McMaster University | HEALTH FORUM

Rapid Evidence Profile

Examining types of misinformation and disinformation and the effectiveness of interventions to combat misinformation and disinformation

17 November 2023

[MHF product code: REP 56]

Context

- Misinformation is false information that is spread, regardless of intent to mislead. Disinformation is the intentional spreading of misinformation.
- Mis/disinformation can have both direct, immediate effects on behaviours and indirect, long-term effects on trust in government, media and science.
- The Global Evidence Commission, drawing on the Broadband Commission for Sustainable Development, proposed a framework for considering the types of mis/disinformation practices and strategies to respond to them.(1)
- Existing investments in Canada's infrastructure to respond to mis/disinformation (e.g., Canadian Digital Media Research Network) have not prioritized evidence synthesis to date.

Questions

- What is known about the types of mis/disinformation practices in democracies, including those which have emerged more recently?
- What is known about the effectiveness of interventions to combat mis/disinformation?

Key findings

- The living evidence synthesis (LES) focused on addressing health-related mis/disinformation included 60 studies, with a predominance of: 1) randomized controlled trials (41 of 60); 2) studies conducted in the U.S. (28 of 51 studies with a geographical focus); and 3) studies examining educational (26 studies) and monitoring and fact-checking (24) interventions.
- The 17 evidence syntheses: 1) addressed health topics exclusively (4), politics exclusively (2) or a combination of topics (11); 2) were low (9) or medium (7) quality; and 3) conducted in 2020 or earlier (6) or in the last three years (11).
- We present key insights in text form below and more detailed findings in Tables 1 and 2 (with findings about 'political institutions' in the second column).

Key insights from evidence syntheses about question 1 - Types of mis/disinformation practices in democracies

- We identified nine evidence syntheses addressing mis/disinformation practices, of which four address cross-cutting topics, three address topics related to political institutions, and two address health topics (none provided insights about marketing or science more generally). None of the evidence syntheses are high quality.
- Key instigators or motivation for disinformation or misinformation include: messages on social media platforms (e.g., from amplification from high-profile users); fake news creation from many sources that generate interest due to eroding trust in mainstream media, a high-choice media environment, and ease of spread through social media; and diverse combinations of agents that collectively play overlapping and sometimes competing roles for producing and amplifying disinformation through these media environments.(2-4)
- Agents or techniques for disinformation or misinformation include fake social media profiles and social bots that produce content and interact with humans on social media, as well as selective censorship, manipulation of search rankings, hacking and releasing sensitive or damaging information and directly sharing disinformation on social media platforms.(3; 5)

- Messages can:
  - include fabricated/fake/false information that is made appealing through a variety of mechanisms (e.g., detailed articles that make it seem more trustworthy, clickbait, imposters, misleading connections, conspiracy theories/theorists, fake reviews, hoaxes, trolling, biased or one-sided, pseudoscience, and rumours), which often require expert judgement to determine accuracy
  - be made more trustworthy or less likely to be questioned through the use of group cues and stereotypes to facilitate acceptance of inaccurate information, anger to make people less likely to distrust inaccurate information, frequent dissemination to ensure continued and recent messages
  - achieve mass spread through connection to central broadcasters or mass media
  - end up being shared more when they are highly controversial
  - be made more persuasive and easy to share in audio-visual format.
- Social media, including Twitter (now X), YouTube and Instagram, are commonly used platforms for spreading misinformation given their low barriers to entry, support of creating easy-to-share media, and rapid dissemination of information, contributing to 'echo chambers' that confirm existing beliefs.(5; 6)
- Conspiracy thinking, religiosity, conservative ideology, conservative party identification, and frequent use of social media were all associated with susceptibility to health-related misinformation, whereas subject knowledge, literacy, numeracy, higher educational attainment and relying on health professionals and scientists for information sources predicted less susceptibility.(7)

Box 1: Approach and supporting materials

We drew from a recently completed draft of the first version of a living evidence synthesis (LES) focused on addressing health-related misinformation (protocol available here), as well as 17 evidence syntheses that were identified while conducting the LES and from supplementary searches addressing non-health topics.

Our searches for the living evidence synthesis (LES) focused on addressing health-related misinformation the LES were conducted on 4 May 2023 using eight databases (Medline, Embase, CINAHL, PsycINFO, COVID-END inventory of best evidence syntheses, Epistemonikos, ProQuest, and MedRxiv) and sources for grey literature (Google Scholar, Open Science Framework and greynet.org). Detailed search strategies are provided in the protocol. Our supplementary searches for relevant evidence syntheses on non-health topics were conducted on 16 November 2023 in PubMed and ProQuest using the terms misinformation OR disinformation AND systematic review. We appraised the methodological quality of evidence syntheses that were deemed to be highly relevant using AMSTAR. AMSTAR rates quality on a scale of 0 to 11, where 11/11 represents a review of the highest quality (and with ratings of 8 or above considered to be high quality).

This rapid evidence profile was prepared in the equivalent of one day of a 'full-court press' by all involved staff. Three caveats to keep in mind are: 1) we may have missed evidence syntheses addressing non-health topics; 2) we had to draw conclusions from evidence syntheses addressing non-health topics, not from the single studies contained in them as we were able to do for health topics, and most were low quality; 3) we did not have the time to: a) conduct risk-of-bias assessments of studies; b) match practices to those attracting political attention in Canada (e.g., targeting elected officials), c) extract information about attribution, or d) complete a jurisdictional scan or engage subject-matter experts.

Key insights from evidence syntheses about question 2 - Effectiveness of interventions to combat mis/disinformation

- We identified one high-quality living evidence synthesis that evaluates the effectiveness of interventions to address health-related mis/disinformation. We also identified six evidence syntheses addressing cross-cutting topics, one addressing topics related to political institutions, and one addressing health topics. None of these eight evidence syntheses are of high quality.
- Overall, the high-quality living evidence synthesis and eight one-off evidence syntheses found that:
  - monitoring and fact-checking have a weak to moderate positive effect in mitigating the influence of misinformation (8)
  - the effect is stronger when it is conducted and provided by experts (9)

- mis/disinformation is more difficult to correct in the context of politics and marketing than for health-related topics, and when shared by peers than from news organizations.(8; 9)
- Higher levels of education are typically a positive predictor for acceptance of health and scientific authority, but correction attempts are less effective in relation to politics, particularly among more educated political partisans.
- Credibility labelling has been found to have the potential to affect user behaviour but may not affect content belief without also providing a truthfulness rating.(10)
- Educational strategies reduce misinformation credibility assessment and reduce the intention of sharing misinformation.(11)
- Counter misinformation campaigns, normative strategies, and technical and algorithmic strategies were found to be effective for stimulating intentions to take protective actions, improving knowledge about a health topic, and reducing beliefs in health-related misinformation.(12)
- Economic strategies and legislative and other policy responses were found to have no effect in changing beliefs or deterring the sharing of health misinformation.(12)

Framework used to organize what we looked for

- Areas of disinformation or misinformation
  - Cross-cutting
  - Political
  - Health
  - Marketing, science or other
- Features of mis/disinformation
  - Instigators or motivation for disinformation or misinformation
  - Agents or techniques for disinformation or misinformation (e.g., bots and fake accounts, false identities)
  - Types of messages (e.g., emotive claims, fabricated or fraudulently altered images and videos, fabricated websites and polluted data sets)
  - Platforms (e.g., dark web, social media) and platform features (e.g., algorithms and business models) for spreading disinformation or misinformation
  - Targets of disinformation or misinformation and how they react
- Interventions to combat mis/disinformation
  - Monitoring and fact-checking responses
  - Counter-misinformation campaigns
  - Credibility labelling
  - Normative responses
  - Educational responses
  - Economic responses
  - Curatorial responses
  - Legislative and other policy responses
  - Technical and algorithmic responses
  - Investigative responses

Velez M, Waddell K, Demaio P, Bhuiya A, Lavis JN, Wilson MG. Rapid evidence profile #56: Examining types of misinformation and disinformation and the effectiveness of interventions to combat misinformation and disinformation. Hamilton: McMaster Health Forum, 17 November 2023.

This rapid evidence profile was funded by the Canadian Institutes of Health Research (grant number PJT-185898). The McMaster Health Forum receives both financial and in-kind support from McMaster University. The views expressed in the rapid evidence profile are the views of the authors and should not be taken to represent the views of the Canadian Institutes of Health Research or McMaster University.

More details about what we found

Table 1. Findings from evidence syntheses related to mis/disinformation practices (with year of literature search and AMSTAR score in brackets after each hyperlinked declarative title)

Features of mis/disinformation	Cross-cutting	Political institutions	Health
Instigators or motivation for disinformation or misinformation	• Messages on social media platforms, (e.g., Twitter), gain amplification because the message or information is associated with certain users or influencers, and exchanging depends on the ratings or the influential users associated with the information. (2018; 4/9)(2) • Fake news creation could be attributed to factors like blurred boundaries between professional journalists and other types of creators, the use of AI technology (i.e., social bots, trolls, Deepfake technology), ideological or monetary purposes, and/or to induce fear or panic; whereas fake news consumption could be related to the fall of mainstream media trust, high-choice media environment, and the use of social media as main news platforms (2021; 4/9)(4)	• A diverse combination of agents including hired and independent trolls, bots, fake-news websites, conspiracy theorists, politicians, highly partisan media outlets, the mainstream media, and foreign governments are all playing overlapping—and sometimes competing—roles in producing and amplifying disinformation in the modern media ecosystem. (2018; 0/9)(3)	
Agents or techniques for disinformation or misinformation (e.g., bots and fake accounts, false identities)	• Fake news is often disseminated through fake social media profiles, as well as social bots that use computer algorithms to produce content and interact with humans on social media (2021; 4/9)(5)	• Misinformation can be spread online by selective censorship (which involves removing some content from online platforms, while leaving other forms of content alone); the manipulation of search rankings (to make certain news stories or sources of misinformation more likely to appear, for example, in a Google search); hacking and releasing (sensitive and/or damaging information, primarily from email	

		accounts, and subsequently selectively leaking the information); and directly sharing disinformation on social media platforms (2018; 0/9)(3)	
Types of messages (e.g., emotive claims, fabricated or fraudulently altered images and videos, fabricated websites and polluted data sets)	• Types of disinformation include fabricated, clickbait, imposter, misleading connection, conspiracy theories, fake reviews, hoaxes, trolling, biased or one-sided, pseudoscience, and rumours (2020; 2/9)(13) • People creating fake news websites and writing false information exploit the non-intellectual characteristics of some people (2020; 4/9)(2) • For social media users to determine if the information they received is true or false, expert judgement of content is needed (2020; 4/9)(2) • Some individuals determine the trustworthiness of information provided to them through social media on how much detail and content it contains (2020; 4/9)(2)	• Messages priming group cues and stereotypes can facilitate the acceptance of inaccurate information about the out-group. (2018; 0/9)(3) • Anger makes people less likely to distrust inaccurate information that supports their views, and more likely to distribute it, while anxiety can have the opposite effect. (2018; 0/9) (3) • People are more likely to be affected by inaccurate information if they see higher volume and more recent messages reporting facts, irrespective of whether they are true (2018; 0/9) (3) • Information achieving mass spread usually relies on central broadcasters in a network and/or amplification by the mass media (2018; 0/9) (3) • Communities of belief, such as conspiracy theorists, are important in generating the kind of sustained attention that is needed for false information to travel (2018; 0/9) (3) • Content that is highly controversial is more likely to be shared by social media users (2018; 0/9) (3) • There is reason to believe that audio-visual messages can be both more persuasive and more easily spread than textual messages (2018; 0/9) (3)	
Platforms (e.g., dark web, social media) and platform features	• Social media's low entry barriers, 'thin slices' content formats, and users'	• One study in the U.S. included in a review reported that with regard only to	• Twitter, Facebook, YouTube and Instagram are critical in

(e.g., algorithms and business models) for spreading disinformation or misinformation	polarization work to increase the spread of misinformation by creating easy-to-share forms of media that can be shared in 'echo chambers' that confirm prior beliefs (2021; 4/9)(5)	Twitter, during the month prior to the elections, each user was exposed to fake news related to the political campaign 10 times on average, and 1% of the study sample consumed 80% of the detected fake news; those big consumers of fake news were mainly conservative in political orientation and were characterized by high consumption of all kinds of news (2021; 1/9)(14)	disseminating rapid and far-reaching misinformation, producing misleading or incorrect interpretations of available evidence, impacting mental health, misallocation of health resources and an increase in vaccination hesitancy (2022; 9/10)(6)
Targets of disinformation or misinformation and how they react	• Fake news can hurt firms' reputations by changing consumers' minds about a firm or product, and firms giving legitimacy to fake news and being contaminated by association (2021; 4/9)(5)		• Subject knowledge, literacy and numeracy, analytical thinking (vs. intuitive thinking), and trust in science confer strong resistance to health misinformation, whereas conspiracy thinking, religiosity, conservative ideology, and conservative party identification are associated with more susceptibility to health misinformation (2022; 7/11)(7) • Demographically, older age and higher educational attainment predict less susceptibility to health misinformation, whereas racial minority status is associated with greater susceptibility (2022; 7/11)(7) • Behaviourally, relying on health professionals or scientists as information sources predicts less susceptibility to health misinformation, whereas social media use is associated with greater susceptibility. (2022; 7/11)(7)

Table 2: Findings from evidence syntheses related to the effectiveness of interventions to combat mis/disinformation (with year of literature search and AMSTAR score in brackets after each hyperlinked declarative title)

Interventions to combat mis/disinformation	General	Political institutions	Health*
Monitoring and fact-checking	• Debunking (or fact-checking) misinformation can be effective when it facilitates an understanding of detailed counter-arguments, while simply labelling misinformation or elaborating on reasons for a particular event can support the persistence of the misinformation (rather than the success of a counter-message) (2015; 5/11)(15) • Corrective attempts have a weak to moderate effect on changing misinformation-related beliefs, being more difficult to correct in the context of politics and marketing than health, and more in real-world than in experiments (2018; 3/11)(8) • While higher levels of education are usually a positive predictor for acceptance of health and scientific authority, when it comes to politics, correction attempts seem to be less effective, particularly among more educated political partisans (2018; 3/11)(8) • Once people are exposed to a coherent message that can explain the chain of events, they will be more likely to substitute the false information with the retraction (2018; 3/11)(8) • Fact-checking was not effective for all individuals (e.g., those with partisan	• Fact-checking messages positively affect beliefs, irrespective of political ideology, preexisting positions, context (campaign vs. routine), and whether it refutes the entire false statement or just parts of a statement; however, the effect is weak, and not all fact-checking attempts are equally effective (2018; 3/11)(8)	• Corrections have a weak to moderate positive effect in mitigating the influence of misinformation; additionally, misinformation shared by peers is more challenging to debunk than misinformation from news organizations, and corrections from experts are more effective than corrections from non-experts (2021; 6/11)(9)

	viewpoints who were more vulnerable to misinformation); fact corrections from people with existing relationships and inoculation effects (e.g., pre-exposure to counterarguments) could be additional strategies to combat misinformation (2022; 2/9)(16)		
Counter-misinformation campaigns			• Misinformation campaigns were found to be effective in five studies for stimulating intentions to take protective actions, improving knowledge about a health topic, and reducing beliefs in misinformation (2023; 9/10)(12)
Credibility labelling	• Veracity labelling, which indicates the truth status (e.g., misleading, truthful, or not-verified) of information and are based on fact-checks, have been found to be effective in correcting misconceptions, especially when details are provided about refutation (2022; 0/9)(10) • Context labelling providing information about the source and/or metadata such as sharing trends along with reminders for critical thinking have the potential to affect user behaviour, but may not have an effect on content belief without an accompanying truthfulness rating (2022; 0/9)(10)		• Fact-checking labels attached to misinformation posts were found in one study to make vaccine attitudes more positive than the misinformation control condition, especially when the labelling was performed by universities and health institutions (2023; 9/10) (12) • Machine learning–based approaches to providing credibility labelling were found to be successful in two studies for classifying reliable information compared to classifying unreliable information (2023; 9/10) (12)
Normative responses	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified	• A mediating or suppressing effect of follower count (in social media) was identified by one study in the relationship between a debunker's identity (celebrity, media, or government) and sharing a behaviour (2023; 9/10)(12) • Ads featuring peer modelling with psychological inoculation yielded a

			significantly higher rate of positive responses than the CDC ads (2023; 9/10) (12)
Educational responses	• Psychological inoculation involves pre-emptively exposing users to weakened forms of misinformation, and has been found to reduce misinformation credibility assessment and improve real information credibility assessment and real information sharing intention. (2023; 6/11)(11)	• No findings from evidence syntheses identified	• Eight US-based studies found educational responses are effective for changing the beliefs of people exposed to misinformation, but not effective for stimulating intentions to take protective actions when compared to not providing education (2023; 9/10)(12) • From international studies, educational responses were effective in stimulating intentions to take protective actions when compared to not providing education, in changing the beliefs of people exposed to misinformation, in improving knowledge about a topic, in changing the willingness to share misinformation, and in enhancing the ability to discriminate misinformation (2023; 9/10)(12)
Economic responses	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified	• Two studies indicated that financial incentives might not have a beneficial effect in reducing the willingness to share misinformation (2023; 9/10)
Curatorial responses	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified
Legislative and other policy responses	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified	• Legislation was found in one study to possibly deter the sharing of healthcare information that users perceive as true, but was noted as not deterring them from sharing the healthcare misinformation they perceive as fake (2023; 9/10)(12) • A Facebook policy to restrict anti-vaccine posting was found by one study to have a small effect in reducing the number of posts, which remained steady after the policy (2023; 9/10)(12)

Technical and algorithmic responses	• Knowledge-based automatic fake news detections have been identified as being needed to combat future misinformation; current fake news detection models involved gathering information from social media, supervised machine learning or deep learning algorithms, use of NLP techniques for data extraction, techniques, and use of neural network-based training models to distinguish fake versus real news (2022; 0/9)(17)	• No findings from evidence syntheses identified	• Interacting with a chatbot for a few minutes was found by two studies to significant increase people's intentions to get vaccinated and positively impacted their attitudes toward COVID-19 vaccination (2023; 9/10)(12)
Investigative responses	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified	• No findings from evidence syntheses identified

*Findings related to health were extracted from a high-quality living evidence synthesis on the effectiveness of strategies to mitigate misinformation

References

1. Global Commission on Evidence to Address Societal Challenges. 4.11 Misinformation and infodemics. 2022. In: The Evidence Commission report: A wake-up call and path forward for decision-makers, evidence intermediaries, and impact-orineted evidence producers [Internet]. Hamilton: McMaster Health Forum.
2. Celliers M, Hattingh M, editors. A systematic review on fake news themes reported in literature Responsible Design, Implementation and Use of Information and Communication Technology. 2020; Cham: Springer International Publishing.
3. Tucker J, Guess A, Barbera P, et al. Social media, political polarization, and political disinformation: A review of the scientific literature. 2018.
4. Kim B, Xiong A, Lee D, Han K. A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLOS ONE* 2021;16(12): e0260080.
5. Domenico GD, Sit J, Ishizaka A, Nunan D. Fake news, social media and marketing: A systematic review. *Journal of Business Research* 2021;124: 329-341.
6. Borges do Nascimento IJ, Pizarro AB, Almeida JM, Azzopardi-Muscat N, Gonçalves MA, Björklund M, Novillo-Ortiz D. Infodemics and health misinformation: A systematic review of reviews. *Bullet of the World Health Organization* 2022;100(9): 544-561.
7. Nan X, Wang Y, Thier K. Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. *Social Science Medicine* 2022;314: 115398.
8. Walter N, Murphy ST. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs* 2018;85(3): 423-441.
9. Walter N, Brooks JJ, Saucier CJ, Suresh S. Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communications* 2021;36(13): 1776-1784.
10. Morrow G, Swire-Thompson B, Polny JM, Kopec M, Wihbey JP. The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the association for information science and technology* 2022;73(10): 1365-1386.
11. Lu C, Hu B, Li Q, Bi C, Ju XD. Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-Analysis. *Journal of Medical and Internet Research* 2023;25: e49255.
12. Vélez C, MG W, Lavis J. The impact of strategies to mitigate misinformation in diverse settigns and populations: Living evidence synthesis. 2023. [Not yet published]
13. Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society* 2021;23(5): 1301-1326.
14. Miró-Llinares F, Aguerri JC. Misinformation about fake news: A systematic critical review of empirical studies on the phenomenon and its status as a 'threat'. *European Journal of Criminology* 2023;20(1): 356-374.
15. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychol Sci* 2017;28(11): 1531-1546.
16. Arcos R, Gertrudix M, Arribas C, Cardarilli M. Responses to digital disinformation as part of hybrid threats: a systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking. *Open Research Europe* 2022;2: 8.
17. Nirav Shah M, Ganatra A. A systematic literature review and existing challenges toward fake news detection models. *Social Network Analysis and Mining* 2022;12(1): 168.