Health Communication

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hhth20

Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis

Nathan Walter , John J. Brooks , Camille J. Saucier & Sapna Suresh

To cite this article: Nathan Walter , John J. Brooks , Camille J. Saucier & Sapna Suresh (2020): Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis, Health Communication, DOI: 10.1080/10410236.2020.1794553

To link to this article: https://doi.org/10.1080/10410236.2020.1794553

HEALTH COMMUNICATION
https://doi.org/10.1080/10410236.2020.1794553

Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis

Nathan Walter, John J. Brooks, Camille J. Saucier, and Sapna Suresh

Department of Communication Studies, Northwestern University

ABSTRACT

Social media poses a threat to public health by facilitating the spread of misinformation. At the same time, however, social media offers a promising avenue to stem the distribution of false claims – as evidenced by real-time corrections, crowdsourced fact-checking, and algorithmic tagging. Despite the growing attempts to correct misinformation on social media, there is still considerable ambiguity regarding the ability to effectively ameliorate the negative impact of false messages. To address this gap, the current study uses a meta-analysis to evaluate the relative impact of social media interventions designed to correct health-related misinformation (k = 24; N = 6,086). Additionally, the meta-analysis introduces theory-driven moderators that help delineate the effectiveness of social media interventions. The mean effect size of attempts to correct misinformation on social media was positive and significant (d = 0.40, 95% CI [0.25, 0.55], p =.0005) and a publication bias could not be excluded. Interventions were more effective in cases where participants were involved with the health topic, as well as when misinformation was distributed by news organizations (vs. peers) and debunked by experts (vs. non-experts). The findings of this meta-analysis can be used not only to depict the current state of the literature but also to prescribe specific recommendations to better address the proliferation of health misinformation on social media.

Social media have become a prominent part of the U.S. health-care process, as more and more Americans use it as a primary source for health information. From peer-to-peer psychosocial support and alerts about the spread of infectious disease to seeking and sharing information about life-changing diets and ways to contact clinics and hospitals, social media platforms such as Twitter and Facebook are ubiquitous hubs for health information. Findings reported by Pew Research indicate that seven-in-ten U.S. adults seek health information online and a growing number of social media users post about health matters, join health-related groups, and follow updates regarding health (9–23%) (Patel et al., 2015; Pew Research Center, 2014). At the same time that social media can facilitate support, complement offline information, and empower patients, the deluge of false and misleading information on these platforms constitutes a substantial threat to public health (Smailhodzic et al., 2016). Beyond confusion and reduced trust in health professionals, unfiltered exposure to misinformation can "delay or prevent effective care, in some cases threatening the lives of individuals" (Y. Wang et al., 2019, p. 1).

Although this danger has previously been exemplified through the growth of the anti-vaccine movement (Broniatowski et al., 2018), the proliferation of naturopathic cancer treatments (Chen et al., 2018), and misleading rumors about the spread of infectious disease (Bode & Vraga, 2018), the duality of social media has never been more apparent than during the COVID-19 pandemic. With shelter-in-place and lockdown orders, social distancing guidelines, and other public health measures disrupting our daily lives, the COVID-19 pandemic has made social media platforms a source of connection as well as confusion. As social media use has continued to grow since the early days of the virus (Hanson, 2020), there have been partnerships between public health organizations and social media companies to get reliable and accurate information to their users. For example, Facebook has updated its interface to feature a COVID-19 info tab and stepped up its content review and fact-checking efforts (Jin, 2020), while celebrities, influencers, and other public figures have taken to social media to spread informative messages of solidarity and amuse the isolated populace (Desaulniers, 2020). However, this boom in social media activity has also accelerated the spread of misinformation. Monthly traffic to Snopes, one of the foremost fact-checking organizations, has risen sharply since the pandemic began, but their small staff has been overwhelmed by the "infodemic" such that they cannot keep up with the deluge of submissions (Leskin, 2020). This work can be a matter of life or death: rumors about hydroxychloroquine and other "cures" (e.g., colloidal silver, oregano oil, sunlight) have spread rapidly online with sometimes fatal consequences (e.g., Trew, 2020).

The growing danger of health-related misinformation on social media has spurred efforts to develop and test various interventions designed to ameliorate the spread of falsehoods and sustain public trust in evidence-based care (Chou et al., 2018). Specifically, interventions have employed a variety of strategies designed to debunk misinformation, including algorithmic corrections (Bode & Vraga, 2015) and user-generated credibility ratings (Pennycook & Rand, 2019). Although each of those strategies has shown some potential in attenuating the impact of misinformation, studies continue to generate

CONTACT Nathan Walter nathan.walter@northwestern.edu Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA
Supplemental data for this article can be accessed here.

conflicting results, ranging from cases where social media is successfully used to correct misinformation (e.g., Vraga & Bode, 2017; 2018) to interventions that backfire and amplify unhealthy tendencies (e.g., Nyhan & Reifler, 2010). Thus, despite a growing number of empirical studies that attempt to correct misinformation, we are far from knowing when and how to best intervene.

The current meta-analysis

The next step in determining how to respond when warranted should involve a synthesis of current data for the purpose of generating general principles and best practices. To date, no systematic efforts have been made to quantify the overall efficacy of attempts to correct health-related misinformation on social media. Although previous meta-analyses that focused on the correction of misinformation included health-related topics (Walter & Murphy, 2018), these studies have only limited utility when trying to address health-related misinformation on social media. First, effects retrieved from studies that explicitly examined health misinformation were combined with effects from a variety of other contexts, including politics and marketing. While this approach can be justified when trying to understand the overall effect of corrective information, health-related misinformation and its correction are not always comparable with misinformation in other contexts. For instance, when directly comparing health and political misinformation, Walter and Murphy's (2018) meta-analysis concluded that attempts to correct health misinformation appear to be more successful compared to those that seek to address political misinformation. This happens, presumably, because motivated reasoning, in-group protective behavior, and confirmation bias pose a greater barrier when challenging political beliefs compared to health-related beliefs. Since all other analyses did not distinguish health from other contexts, it is difficult to determine what other factors (e.g., message and source characteristics) contribute to the observed differences. In addition, past meta-analyses offer little information with respect to the role played by social media in helping spread, as well as tag and debunk health-related falsehoods. Acknowledging that many individuals use social media as their primary source of health information (Chou et al., 2018) and that these platforms differ from other media in authorship, oversight, and algorithms (Southwell et al., 2019), assessing the efficacy of corrections on social media can go a long way in developing a proactive approach to fighting misinformation.

In order to address these challenges, we systematically collected and analyzed causal evidence regarding the correction of health-related misinformation on social media, considering the potential moderating role of sample (type, geographical region, issue-involvement) and message characteristics (topic, source of misinformation, correction type, source of correction, format, and social media platform). Drawing on the World Health Organization (2020) definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity," the current study adopts a broad understanding of health misinformation. Hence, the current meta-analysis brings together topics that fit the understanding of health as physical well-being (e.g., infectious disease, obesity, reproductive health), as well as mental and social well-being (e.g., nutrition and developmental health).

In sum, the current study offers a meta-analytic perspective on some of the most essential benefits, challenges, and opportunities that social media presents to public health by assessing the average effect of interventions and possible factors underlying their success. The findings of such an effort will be critical for policymakers, practitioners, and researchers alike as they aim to remedy some of the negative outcomes produced by social media. Thus, a meta-analytic approach to the existing literature may be particularly helpful in determining the best way to address the challenges presented by health misinformation.

Method

Search strategy, selection criteria, and data extraction

Relevant studies were obtained using the following procedures: (a) Seven electronic databases (i.e., *Communication Source, Educational Resources Information Center, JSTOR, Medline, ProQuest, PubMed*, and *Web of Science*) were searched from June 2019 to August 2019, with the following Boolean search strings "misinformation AND health AND social media OR Facebook OR Twitter OR YouTube," "disinformation AND health AND social media OR Facebook OR Twitter OR YouTube," "conspiracy theor* AND health AND social media OR Facebook OR Twitter OR YouTube," "literacy AND health AND social media OR Facebook OR Twitter OR YouTube," "correction AND health AND social media OR Facebook OR Twitter OR YouTube," and "retraction AND health AND social media OR Facebook OR Twitter OR YouTube." The same search strings were used in August 2019 with Google Scholar to identify additional articles. More information about the specific search strings and the number of results is included in the appendix; (b) Reference lists of included studies were manually searched; and (c) Seven experts in the field of health misinformation and social media were contacted to review the list of included studies, identify omissions, and share unpublished data. This search strategy generated 110 (after removal of duplicates) research reports, which were then assessed for eligibility against the following inclusion criteria: (a) Studies had to include an attempt to correct health-related misinformation; (b) Studies had to employ experimental designs where participants were randomly assigned either to receive corrective information or to a no-correction condition (either misinformation-only or neutral control[1]); (c) Study material had to explicitly name a social media platform as the place where the correction attempt occurred; and (d) Studies had to measure the influence of correction on attitudes, behavioral intent, or behavior. When studies failed to report on appropriate statistics (e.g., t-values, means, standard deviations, counts, frequencies, zero-order correlations, exact p-values), relevant information was successfully obtained from the corresponding authors ($k = 3$). All research reports were independently reviewed for potential inclusion in the meta-analysis and disagreements were resolved through discussion. Adhering to this screening process, 19 research reports that documented the results from 24 individual studies were included in the meta-analysis (~21% unpublished), with a total sample size of 6,086 (M = 253.58, Med = 221, SD = 148.86) (see Figure 1 for a PRISMA flow diagram).

Type of intervention

To assess the effectiveness of health misinformation correction on social media platforms, the current meta-analysis specifically examines randomized experiments that compared attitudes, intentions, and behaviors of individuals who were exposed to misinformation that subsequently was corrected and those who were exposed to the misinformation but not to its correction, or to neutral control conditions. This contrast allows to determine the extent to which social media interventions designed to correct misinformation can influence people's health-related decision-making.

Coding of outcomes and moderators

Two independent coders were trained on a subset of studies that did not meet all the inclusion criteria until satisfactory level of agreement was reached (*kalpha* > .80). Then, reliability was calculated on 50% of the actual dataset, resulting in agreement ranging from .82 to 1.00. A single effect size of exposure to the correction of health-related misinformation was calculated per sample. Following recommendations by Borenstein et al. (2009), in cases where studies reported on several relevant outcomes (e.g., Kim, 2019), all effect sizes were recorded and then averaged into a single outcome (for data in support of the invariance of attitudes, intent, and behavioral outcomes, O'Keefe, 2015). Likewise, in cases where studies employed multiple types of corrections in the same study (e.g., Smith & Seitz, 2019), all relevant effect sizes were retrieved from the study and then averaged for the analysis.

Given the focus of the current meta-analysis on experimental designs that directly compare two conditions, reported effects from primary studies were transformed to Cohen's d, allowing easy interpretation of directionality and magnitude. Each study was coded for the following moderating characteristics: outcome type (attitude/intent/behavior), sample type (college students/general adults), geographical region (U.S./other), social media platform (Facebook/Twitter), correction format (text/text and image), and health topic.

Additionally, each study was coded based on the source of the misinformation, distinguishing between messages that attributed the misinformation to a news agency (for example, Washington Post in Smith & Seitz, 2019) or to other private social media users/peers[2] (for example, mother of a kindergartner in Gesser-Edelsburg et al., 2018). Using the distinction proposed by van

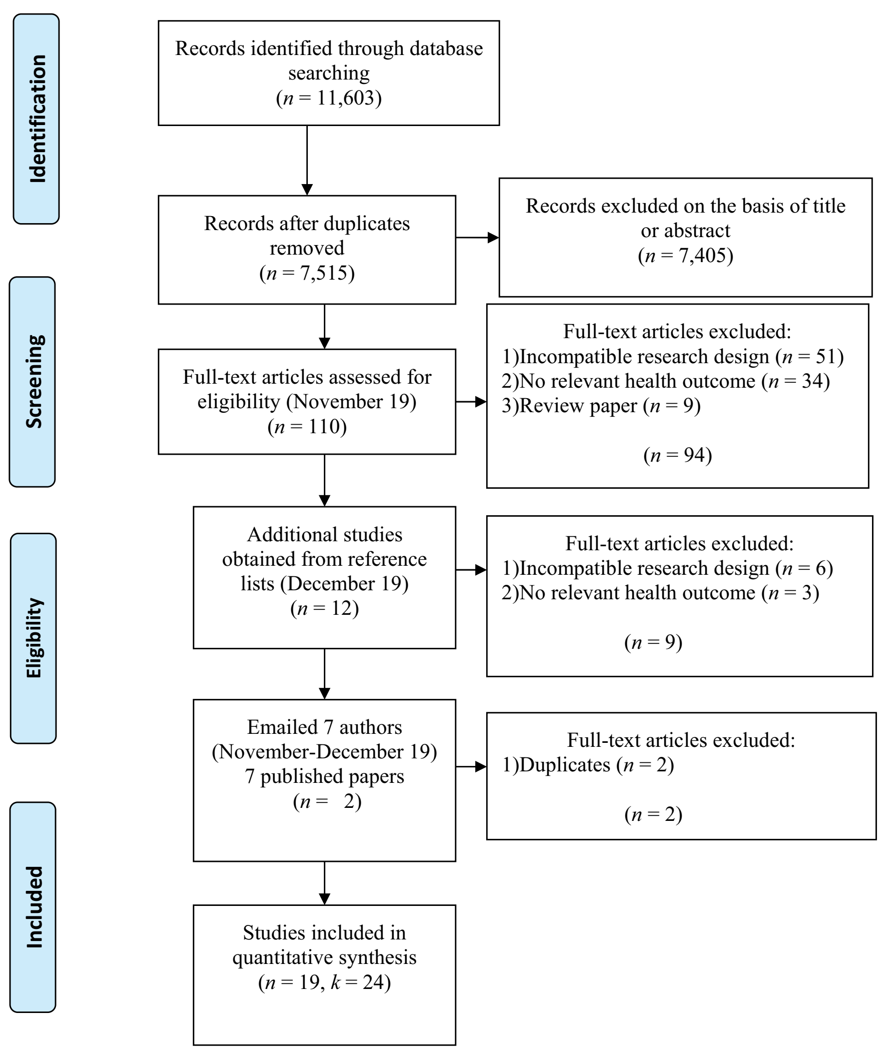

Figure 1. Flow diagram for identifying experimental studies reporting on attempts to correct health misinformation on social media.

der Meer and Jin (2020), correction attempts were classified as either factual elaboration, negating the misinformation and reinforcing the correct facts (for example, debunking misconceptions surrounding the flu vaccine in Sullivan, 2019) or simple rebuttals, negating the misinformation without reinforcing the correct facts (for example, simply indicating that the CDC did not find a link between GM mosquitoes and Zika in Bode & Vraga, 2018). Further, the source of correction was identified as either an expert referring to health professionals/official health agencies (for example, CDC in Bode & Vraga, 2015) or a source without an evident expertise in health (for example, anonymous social media user in Lee, 2019). Finally, samples were coded as highly involved with the health-related topic if the study explicitly suggested that participants were chosen due to their relevance to the topic (for example, public health professionals in Gesser-Edelsburg et al., 2018). In cases where the study did not indicate the rationale for recruiting the specific sample, studies were coded as having low-involvement (see appendix for a complete version of the coding sheet).

To protect against potential violations of independence of effect sizes (for an explanation, see Borenstein et al., 2009), all moderators were coded at the level of the study. Notably, there are different approaches to handle dependent effect sizes (e.g., Cheung, 2014), such as cases where the same study includes different treatments that are associated with distinct effect sizes. For example, in the current meta-analysis, Vraga and Bode (2017) used corrections associated with experts and non-experts while Vraga and Bode (2018) employed corrective information delivered through Facebook and Twitter. Yet, given the limited number of studies that reported on several relevant effect sizes, only one effect size per study was selected, which addressed concerns regarding interdependence (Cooper, 2010). In particular, the guiding principle was that when an overall estimate of the pooled effects was required, all relevant effect sizes within studies were averaged. When moderation analyses involved effect sizes of specific subgroups within studies, only those effects were used. If studies included subsamples associated with several different values of the same moderator (e.g., both expert and non-expert correction in Vraga & Bode, 2017), the study was removed for that particular moderation analysis. Keeping in mind that each study contributed only one effect size, the pooled effects were always calculated using independent samples (Cheung, 2014). Based on recommendation by Jackson and Turner (2017), moderation tests were performed only when the data included at least five cases for each level of the moderator (see Table 1 for an overview of studies included in the meta-analysis).

Table 1. Study and intervention features.

Study	Year	Study number	N	Platform	Correction source	Misinformation source	Misinformation topic	Primary outcome
Bode & Vraga[adfhk]	2015	1	99	Facebook	Expert	News agency	GMO myths	Belief in the link between GMOs and illness
Bode & Vraga[aehk]	2015	2	156	Facebook	Expert	News agency	GMO myths	Belief in the link between GMOs and illness
Bode & Vraga[adfhk]	2018	1	136	Facebook	Expert	News agency	Infectious disease	Belief in misinformation about Zika
Du et al. [ad]	2019	1	257	Other	Non-expert	-	Obesity misinformation	Credibility of debunked report on obesity rates
Gesser-Edelsburg et al.[adgil]	2018	1	228	Facebook	Expert	Peer	Infectious disease	Belief in the accuracy of reports on a measles outbreak
Hoel[aehijl]	2019	2	157	Other	Expert	-	Nutrition myths	Belief in food safety myth
Kim[abefgijl]	2019	3	400	Twitter	-	Peer	Infectious disease	Belief in anti-vaccination claims
Kim et al.[aefhijk]	2017	1	392	Facebook	-	News agency	Reproductive health	Believability of misleading information about Planned Parenthood
Lee[aehjl]	2019	1	205	Facebook	Non-expert	Peer	Infectious disease	Belief in disinformation regarding a "fatal virus"
Lee[aehjl]	2019	2	214	Facebook	Non-expert	Peer	Infectious disease	Belief in disinformation regarding a "fatal virus"
Li & Sakamoto[aegk]	2014	1	122	Twitter	-	-	Health myths	Truth judgments regarding various health rumors
Li & Sakamoto[cegk]	2014	2	101	Twitter	-	-	Health myths	Truth judgments regarding various health rumors
Ozturk et al.[cegk]	2015	1	104	Twitter	Non-expert	-	Health myths	Truth judgments regarding various health rumors
Pal et al.[begk]	2019	2	206	Other	Expert	-	Health myths	Intention to share expert denials of false rumors
Smith & Seitz[aefhk]	2019	1	280	Facebook	Expert	News agency	Neuroscience myths	Trust in false statements about neuroscience
Sullivan[aefhl]	2019	1	603	Facebook	-	Peer	Infectious disease	Belief in misinformation about the seasons flu vaccine
Tully et al.[aeh]	2020	1	241	Twitter	Non-expert	News agency	GMO myths	Credibility of false Tweet about GMOs
Tully et al.[aeh]	2020	2	303	Twitter	Non-expert	News agency	Infectious disease	Credibility of false Tweet about the flu vaccine
van de Meer & Jin[befg]	2020	1	700	Facebook	-	News agency	Infectious disease	Perceived severity of a fictitious infectious disease
Vraga & Bode[adfgk]	2017	1	308	Twitter	-	News agency	Infectious disease	Belief in misinformation about Zika
Vraga & Bode[adfgk]	2018	1	271	-	-	News agency	Infectious disease	Belief in misinformation about Zika
Vraga et al. [aegil]	2019	1	135	Twitter	Non-expert	Peer	Infectious disease	Belief in misinformation about the HPV vaccine
Weber et al.[abefgk]	2019	1	129	Facebook	Non-expert	Peer	Infectious disease	Belief in misinformation regarding childhood vaccines
Weber et al.[abdfgk]	2019	3	339	Facebook	Non-expert	Peer	Infectious disease	Belief in misinformation regarding childhood vaccines

[a]attitudes; [b]intent; [c]behavior; [d]college students; [e]general adults; [f]U.S. sample; [g]text; [h]text and image; [i]high-involvement; [j]unpublished; [k]simple rebuttal; [l]factual elaboration.

Quality and bias assessment

The quality of each study was appraised with the study quality assessment tools utilized by the National Heart, Lung, and Blood Institute (National Heart, Lung and Blood Institute [NHLBI], 2019), developed specifically to evaluate the rigor of experimental studies in the context of public health. Each study was assessed according to their performance on a 7-item checklist, encompassing the following criteria: randomization, blinding, baseline measures, attrition, intervention, outcome measures, and power calculation. Quality assessment coding was conducted by three independent raters (*kalpha* = .84). Further, the findings were subjected to several different publication bias tests, including Egger's regression test (Egger & Smith, 1998), Begg and Mazumdar's ran correlation test (Begg & Mazumdar, 1994), and generation of a funnel plot with effect sizes plotted against a corresponding standard error. Moreover, to evaluate the stability of the findings, we performed a "leave-one-out" sensitivity analysis. This approach allows to assess the relative influence of individual studies by examining changes to the pooled-estimate in the absence of each study.

Data analysis

Individual effect sizes, averaged effect sizes, homogeneity statistics, moderation analyses, sensitivity analysis, and publication bias tests were calculated using Comprehensive Meta-Analysis (version 3) software (Borenstein et al., 2015). Random effects models were employed in the current study, which allowed generalizing the results beyond the specific population of studies being analyzed (Hedges & Vevea, 1998). Heterogeneity was assessed with Cochrane's Q and I^2 statistics. To explore potential causes of significant heterogeneity, we undertook Q-test subgroup analysis. The appendix presents the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Liberati et al., 2009).

Results

The mean effect size for correction of health-related misinformation on social media was positive and significant (d = 0.40, 95% CI [0.25, 0.55], p = .0005, k = 24), with substantial heterogeneity, $Q(23)$ = 198.655, I^2 = 88.42%, p = .0005 (see Figure 2 for a forest plot that includes the retrieved effect size, p-values, and 95% confidence interval for each study). Notably, there was no significant difference ($Q(1)$ = 0.71, p = .401) between effect sizes retrieved from studies that examined influence on attitudes (d = 0.37, 95% CI [0.20, 0.53], p = .0005, k = 20) and behavioral intent (d = 0.24, 95% CI [−.01, 0.49], p = .062, k = 5).

Moderation results

In terms of sample characteristics, the analysis did not record a significant difference between subgroups based on sample type (college students/general adults; $Q(1)$ = 1.37, p = .242) and geographical region ($Q(1)$ = 0.01, p = .992). Yet, there was a significant difference ($Q(1)$ = 3.14, p = .039) between lowly involved (d = 0.31, 95% CI [0.26, 0.36], p = .0005, k = 18) and highly involved samples (d = 0.63, 95% CI [0.34, .97], p = .001, k = 6), such that interventions designed to correct health-related misinformation were more successful for the latter.

The results indicated that the source of misinformation emerged as a significant moderator ($Q(1)$ = 4.46, p = .001). Specifically, it is more challenging to correct misinformation when it is delivered by our peers (d = 0.24, 95% CI [0.11, 0.36], p = .0005, k = 8) as opposed to news agencies (d = 0.48, 95% CI [0.15, 0.81], p = .001, k = 10). Also, the source of the correction played a significant role ($Q(1)$ = 4.36, p = .031), resulting in stronger effects when corrective messages delivered by experts (d = 0.42, 95% CI [0.28, 0.55], p = .0005, k = 7) compared with non-experts (d = 0.24, 95% CI [0.13, 0.34], p = .0005, k = 9). With regard to the type of the corrective format, there was no significant difference between corrections that addressed the misinformation with a simple rebuttal and those that

Study name	Statistics for each study			
	Std diff in means	Lower limit	Upper limit	p-Value
Weber, Muehling, & Kareklas (2019) - Study 1	-0.14	-0.49	0.21	0.43
Vraga & Bode (2017)	0.08	-0.15	0.30	0.50
Kim, J. (2019)	0.11	0.01	0.21	0.02
Bode & Vraga (2015) - Study 1	0.16	-0.24	0.56	0.44
Du, Zhu, & Cheng (2018)	0.18	-0.13	0.48	0.26
Lee (2019) - Study 2	0.18	0.04	0.31	0.01
Tully, Vraga, & Bode (2019) - Study 1	0.21	-0.04	0.47	0.10
Sullivan (2019)	0.22	0.02	0.42	0.03
Lee (2019) - Study 1	0.22	-0.06	0.49	0.12
Smith & Seitz (2019)	0.24	0.00	0.47	0.05
Weber, Muehling, & Kareklas (2019) - Study 3	0.24	0.03	0.46	0.03
Vraga & Bode (2018)	0.25	0.01	0.50	0.04
Tully, Vraga, & Bode (2019) - Study 2	0.29	0.06	0.52	0.01
Hoel (2019) - Study 2	0.34	0.01	0.67	0.04
Bode & Vraga (2015) - Study 2	0.45	0.12	0.77	0.01
Pal, Chua, & Goh (2019) - Study 2	0.49	0.15	0.84	0.01
Vraga, Kim, & Cook (2019)	0.52	0.16	0.88	0.00
Li & Sakamoto (2014) - Study 1	0.53	0.17	0.89	0.00
Bode & Vraga (2018)	0.55	0.19	0.90	0.00
Van de Meer & Jin (2019)	0.60	0.38	0.81	0.00
Ozturk, Li, & Sakamoto (2015)	0.62	0.23	1.01	0.00
Gesser-Edelsburg et al. (2018)	0.65	0.38	0.92	0.00
Li & Sakamoto (2014) - Study 2	0.82	0.41	1.23	0.00
Kim, Moravec, & Dennis (2017)	1.98	1.70	2.26	0.00
	0.40	0.25	0.55	0.00

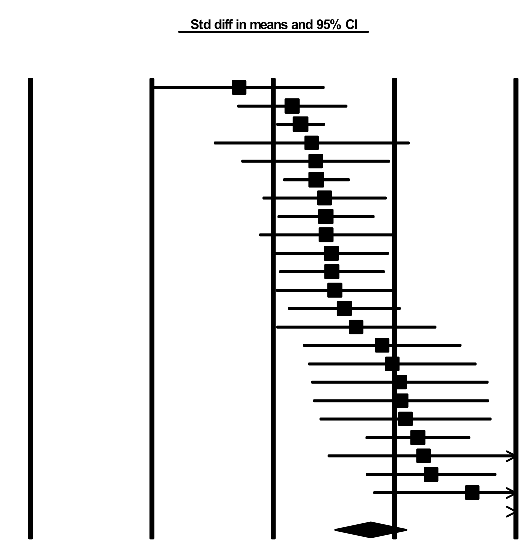

Figure 2. Effects of exposure to correction regarding health-related misinformation on social media (Overall effect).

elaborated on the information ($Q(1) = 1.47$, $p = .225$). Further, there was no significant difference ($Q(1) = 0.14$, $p = .713$) between text messages and information that combined text and images. Additionally, there was no significant difference ($Q(1) = 0.07$, $p = .787$) between interventions that employed Facebook rather than Twitter. Finally, the results suggest that it is more difficult ($Q(1) = 3.17$, $p = .014$) to correct misinformation in the context of infectious disease ($d = 0.28$, 95% CI [0.17, 0.39], $p = .0005$, $k = 13$) as opposed to other health-related issues ($d = 0.55$, 95% CI [0.31, 0.79], $p = .0005$, $k = 11$). Table 2 outlines the findings of the moderation analysis.

Examination of quality assessment and publication bias

Assessment of the 24 studies included in the meta-analysis indicated high risk of bias among 11 (45.8%) and unclear risk among 13 (54.2%) studies. In particular, 11 (45.8%) studies did not assess baseline measures prior to the intervention, making it impossible to gauge whether the groups were similar at baseline. Further, 4 (16.7%) studies suffered from attrition that exceeded 20% and 1 (4.2%) study was statistically underpowered. Additionally, 4 (16.7%) studies included several interventions in each condition thus raising concerns over possible confounds. Likewise, although all studies employed adequate randomization, none of the studies used double-blind procedures such that research personnel was aware of the intervention and group allocation. Yet, this threat is somewhat mitigated by the fact that both the interventions and the outcome measurements were administered online rather than in-person (see online supplement for summary of quality assessment).

The funnel plot of effect sizes against their standard error (see appendix) showed a substantial clustering of studies in the upper left part of the funnel, indicating a potential underrepresentation of studies with stronger effects. Additionally, both Egger's test for asymmetry ($p = .047$) and Begg and Mazumdar's rank correlation test ($p = .018$) reached significance, indicating a possible publication bias. Notably, in all cases, the recorded direction of bias suggested that available studies are likely to underestimate the benefit of interventions used to correct health misinformation on social media. On the one hand, this finding may alleviate concerns regarding a possible "file-drawer problem" in the current sample of studies. Indeed, there is no evidence to suggest that studies with null findings or weak effects are less likely to be published in this particular literature. On the other hand, the fact that the sample of studies suffers from underrepresentation of stronger effects is also concerning since it reduces the certainty in evidence (Murad et al., 2018). Finally, the sensitivity analysis indicated that the pooled-estimate was fairly stable (see appendix), ranging from $d = 0.31$, 95% CI [.23, .40], $p = .0005$ (when removing Kim et al., 2017) to $d = 0.42$, 95% CI [0.27, 0.57], $p = .0005$ (when removing Weber et al., 2019). Taken together, the examination of quality assessment and the potential publication bias point to the need to enhance research standards through better-powered designs and preregistration.

Discussion

This study reports on a meta-analysis that estimated the efficacy of correcting health-related misinformation on social media. To better understand the causal relationship between exposure to corrective messages on social media and subsequent effects on health-related outcomes, we restricted the meta-analysis to experimental designs that attempted to debunk health misinformation. The findings indicate that, on the whole, correction can successfully mitigate the influence of misinformation. Importantly, the pooled-effect retrieved in this study was comparable with previous meta-analyses that

Table 2. The effects of correction by intervention characteristics.

Variable	d	k	N	Q	p	95% CI
Main effect	0.40	24	6086			[.25,.55]
Sample type				1.37	.242	
College students	0.29	7	1638			[.14,.45]
General population	0.44	17	4448			[.25,.64]
Geographical region				0.01	.992	
U.S.	0.39	11	3657			[.11,.67]
Other	0.39	13	2429			[.27,.50]
Issue-involvement				3.14*	.039	
High	0.63	6	1312			[.34,.97]
Low	0.31	18	4517			[.26,.36]
Misinformation source				4.46***	.001	
News agency	0.48	10	2886			[.15,.81]
Peer	0.24	8	2253			[.11,.36]
Correction source				4.36*	.031	
Experts	0.42	7	1262			[.28,.55]
Non-experts	0.24	9	1927			[.13,.34]
Correction type				1.47	.225	
Factual elaboration	0.28	7	1942			[.15,.41]
Simple rebuttal	0.48	13	2643			[.19,.78]
Correction format				0.14	.713	
Text	0.38	12	3043			[.22,.53]
Text and image	0.44	11	2786			[.15,.72]
Platform				0.07	.787	
Facebook	0.44	13	3481			[.19,.70]
Twitter	0.40	8	1714			[.27,.55]
Topic category				3.17*	.014	
Infectious disease	0.28	13	3971			[.17,.39]
Other	0.55	11	2115			[.31,.79]

Note. *p < .01; **p < .05; ***p < .001

focused on the correction of misinformation in contexts such as crime, politics, and science (Walter et al., 2020; Walter & Murphy, 2018). Not only that but the results present no evidence of the so-called "boomerang" or backfire effect, whereby attempts to correct misinformation can unintentionally increase people's acceptance of the falsehood. With less concern about inadvertently perpetuating misinformation, the public health community should feel encouraged to pursue further (and possibly more extensive) corrective efforts on social media.

Since the average effect of correction is of weak-moderate magnitude, it is especially prudent for this analysis to highlight the significant moderators of corrective interventions that may enhance the efficacy of such efforts. Simply put, a more complete understanding of the circumstances under which misinformation is most effective – as well as the conditions that will facilitate successful correction – will be necessary to successfully act against health misinformation. As one might expect, the source of the misinformation makes a difference. Misinformation shared by peers is more challenging to root out than misinformation from news organizations. This conforms to our understanding of how social ties influence information processing: individuals are more likely to trust information provided by those who we perceive to be similar to us, even when they do not possess medical expertise (Z. Wang et al., 2008). Because we are more likely to perceive information sharing as an act of goodwill rather than persuasion from these sources, we tend to be less skeptical of the content. The source of the correction is another important consideration: for health misinformation, correction from experts is more effective than correction from non-experts. Despite recent concerns about dwindling belief in the value of expert opinion and low trust in official organizations (Ortiz & Rosenthal, 2019), it appears that their efforts to contradict inaccurate health information are more likely to be believed than those from peers or other inexpert sources. This finding is consistent with previous studies that demonstrated the ability of expert organizations like the CDC to independently rebut misinformation on social media (Vraga & Bode, 2017). As such, their clout can be deployed to address serious threats through official corrections, which can have considerable reach on social media.

The moderation analysis did not identify a significant difference between simple rebuttals and factual elaborations. Although this finding somewhat deviates from previous studies where factual elaborations outperformed simple rebuttals (van der Meer & Jin, 2020), it is important to situate this contradiction within the scope of the current meta-analysis. Following the literature on the continued influence effect and mental models (Walter & Tukachinsky, 2020), one can expect simple rebuttals to be more effective when people have relatively limited knowledge on the topic (no pre-established mental models) whereas cognitive elaborations are more likely to reduce misconceptions when they provide contextual information that both debunks the falsehood and fills gaps in existing mental models. To this end, a more accurate test for this hypothesis should involve not only a manipulation of correction type (simple rebuttal vs. factual elaboration) but also belief certainty (certain vs. uncertain).

Unsurprisingly, the content of the misinformation is another critical consideration. Some misinformation topics are of broader interest (or perhaps more urgent necessity) than others, and therefore more likely to hold one's attention – in this case, we found that infectious disease presented a special category of misinformation that was particularly resistant to correction. The effects of myths about genetically modified produce, nutrition, and reproductive health were more effectively attenuated by corrective interventions than misinformation about Zika virus, measles, HIV, and other communicable diseases. The diseases featured in social media rumors often have life-threatening outcomes; the perception of severity, in tandem with feelings of susceptibility tied to their infectious nature, can motivate an individual to attend to a message more closely (van der Meer & Jin, 2020) – even if that message is inaccurate. However, the greater levels of engagement that can help misinformation stick can also facilitate its correction: Our findings indicate that individuals are more receptive to correction for issues in which they are more involved. In essence, the more one cares about an issue, the more likely they are to continue to update their knowledge of an issue, particularly when a factual correction is offered.

Strengths and limitations

The current study has a number of strengths. In particular, data collection and coding were done systematically, yielding a large corpus of studies that also included unpublished data. Relatedly, the decision to focus on experimental designs has contributed to our ability to discuss the influence of corrections in terms of causality as opposed to mere association. Likewise, the inclusion of various sensitivity tests and quality checks has contributed to more robust inference procedure. Despite these important strengths, the current study is not without limitations. One of the most significant obstacles to our work is the limited number of studies focusing on health misinformation on social media. Notably, only recently have a sufficient number of studies been published to allow for meta-analytic study – our findings are heartening, but there is still much that is unknown about the interaction between social media and false health content that could be answered with further research. For instance, additional studies are needed to fully address the question of multiple sources on social media, as the same information can be attributed to the person sharing it, the organization they are sharing it from, and the friends who like the content. Relatedly, the findings are limited to the type of comparison conditions included in the meta-analysis. Although our original plan was to include comparisons between misinformation conditions and the neutral control conditions (e.g., Weber et al., 2019), there were not enough studies in these categories for a statistically meaningful analysis. While such comparisons provide important insights, they appear to be less common when addressing misinformation on social media. Perhaps, the misinformation-only condition is less common in this literature because many social media corrections occur at real-time (e.g., Bode & Vraga, 2018) or through tagging of misinformation (e.g., Li & Sakamoto, 2014) thus exposure to misinformation and its correction occurs at the same stage as opposed to two separate events. In a similar vein, the relatively small number of studies limited our findings to broader categorical organization when evaluating moderating factors. Studies focusing on discrete health issues will allow for more granular comparisons between topics, and studies examining the efficacy of corrective methods between platforms may offer more tailored strategies for addressing misinformation. Consequently, future work is a matter of both quantity and quality: advancing our knowledge of this subject will require an expanded corpus of high-quality studies with diverse approaches to message design.

Implications

By identifying the circumstances under which misinformation is most resistant or susceptible to correction, our findings can assist public health organizations and advocates in developing successful interventions. Addressing every item of misinformation distributed on social media would be both time-consuming and a drain on resources (Southwell et al., 2019), so the conclusions of this paper may help these actors develop more targeted strategies for addressing the myriad rumors on social platforms. Further, the finding that infectious disease rumors were more resilient to correction than other kinds of health myths highlights the potential for social media to complicate potential public health crises. As evidenced by the early days of the COVID-19 pandemic, it is crucial that public health agencies develop media strategies to address misinformation quickly and efficiently.

Conclusion

Although there is still much to be learned, the current study's results are cause for optimism. The vast majority of corrective interventions are at least somewhat successful in diminishing the impact of misinformation, and our findings regarding moderating factors should inform future research into designing effective countermeasures. The continued efforts of the broader research community will only further refine our understanding of best practices to address the threat presented by health misinformation on social media.

Notes

1. Rather than directly comparing exposure to correction of misinformation with a misinformation-only control condition, three studies in the sample (Bode & Vraga, 2018; Vraga & Bode, 2017; 2018) compared the effect of exposure to correction of misinformation with a neutral control condition. Similarly to fact-checking messages, these studies introduced the false statements and then immediately debunked them. When studies did not include a no-correction condition, the contrast between exposure to misinformation and exposure to correction was assessed with within-subject effects focusing on participants attitudes/intentions/behavior after exposure to misinformation and then following its correction (e.g., Study 3 in Kim, 2019; Kim et al., 2017; Study 2 in Lee, 2019).
2. When the source of misinformation was a news agency (e.g., Washington Post) but the message was shared by a private social media user (e.g., random person's Newsfeed), the misinformation was attributed to the news agency (e.g., Bode & Vraga, 2015). This decision was based on the fact that (a) the true source of the message was the news agency and (b) the social media user was unfamiliar to the participants in the study.

References

An asterisk precedes references included in the meta-analysis.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

*Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, *65*(4), 619–638. https://doi.org/10.1111/jcom.12166

*Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140. https://doi.org/10.1080/10410236.2017.1331312

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2015). *Comprehensive meta-analysis*. (version 3.0) Biostat. https://www.meta-analysis.com/

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley-Blackwell.

Broniatowski, D. A., Jamison, A. M., Qi, S. H., Alkulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *Health*, *108*(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Chen, L., Wang, X., & Peng, T. Q. (2018). Nature and diffusion of gynecologic cancer-related misinformation on social media: Analysis of tweets. *Journal of Medical Internet Research*, *20*(10), e11515. Article. https://doi.org/10.2196/11515

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Chou, W. S., Oh, A., & Klein, W. M. P. (2018). Addressing health-related misinformation on social media. *JAMA*, *320*(23), 2417–2418. https://doi.org/10.1001/jama.2018.16865

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Sage.

Desaulniers, S. (2020, April 10). *How Instagram influencers are trying to slow the spread of coronavirus*. CNBC. https://www.cnbc.com/2020/04/10/coronavirus-instagram-influencers-try-to-slow-covid-19-spread.html

*Du, Y. R., Zhu, L., & Cheng, B. K. L. (2019). Are numbers not trusted in a "post-truth" era? An experiment on the impact of data on news credibility. *Electronic News*, *13*(4), 179–195. https://doi.org/10.1177/1931243119883839

Egger, M., & Smith, G. D. (1998). Meta-analysis bias in location and selection of studies. *BMJ*, *316*(7124), 61–66. https://doi.org/10.1136/bmj.316.7124.61

*Gesser-Edelsburg, A., Diamant, A., Hijazi, R., & Mesch, G. S. (2018). Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. *PLoS One*, *13*(12), e0209505. Article. https://doi.org/10.1371/journal.pone.0209505

Hanson, A. (2020, April 2). *How social networks are surging and changing during the COVID-19 outbreak*. Business 2 Community. https://www.business2community.com/social-media/how-social-networks-are-surging-and-changing-during-the-covid-19-outbreak-02297720

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

*Hoel, E. (2019). *Funny myth busting: Are humorous messages effective at correcting misinformation regarding food safety?* [Unpublished master's thesis]. Handelshøyskolen BI.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods*, *8*(3), 290–302. https://doi.org/10.1002/jrsm.1240

Jin, K.-X. (2020, June 10). *Keeping people safe and informed about the coronavirus*. Facebook Newsroom. https://about.fb.com/news/2020/04/coronavirus/

*Kim, A., Moravec, P., & Dennis, A. R. (2017). Behind the stars: The effects of news source ratings on fake news in social media. *Kelley School of Business Research*, *18*(3), 1–38.

*Kim, J. W. (2019). *Countering anti-vaccination rumors on Twitter* (Publication No. 22619077) [Doctoral dissertation], Syracuse University. ProQuest Dissertations and Theses Global.

*Lee, J. (2019). *Correcting fear-arousing disinformation on social media in the spread of a health virus: A focus on situational fear, situational threat appraisal, belief in disinformation, and intention to spread disinformation on social media* (Publication No. 13880914) [Doctoral dissertation], Syracuse University. ProQuest Dissertations and Theses Global.

Leskin, P. (2020, March 31). *One of the internet's oldest fact-checking organizations is overwhelmed by coronavirus misinformation — And it could have deadly consequences*. Business insider. https://www.businessinsider.com/coronavirus-snopes-misinformation-fact-checking-overwhelmed-deadly-consequences-2020-3

*Li, H., & Sakamoto, Y. (2014). Social impacts in social media: An examination of perceived truthfulness and sharing of information. *Computers in Human Behavior*, *41*, 278–287. https://doi.org/10.1016/j.chb.2014.08.009

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), e1–e34. https://doi.org/10.1136/bmj.b2700

Murad, M. H., Chu, H., Lin, L., & Wang, Z. (2018). The effect of publication bias magnitude and direction on the certainty in evidence. *BMJ*, *23*(3), 84–86. https://doi.org/10.1136/bmjebm-2018-110891

National Heart, Lung, and Blood Institute (NHLBI). (2019). *Study quality assessment tools*. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

O'Keefe, D. J. (2015). Message generalizations that support evidence-based persuasive message design: Specifying the evidentiary requirements. *Health Communication*, *30*(2), 106–113. https://doi.org/10.1080/10410236.2014.974123

Ortiz, S. E., & Rosenthal, M. B. (2019). Medical marketing, trust, and the patient-physician relationship. *JAMA*, *321*(1), 40–41. https://doi.org/10.1001/jama.2018.19320

*Ozturk, P., Li, H., & Sakamoto, Y. (2015). Combating rumor spread on social media: The effectiveness of refutation and warning. In *Proceedings of the annual Hawaii international conference on system science* (pp. 2406–2414). Hawaii. https://doi.org/10.1109/HICSS.2015.288

*Pal, A., Chua, A. Y. K., & Hoe-Lian Goh, D. (2019). Debunking rumors on social media: The use of denials. *Computers in Human Behavior*, *96*, 110–122. https://doi.org/10.1016/j.chb.2019.02.022

Patel, R., Chang, T., Greysen, S. R., & Chopra, V. (2015). Social media use in chronic disease: A systematic review and novel taxonomy. *American Journal of Medicine*, *128*(12), 1335–1350. https://doi.org/10.1016/j.amjmed.2015.06.015

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116

Pew Research Center. (2014, January 15). *The social life of health information*. https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/

Smailhodzic, E., Hooijsma, W., Boonstra, A., & Langley, D. J. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Services Research*, *16*(1), 1–14. https://doi.org/10.1186/s12913-016-1691-0

*Smith, C. N., & Seitz, H. H. (2019). Correcting misinformation about neuroscience via social media. *Science Communication*, *41*(6), 790–819. https://doi.org/10.1177/1075547019890073

Southwell, B. G., Niederdeppe, J., Cappella, J. N., Gaysynsky, A., Kelley, D. E., Oh, A., Peterson, E. B., & Chou, W.-S. (2019). Misinformation as a misunderstood challenge to public health. *American Journal of Preventive Medicine*, *57*(2), 282–285. https://doi.org/10.1016/j.amepre.2019.03.009

*Sullivan, M. C. (2019). Leveraging library trust to combat misinformation on social media. *Library & Information Science Research*, *41*(1), 2–10. https://doi.org/10.1016/j.lisr.2019.02.004

Trew, B. (2020, March 27). *Coronavirus: Hundreds dead in Iran from drinking methanol amid fake reports it cures disease*. Independent.

*Tully, M., Vraga, E. K., & Bode, L. (2020). Designing and testing news literacy messages for social media. *Mass Communication & Society*, *23*(1), 22–46. https://doi.org/10.1080/15205436.2019.1604970

*van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, *35*(5), 560–575. https://doi.org/10.1080/10410236.2019.1573295

*Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621–645. https://doi.org/10.1177/1075547017731776

*Vraga, E. K., & Bode, L. (2018). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication, & Society*, *21*(10), 1337–1353. https://doi.org/10.1080/1369118X.2017.1313883

*Vraga, E. K., Kim, S. C., & Cook, J. (2019). Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *Journal of Broadcast & Electronic Media*, *63*(3), 393–414. https://doi.org/10.1080/08838151.2019.1653102

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, *47*(2), 155–177. https://doi.org/10.1177/0093650219854600

Wang, Y., Mckee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*, e112552. Article. https://doi.org/10.1016/j.socscimed.2019.112552

Wang, Z., Walther, J. B., Pingree, S., & Hawkins, R. P. (2008). Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health Communication*, *23*(4), 358–368. https://doi.org/10.1080/10410230802229738

*Weber, T. J., Muehling, D. D., & Kareklas, I. (2019). How unsponsored, online user-generated content impacts consumer attitudes and intentions toward vaccinations. *Journal of Marketing Communications*, 1–26. Advance online publication. https://doi.org/10.1080/13527266.2019.1671479

World Health Organization. (2020). *Definition of health*. https://www.who.int/about/who-we-are/constitution