand on these analyses by further validating and improving the method for confidence, comparing results to hand-coded analyses, exploring popular press articles to get a better understanding of the entire scientific communication process, and investigating the impacts of writing style on interpretation.

Code & Data Availability

Project code and the collected dataset are available at:
<https://github.com/wdfox/ConfidenceScanner>

Acknowledgments

We would like to thank Sean Trott, Shuai Tang, Scott Cole, Sydney Smith, Federico Rossano, Bradley Voytek and the reviewers for helpful commentary and feedback.

References

- Besley, J. C. (2016). The National Science Foundation's science and technology survey and support for science funding, 2006–2014. *Public Understanding of Science*.
- Bird, S., & Loper, E. (2004). NLTK: the natural language toolkit. *Association for Computational Linguistics*.
- Gilbert, C. H. E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International Conference on Weblogs and Social Media*.
- Kacewicz, E., Pennebaker, J.W., Davis, M., Jeon, M., & Graesser, A.C. (2013). Pronoun use reflects standings in social hierarchies. *Language & Social Psychology*.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas for navy enlisted personnel (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch.
- Lazarus, C., Haneef, R., Ravaud, P., ... & Boutron, I. (2016). Peer-reviewers identified spin in manuscripts of non-randomized studies assessing therapeutic interventions, but their impact on spin in abstract conclusions was limited. *Journal of Clinical Epidemiology*.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining: Synthesis Lectures on Human Language Technologies, vol. 16. *Morgan & Claypool Publishers, San Rafael*.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71, 2001*.
- Phillips, D. P., Kanter, E. J., Bednarczyk, B., & Tastad, P. L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *New England Journal of Medicine, 325(16), 1180-1183*.
- Retzbach, A., & Maier, M. (2015). Communicating scientific uncertainty: Media effects on public engagement with science. *Communication Research, 42(3), 429-456*.
- Stryker, J.E. (2002). Effects of press releases and newsworthiness on medical journal articles visibility in the news media. *Preventive Medicine, 35(5)*.
- Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., ... & Boy, F. (2014). The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ, 349*.
- Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Bott, L., ... & Chambers, C. D. (2016). Exaggerations and caveats in press releases and health-related science news. *PloS one, 11(12), e0168217*.
- Taylor, J. W., Long, M., Ashley, E., Denning, A., ... & Wojtowicz, A. (2015). When Medical News Comes from Press Releases—A Case Study of Pancreatic Cancer and Processed Meat. *PloS one, 10(6)*.
- Woloshin, S., & Schwartz, L. M. (2002). Press releases: translating research into news. *JAMA, 287(21)*.
- Yavchitz, A., Boutron, I., Bafeta, A., Marroun, I., Charles, P., Mantz, J., & Ravaud, P. (2012). Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med, 9(9), e1001308*.