Title: Supplemental methods appendix for “Misinformation more likely to use non-specific authority references: Twitter analysis of two COVID-19 myths.”
Authors: Joseph McGlynn (1), Maxim Baryshevtsev (2), Zane A. Dayton (3)
Date: September 4th, 2020
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix: Supplementary Methods

Twitter search query

We limited our search to English tweets only and variations of the term “COVID-19”. Specifically, the API queries we used were:

Gargle Myth

lang:en (gargle OR gargling)
(covid19 OR coronavirus OR covid-19 OR c19 OR cov OR covid_19 OR covid)

Hold Breath Myth

lang:en hold (breath OR breathe) seconds
(covid19 OR coronavirus OR covid-19 OR c19 OR cov OR covid_19 OR covid)

Data analysis

To test whether authority appeals predicted tweet engagement, we fit a negative binomial model. The negative binomial model accounted for the large variance (i.e., overdispersion) of retweet and like counts. Part of the overdispersion was caused by one extreme outlier (i.e., one tweet with greater than 10,000 likes), which caused the model fit to be unreliable. As a result, we restricted the number of likes to fewer than 10,000. Control variables for the analysis included whether or not the account was verified and the number of followers for the account.

Coding process and creation of the codebook

With regard to the coding of the tweets for content analysis, the first two authors coded the first 300 tweets collaboratively (approximately 20% of the dataset). During the initial coding of the tweets, a codebook was developed to create categories, definitions, and examples for each code. Upon creation of the codebook, the first 300 tweets were then recoded to account for any potential changes in coding decisions that may have developed throughout the coding process. This step of recoding helps to promote stability (Krippendorff, 2004) and ensure the coding decisions did not change over time. The first two researchers then divided the remainder of the dataset evenly and coded the tweets individually. If one of the researchers was uncertain on how to code a particular tweet, the two researchers discussed the coding of the tweet collaboratively. All disagreements were discussed until agreement was met. This process created the set of coded tweets for the analysis.

Intercoder reliability

To assess intercoder reliability, a researcher blind to the purpose of the study was trained using the codebook created for the study. Reliability between coders was then assessed through an open-source calculation tool for reliability analysis (Freelon, 2010). As our data included two sets of observations and binary coding categories, scores for Scott’s pi, Cohen’s kappa, and Krippendorff’s alpha were asymptotically equal (Hayes & Krippendorff, 2007). We used percent agreement and Krippendorrf’s alpha as a guideline for acceptable reliability (Hayes & Krippendorff, 2007). Percent agreement was above 90% for each category, and Krippendorff's alpha indicated acceptable levels of reliability, ranging from .70 to .89 (Krippendorff, 2004). Krippendorff's alpha accounts for the possibility of chance agreements, which is a reason why a coding category with a higher percent agreement may have a lower Krippendorff's alpha. Table 4 illustrates the reliability statistics for each tweet category. The codes assigned to each tweet are available in the replication materials. The codebook created for the current study is also available in the replication materials. The codebook includes descriptions of the coding categories, definitions for each category, and example tweets for each coding category.

Table 4. *Intercoder reliability statistics for the coded tweets (N = 1,493)*

Category of Tweet	Percent Agreement	N Agreements	N Disagreements	Scott's π	Cohen's κ	Krippendorff's α
Irrelevant	93.64%	1398	95	0.77	0.77	0.77
Misinformation	92.63%	1383	110	0.85	0.85	0.85
Debunking – Overt	94.98%	1418	75	0.89	0.89	0.89
Debunking – Suspicion	94.31%	1408	85	0.88	0.88	0.88
Authority Appeal	90.82%	1356	137	0.79	0.79	0.79
Non-specific Authority Appeal	96.45%	1440	53	0.70	0.70	0.70
Specific Authority Appeal	93.77%	1400	93	0.83	0.83	0.83