McMaster University | HEALTH FORUM

Living Evidence Synthesis

Impact of strategies to mitigate health-related misinformation in diverse settings and populations

15 April 2024

[MHF product code: LES 22.1]

Protocol registered with PROSPERO (CRD42023421149) and published in BMJ Open

Context

- **Impact of strategies to mitigate health-related misinformation in diverse settings and populations**
- Misinformation, which refers to "information that is false, inaccurate, or misleading according to the best available evidence at the time," (1) can delay or prevent effective care,(2) affect mental health,(3) lead to misallocation of health resources (3) and/or create or exacerbate public-health crises.(3; 6)
- In addition, misinformation can affect some members of society more than others (e.g., those with lower digital, numerical and health literacy and/or cognitive skills are more vulnerable to misinformation),(7; 8) and therefore these groups may be more exposed to health threats, leading to greater social and health inequities.(9)
- Groups most vulnerable to health-related misinformation include younger people, those with lower educational attainment, racial minorities, and social media users,(10) and a disproportionate impact on women, trans, and nonbinary people has also been highlighted.(11)
- Extensive COVID-19-related misinformation and a constantly evolving social media landscape have spurred efforts to mitigate the spread of falsehoods that undermine public trust in evidence-based care.
- While there are many strategies that have been identified to address misinformation, there is a need to evaluate their effects.

Question

What is the impact of strategies to mitigate misinformation in diverse settings, and across diverse populations?

High-level summary of key findings

Overview of evidence identified

- We identified 2,086 unduplicated articles and included 60 studies (mostly published in the last four years), which included 41 randomized controlled trials, six quasi-experimental studies, six that used machine learning approaches, three implementation research studies, two cross-sectional studies and one other type of observational study.
- The included studies were conducted online not limited to any geographic region (n=9); in more than one country (n=3) (one in Kyrgyzstan, India, and the U.S.; a second in Germany, Mexico, Spain, the U.K. and the U.S.; and the third in Australia, Canada and the U.S.); and in Australia (n=2), Brazil (n=2), Canada (n=1), China (n=3), France (n=1), Ghana (n=1), Guatemala (n=1), Hong Kong (n=3), Israel (n=1), Italy (n=2), Korea (n=1), Nigeria (n=1), Sierra Leone (n=1), the Netherlands (n=2), U.S. (n=25), U.K. (n=1) and Zimbabwe (n=1).
- We identified evidence for eight of the 10 types of responses/strategies to counter misinformation, but no evidence was identified for curatorial and investigative responses/strategies.
  - The interventions addressed by studies included monitoring and fact-checking (n=24), counter-misinformation campaigns (n=5), credibility labelling (n= 6), educational (n=26), narrative (n=2), technical and algorithmic (n=7), economic (n=2), and legislative and other policy (n=2).
- Topics addressed by studies include COVID-19 (n=31), vaccination (n=14), human papilloma virus (n=4), and other health topics (n=18).

Strategies identified as effective at reducing misinformation

- **Educational strategies (n=26 studies)**
  - Overall, the eight experimental randomized studies conducted in the U.S. found that educational strategies 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.
  - The twelve experimental randomized studies conducted in different countries (e.g., Australia, Brazil, Canada, China, France, Germany, Hong Kong, India, Italy, Mexico, Sierra Leone, Spain, and the U.K.) found educational strategies 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.
  - Among the four quasi-experimental studies, two found educational strategies effective for stimulating intentions to take protective actions when compared to not providing education, and for improving knowledge about a topic; the other two studies found educational interventions ineffective in increasing belief accuracy.
- **Monitoring and fact-checking (n=24 studies)**
  - Overall, the identified studies found that monitoring and fact-checking strategies are effective for stimulating intentions to take protective actions when compared to not providing corrections, and that those strategies might change the beliefs of people exposed to misinformation, as well as their willingness to share misinformation.
  - It was also reported that the positive effect of monitoring and fact-checking is stronger if the source of the fact-checking is provided (one study), that the format of the correction does not make a considerable difference in effectiveness (one study), and that humorous corrections might produce more attention to the misinformation text than non-humorous corrections, but not improve credibility of the correction (two studies).
  - Two studies explored the familiarity backfire effect and reported that corrections that exposed participants to novel misinformation did not lead to stronger misconceptions compared to people never exposed to false claims or corrections, suggesting that it is safe to repeat misinformation when correcting it, even when the audience might be unfamiliar with the misinformation.
- **Technical and algorithmic strategies (n=7 studies)**
  - The six machine learning studies reported the effectiveness of different models in identifying misinformation.
  - Two studies, one experimental randomized and one machine learning study, found that interacting with a chatbot for a few minutes significantly increased people's intentions to get vaccinated and positively impacted their attitudes toward COVID-19 vaccination.
- **Credibility labelling (n=6 studies)**
  - Overall, three experimental randomized studies reported the effectiveness of credibility labelling on the ability to critically evaluate a given message, or in accurately identifying misinformation.
  - Only one experimental randomized study in the U.S. reported that fact-checking labels attached to misinformation posts made vaccine attitudes more positive than the misinformation control condition, especially when the labelling was performed by universities and health institutions.
  - Two studies of machine learning–based approaches found those strategies successful in classifying reliable information compared to classifying unreliable information.
- **Counter-misinformation campaigns (n=5 studies)**
  - Overall, misinformation campaigns were effective for stimulating intentions to take protective actions, improving knowledge about a health topic, and reducing beliefs in misinformation.

Strategies found to have limited or no evidence for reducing misinformation

- **Narrative strategies (n=2 studies)**
  - One experimental randomized study conducted in China found a mediating or suppressing effect of follower count (in social media) in the relationship between a debunker's identity (celebrity, media, or government) and sharing behaviour.
  - Another experimental randomized study that was conducted in the U.S. used advertisements for Facebook providing video testimonials from peer role models promoting vaccination, and found that ads featuring peer modelling with psychological inoculation yielded a significantly higher rate of positive responses than the Center for Disease Control ads.
- **Legislative and other policy strategies (n=2 studies)**
  - The experimental randomized study in Hong Kong found that legislation may deter the sharing of healthcare information that users perceive as true but cannot deter them from sharing the healthcare misinformation they perceive as fake.
  - An interrupted time series study found that a Facebook policy to restrict anti-vaccine posting had a small effect in reducing the number of posts, which remained steady after the policy.
- **Economic strategies (n=2 studies)**
  - Two experimental randomized studies conducted in Hong Kong and the U.K. found that financial incentives might not have a beneficial effect in reducing the willingness to share misinformation.
- **Curatorial strategies (n=0 studies)**
  - No evidence identified.
- **Key findings in relation to investigative strategies (n=0 studies)**
  - No evidence identified.

Background

Increasing digitalization and use of social media is a double-edged sword.(12) It creates opportunities to rapidly communicate and disseminate information to address social challenges, and is therefore an important tool for reaching individuals and communities.(2; 12) However, as emphasized by the United Nations (UN),(13) digital technologies and social media also have the potential of introducing misinformation to citizens.(12) Misinformation, which refers to "information that is false, inaccurate, or misleading according to the best available evidence at the time," (1) can delay or prevent effective care,(2) affect mental health,(3) lead to misallocation of health resources (3) and/or create or exacerbate public-health crises.(3; 6) Disinformation or malinformation are other common terms, but refer to instances "…when misinformation is used to serve a malicious purpose, such as to trick people into believing something for financial gain or political advantage." (1)

Misinformation can affect some members of society more than others (e.g., those with lower digital, numerical and health literacy and/or cognitive skills are more vulnerable to misinformation),(7; 8) and therefore these groups may be more exposed to health threats, leading to greater social and health inequities.(9) A systematic review conducted in 2021 found the groups most vulnerable to health-related misinformation include younger people, those with lower educational attainment, racial minorities, and social media users.(10) The disproportionate impact on women, trans, and nonbinary people has also been highlighted. For instance, a report published by Plan International Australia shows that those groups are bombarded with stereotypes and misleading facts about their bodies

Box 1: Approach and supporting materials

We retrieved candidate studies by searching seven electronic databases: 1) Medline, 2) Embase, 3) CINAHL, 4) PsycINFO, 5) COVID-END inventory of best evidence syntheses, 6) Epistemonikos, and 7) pre-print servers (MedRxiv); as well as sources for grey literature (Google Scholar, Open Science Framework and greynet.org). Search terms were developed with the collaboration of a library scientist using medical subject headings (MeSH) and text words related to forms of misinformation and interventions. Searches were focused on studies conducted with humans and published since database inception until 4 May 2023. Searches will be updated at six and nine months after the original search. Our detailed search strategy is included in **Appendix 1**.

We included original articles without language restrictions that evaluate one or more of the potential responses to health-related misinformation listed in Table 3. We included experimental, quasi-experimental, and observational studies for any populations, settings, and diseases (i.e., we will not limit to only COVID misinformation). Outcomes considered included change in attitudes/behaviour, health benefits, harms, and costs. We excluded evidence syntheses but reviewed their references to identify additional studies to include. A full list of included studies is provided in **Appendix 2**. Studies excluded at the last stages of reviewing are provided in **Appendix 3**.

Population of interest: general population (stratified by age, gender and sex, users of different social media and platforms).

Intervention and control/comparator: different strategies including, monitoring and fact-checking, counter-misinformation campaigns, credibility labelling, educational, curatorial, narrative, technical and algorithmic, economic, legislative and other policy, and investigative.

Outcomes: Change in attitudes/behaviour, health benefits, harms, costs.

Data extraction: Data extraction was conducted by one team member and checked by another.

Critical appraisal: We have not yet conducted critical appraisal of included studies for risk of bias and certainty of evidence, but this is prioritized for the next version of the LES. For risk of bias, we will use the Cochrane risk of bias tool (RoB 2) for any experimental studies. For observational study designs, we will use a version of ROBINS-I that was enhanced for assessment of cohort studies in a series of living evidence syntheses evaluating COVID-19 public health and social measures.(4; 5) We will also use the GRADE approach for assessing the certainty of evidence for the outcomes identified. One reviewer independently conducted the assessments, then was checked for accuracy by another reviewer. Any discrepancies between reviewers were solved through consensus.

Summaries: We summarized the evidence by presenting narrative evidence profiles across studies by outcome measure. When appropriate, statistical pooling of results was performed.

The next update to this document will be provided in Summer 2024.

and their health.(11) This was found to lead to feeling unsafe because of online health information and questioning whether to get COVID-19 vaccines.

Extensive COVID-19-related misinformation and a constantly evolving social media landscape have spurred efforts to mitigate the spread of falsehoods that undermine public trust in evidence-based care. has spurred efforts to mitigate the spread of falsehoods that undermine public trust in evidence-based care. Such efforts were a focus in a report from the Broadband Commission for Sustainable Development (14) and in the Global Commission on Evidence.(13) Individuals can engage with misinformation through different sources.(15) In particular, while social-media platforms are a key driver of misinformation,(3; 16) it is not well understood since data is not publicly available for analysis, and because many popular platforms (e.g., Instagram, YouTube, TikTok, Facebook, and Pinterest), use visual content instead of text.(15) For example, a study that analyzed 800 vaccine-related Pinterest posts found that 74% were anti-vaccine in sentiment.(17; 18)

Although vaccines were the most common topic of misinformation before COVID-19,(19) other common topics for misinformation include reproductive health, substance use or smoking, non-communicable diseases, pandemics, eating disorders, and medical treatments.(19) Governments have employed a variety of strategies designed to debunk misinformation, including monitoring and fact-checking, economic incentives, and legislative policies.(20-23) These strategies must be assessed and compared in terms of impacts and effects on health outcomes and behaviour change. For instance, one older evidence synthesis found that correcting misinformation has a moderate influence on belief in misinformation, rebuttals are more effective than forewarnings, and appeals to coherence are more effective than fact-checking and appeals to credibility.(24) However, the rapid evolution of platforms for information sharing, and growth and innovation of misinformation actors means that previously synthesized evidence may no longer be valid to understand current misinformation challenges. In particular, the rapid expansion of artificial intelligence (AI) has many implications for misinformation, which will need to be better understood. This includes the potential for AI to amplify or propagate misinformation, but also for it to be used as a tool to address misinformation (e.g., through automated fact checking and credibility labelling.(25) Given this, there is a need for new high-quality and routinely updated evidence syntheses from trusted sources that assess the comparative impact of different strategies.

In 2020, the International Telecommunication Union (ITU) and United Nations Educational, Scientific and Cultural Organization (UNESCO) sponsored the Broadband Commission for Sustainable Development. This Commission developed a report about countering digital misinformation while respecting freedom of expression.(14) The report provides a framework for ten potentially effective responses to misinformation and the possible intersections with freedom-of-expression rights,(14) which is outlined in Table 1 along with a classification of five categories of governmental strategies to address COVID-19 misinformation identified in a non-systematic review conducted in 2021.(26)

Table 1: Potential responses to misinformation (table adapted from: "Broadband Commission research report on 'Freedom of Expression Addressing Disinformation on the Internet'" and "Governmental actions to address COVID-19 misinformation") (14; 26)

Response/ strategy	Description	Purpose of the strategy	Intersections with freedom-of-expression rights
Monitoring and fact-checking	Ongoing monitoring and timely exposing misinformation (e.g., debunked claims) and fact-checking new claims Judgement of trained professionals employed by independent organizations, even when helped by automation	Mitigating dissemination of disinformation, false information, and misinformation	Can mitigate the risk of infringing on freedom-of-expression rights
Counter-misinformation campaigns	Specialized units to develop counter-narratives to challenge misinformation and mobilizing online communities to spread high-quality evidence		

Credibility labelling	Content-verification tools, web-content indicators, signposting to credible evidence sources, and website-credibility labelling	Disseminating and increasing access to accurate information	
Educational	Develop citizens' media/information literacy for critical-thinking and digital-verification, and journalists' information literacy		
Curatorial	Point users to credible evidence sources, which can be used by news media, social media, messaging and search platforms		
Narrative	Public condemnations of misinformation and recommendations to address it, often by political and societal leaders	Restricting access to inaccurate information	
Technical and algorithmic	Ranges from human learning to machine learning and other artificial-intelligence approaches to identify misinformation, provide additional context, and limit spread		Automation of appeal processes can infringe on freedom-of-expression rights
Economic	Advertising bans, demonetizing specific content (e.g., for COVID-19) and approaches to remove misinformation incentives	Addressing commercial fraud	Can be misused as a form of private censorship
Legislative and other policy	Criminalize acts of misinformation, directing Internet communication companies to take down content, and providing material support for credible information sources	Criminalizing expressions of disinformation	Can be misused to weaken legitimate journalism and infringe on freedom-of-expression rights
Investigative	Examine instigators, degree and means of spread, money involved, and affected communities		Can inform legislative and other responses

Objective

To synthesize and continually update empirical evidence on strategies to address health-related misinformation in different settings and across diverse populations.

What we found

After removing duplicates, we screened 2,085 titles and abstracts; and selected 93 as eligible. After full-text review, 60 studies were included in this LES (see Figure 3 for the PRISMA chart, and Appendix 3 for the list of studies excluded). Most studies were published in 2020 and after, only four studies were published before 2020.(27-30) See Table 2 for details of all included studies.

Some of the studies were conducted online not limited at any geographic region (n=9), the rest of studies were conducted in Australia (n=2), Brazil (n=2), Canada (n=1), China (n=3), France (n=1), Ghana (n=1), Guatemala (n=1), Hong Kong (n=3), Israel (n=1), Italy (n=2), Korea (n=1), Nigeria (n=1), Sierra Leone (n=1), the Netherlands (n=2), U.S. (n=25), U.K. (n=1), Zimbabwe (n=1), three multi-country, one in Kyrgyzstan, India, and the U.S.; other in the U.S., Mexico, the U.K., Germany, and Spain; and the last one in U.S., Australia, and Canada.

Most studies (n=39) came from behavioural sciences. The specific study designs were experimental randomized studies (n=41), machine learning–based approaches (n=6),(31-36) quasi experimental (n=5),(28; 37-40) implementation research (n=3),(41-43) cross-sectional (n=2),(44; 45) qualitative research (n=2),(46; 47) observational study (n=1),(48) and interrupted time-series (n=1).(49)

The interventions addressed in the included studies (some studies addressed more than one strategy) were monitoring and fact-checking (n=24),(27; 29; 30; 41-43; 46; 48; 50-65) counter-misinformation campaigns (n=5),(37; 42; 44; 47; 66) credibility labelling (n= 6),(32; 34; 67-70) educational (n=26),(28; 35; 38-40; 45; 53; 55; 58; 60-62; 64; 65; 71-82) narrative (n=2),(83; 84) technical and algorithmic (n=7),(31-36; 71) economic (n=2),(20; 69) and legislative and other policy (n=2).(20; 49)

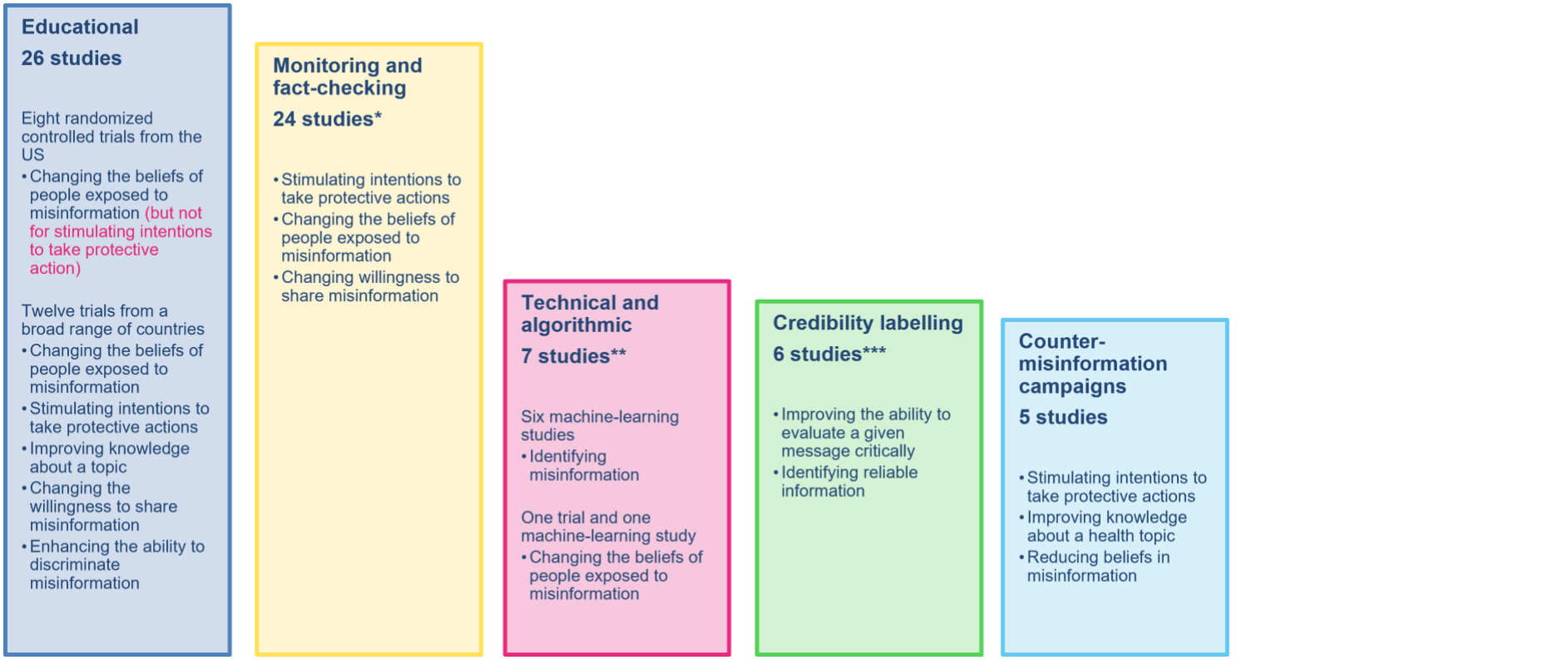
The topics addressed (some studies addressed more than one topic) were COVID-19 (n=31),(32; 33; 35-37; 39-42; 44-46; 48; 50; 53; 55-57; 59; 66; 67; 71-73; 76-78; 81; 83; 84) vaccination (excluding for COVID-19) (n=14),(30; 32; 34; 40-42; 49; 52; 53; 58; 60; 62; 70; 83) human papilloma virus (n=4),(30; 31; 54; 74) and other health topics (n=18).(20; 27-29; 38; 43; 47; 61; 63-65; 68; 69; 75; 79; 80; 82; 85)

Seven studies provided disaggregated analysis by gender. Studies were conducted in Brazil (n=2), Guatemala (n=1), Hong Kong (n=1), the U.S. (n=1), Zimbabwe (n=1), and a multi-country in Kyrgyzstan, India, and the U.S. (n=1).

All studies identified were mapped according to the 10 responses/strategies to counter misinformation presented in Table 1. We identified evidence for eight of the ten types, including evidence for monitoring and fact-checking, counter-misinformation campaigns, credibility labelling, educational, narrative, technical and algorithmic, economic, and legislative and other policy. There was no evidence identified for curatorial and investigative responses/strategies.

We provide a high-level overview of the key findings for each strategy in Figures 1 and 2 below, which is followed by a detailed description of the principal findings for each response/strategy (see Table 3 for details). The findings are presented by those that we found to be effective at reducing misinformation, followed by those for which we found limited or no evidence for reducing misinformation.

Figure 1: Strategies that have supporting evidence about their effectiveness in addressing misinformation



***Monitoring and fact-checking**

1) Studies reported that the positive effect of monitoring and fact-checking is stronger if the source of the fact-checking is provided (one study), that the format of the correction does not make a considerable difference in effectiveness (one study), and that humorous corrections might produce more attention to the misinformation text than non-humorous corrections, but not improve credibility of the correction (two studies)
2) Two studies explored the familiarity backfire effect and reported that corrections that exposed participants to novel misinformation did not lead to stronger misconceptions compared to people never exposed to false claims or corrections, suggesting that it is safe to repeat misinformation when correcting it, even when the audience might be unfamiliar with the misinformation

****Technical and algorithmic strategies**

1) Two studies, one trial and one machine learning study, found that interacting with a chatbot for a few minutes significantly increased people's intentions to get vaccinated and positively impacted their attitudes toward COVID-19 vaccination

***** Credibility labeling**

1) One trial in the U.S. reported that fact-checking labels attached to misinformation posts made vaccine attitudes more positive than the misinformation control condition, especially when the labelling was performed by universities and health institutions.

Figure 2: Strategies that have little or no available evidence

Narrative

2 studies

- Changing the willingness to share misinformation

Legislative and policy*

2 studies

- Changing the willingness to share misinformation

Economic

2 studies**

- Changing the willingness to share misinformation

Curational

No studies identified

Investigative

No studies identified

***Legislative and policy strategies**

1) One experimental randomized study conducted in China found a mediating or suppressing effect of follower count (in social media) in the relationship between a debunker's identity (celebrity, media, or government) and sharing behaviour.
2) Another experimental randomized study that was conducted in the U.S. found that ads featuring peer modelling with psychological inoculation yielded a significantly higher rate of positive responses than the Center for Disease Control ads.

****Economic strategies**

1) An experimental randomized study in Hong Kong found that legislation may deter the sharing of healthcare information that users perceive as true but cannot deter them from sharing the healthcare misinformation they perceive as fake.
2) An interrupted time series study found that a Facebook policy to restrict anti-vaccine posting had a small effect in reducing the number of posts, which remained steady after the policy.

Strategies identified as effective at reducing misinformation

Educational

We identified 26 studies,(28; 35; 38-40; 45; 53; 55; 58; 60-62; 64; 65; 71-82) which were conducted in Australia (n=2), Brazil (n=2), China (n=1), Hong Kong (n=2), France (n=1), Italy (n=1), Nigeria (n=1), Sierra Leone (n=1), the U.S. (n=11), multi-country (n=3), and online not linked to a geographic region (n=1). The health topics addressed were COVID-19 (n=13), vaccines (n=4), and other topics (n=9). The study designs were experimental randomized studies (n=20), quasi experimental (n=4), cross-sectional (n=1), and machine learning–based approaches (n=1). Overall, 23 studies enrolled 39,782 participants, 19,956 were women (50.2%).

Among the experimental randomized studies, eight were conducted exclusively in the U.S. Overall, those studies found that educational strategies are effective for changing the beliefs of people exposed to misinformation,(60; 62; 64; 80; 82) but not effective for stimulating intentions to take protective actions when compared to not providing education.(64; 77; 79) For instance, one study that tested the efficacy of a structured reading support intervention for evaluation and critique on cultivating a critical awareness of flawed scientific claims in an online setting, found no difference with people not receiving the intervention.(79) Additionally, one study reported that news literacy messages can alter misinformation perceptions and beliefs, but not with a single message.(80) Another study found that educational messages were more persuasive when delivered by a non-expert, and that an expert speaker increased the persuasiveness of videos only when the evidence provided was statistical.(75)

The other 12 experimental randomized studies were conducted in different countries (e.g., Australia, Brazil, Canada, China, France, Germany, Hong Kong, India, Italy, Mexico, Sierra Leone, Spain, and the U.K.). Those studies found educational strategies to be effective for stimulating intentions to take protective actions when compared to not providing education,(53; 61; 71; 74) changing the beliefs of people exposed to misinformation,(61; 65; 71; 78; 81) improving knowledge about a topic,(55; 74; 81) changing the willingness to share misinformation,(73; 78) and enhancing the ability to discriminate misinformation.(76) Additionally, one study in Australia found no evidence that repeating myths increased agreement with myths compared with other debunking strategies.(58) Another study conducted in Italy reported that among participants with higher levels of conspiracy mentality, those exposed to counterfactual pre-bunking rated fake news headlines less plausible than those in a control condition.(72)

Among the four quasi experimental studies, two found educational strategies to be effective for stimulating intentions to take protective actions when compared to not providing education,(40) and for improving knowledge about a topic.(28) The other two studies found educational interventions not to be effective for increasing belief accuracy.(38; 39)

Additionally, one cross-sectional study in Brazil examined the extent to which WhatsApp users might be willing to correct their peers who might share COVID-19 misinformation. The study found a pattern of how different demographics influenced the three types of social correction behaviours. Younger participants exhibited greater passivity in engaging with social correction, while higher educational attainment was associated with providing correction to the original sender, and male participants were more likely to send the correction to an entire group.(45) One machine learning study in the U.S. found that providing factual information on Twitter leads to a decrease in misinformation.(35)

Monitoring and fact-checking

We identified 24 studies,(27; 29; 30; 41-43; 46; 48; 50-65) which were conducted in Australia (n=2), China (n=1), Ghana (n=1), Hong Kong (n=1), Italy (n=1), Sierra Leone (n=1), the U.S. (n=13), a multi-country study in Australia, Canada and the U.S. (n=1), and studies performed online that were not linked to any geographical region (n=3). Studies addressed different health misinformation topics, with COVID-19 being the most common (n=8), followed by vaccines (n=6). The study designs were experimental randomized studies (n=19), implementation

research (n=3), and qualitative research (n=1). Overall, 21 studies enrolled 12,664 participants, 5,439 were women (42.9%).

Among the experimental randomized studies, eleven were conducted exclusively in the U.S. Overall, those studies found that monitoring and fact-checking strategies are effective for stimulating intentions to take protective actions when compared to not providing corrections,(29; 50; 52; 62; 86) and that those strategies might change the beliefs of people exposed to misinformation,(56; 59; 64) as well as their willingness to share misinformation.(57) Studies also reported that the positive effect of monitoring and fact-checking is stronger if the source of the fact-checking is provided,(29) that the format of the correction does not make a considerable difference in effectiveness,(60) and that humorous corrections might produce more attention to the misinformation text than non-humorous corrections, but not improve credibility of the correction.(30; 54) Additionally, two studies explored the familiarity backfire effect, and reported that corrections that exposed participants to novel misinformation did not lead to stronger misconceptions compared to people never exposed to false claims or corrections; suggesting that it is safe to repeat misinformation when correcting it, even when the audience might be unfamiliar with the misinformation.(51; 60)

The other eight experimental randomized studies were conducted in different countries (e.g., Australia, Canada, Hong Kong, Israel, Sierra Leone). Those studies also found monitoring and fact-checking strategies effective for stimulating intentions to take protective actions when compared to not providing corrections,(53; 61) improving knowledge about a health topic,(61) and reducing beliefs in misinformation.(27; 58; 65) Additionally, one study in Australia investigated the impact of misinformation on the willingness-to-pay for an unproven treatment and the propensity to share misinformation online. The study found that both tentative and enhanced refutations reduced demand for the treatment (18% and 25%, respectively) and misinformation promotion (29% and 55%).(55)

The three implementation research studies described programs for monitoring and fact-checking in the U.S.,(42) Ghana,(41) and Italy.(43) The effectiveness of those programs was not evaluated; however, they reported that the targeted population was reached in each country.

One qualitative study conducted in China found that since rumours in public-health crises often involve different objects, rumour refutation requires various information sources and, therefore, different rumour-debunking models apply. Such socialized rumour-debunking models can be categorized into government-led, media-led, scientific community-led, rumour-debunking platform-led and multi-agent collaborative models.(46) The government-led model features authenticity but has limited scope of dissemination. The media-led model utilizes its resources to quickly contact relevant departments and parties and verify the rumour before releasing rumour-debunking information, and the advantage of this is approach is the ability to produce instantaneous responses. Scientific community-led models debunk information through means such as knowledge exchange, joint publication, and mutual reviews, which are then followed by releasing rumour-debunking articles. This model is scientifically viable but limited in scope. The rumour-debunking platform-led model collects clarifications on local rumours released by departments and media platforms, and normally has a regionally limited scope. The multi-agent collaborative model promotes the transition of the rumour-debunking model from the traditional path of "rumour emerges–government and media dispel the rumour" to "rumour emerges–users report the rumour–the rumour is dispelled jointly".(46)

Technical and algorithmic

We identified seven studies,(31-36; 71) which were conducted in France (n=1), the Netherlands (n=1), U.S. (n=1), and online not linked to a geographic region (n=4). The health topics addressed were COVID-19 (n=5) and HPV (n=2). The study designs included machine learning–based approaches (n=6) and an experimental randomized study (n=1). Overall, four studies enrolled 8,040 participants, 4,737 were women (58.9%).

The six machine learning studies reported effectiveness of different models in identifying misinformation.(32-36) Specifically, one study developed a chatbot named DR-COVID with an ensemble Natural Language Processing

(NLP) model on the Telegram platform and evaluated various performance metrics and multi-lingual text-to-text translation to Chinese, Malay, Tamil, Filipino, Thai, Japanese, French, Spanish, and Portuguese.(36) This chatbot responded accurately to open-ended, COVID-19 related questions, achieving an overall accuracy of 0.838 [95% CI: 0.826-0.851].(36) Two studies reported effectiveness in the identification and classification of HPV vaccine misinformation on Reddit.(31; 34) One study found a superior performance of credibility labelling when using the deep learning models compared with other machine learning models (XGBoos) for a relatively larger training set, and the study recommended machine learning BERT because was able to predict most of the misinformation.(32)

Two studies, one experimental randomized study and one machine learning study, found that interacting with a chatbot for a few minutes significantly increased people's intentions to get vaccinated and positively impacted their attitudes toward COVID-19 vaccination,(71) and that providing factual information on Twitter leads to a decrease in misinformation (i.e., suppression) with a time lag.(35)

Credibility labelling

We identified six studies,(32; 34; 67-70) which were conducted in the Netherlands (n=2), U.S. (n=2), U.K. (n=1), and online not linked to a geographic region (n=1). The health topics addressed were COVID-19 (n=2), vaccines (n=2), and other topics (n=2). The study designs were experimental randomized studies (n=4), and machine learning–based approaches (n=2). Overall, four studies enrolled 8,040 participants, 4,737 were women (58.9%).

Among the four experimental randomized studies, three reported no evidence of effectiveness for credibility labelling on the ability to critically evaluate a given message,(67) or accuracy in identifying misinformation.(68; 69) Only one study in the U.S. reported that fact-checking labels attached to misinformation posts made vaccine attitudes more positive than the misinformation control condition, especially when the labelling was performed by universities and health institutions.(70)

Two studies of machine learning–based approaches found those strategies successful for classifying reliable information compared to unreliable information.(32; 34) One study found a superior performance of credibility labelling when using deep learning models compared with XG Boost for a relatively larger training set, and the study recommended BERT (Bidirectional Encoder Representations from Transformers) because it was able to predict most of the misinformation.(32)

Counter-misinformation campaigns

We identified five studies,(37; 42; 44; 47; 66) which were conducted in Canada (n=1), Guatemala (n=1), U.S. (n=1), Zimbabwe (n=1), and Korea (n=1). The health topics addressed were COVID-19 (n=4), and cancer (n=1). The study designs were experimental randomized studies (n=1), quasi experimental (n=1), cross-sectional (n=1), implementation research (n=1), and qualitative research (n=1). Overall, three studies enrolled 2,470 participants, 1,411 were women (57.1%).

Overall, misinformation campaigns were reported effective for stimulating intentions to take protective actions,(37) improving knowledge about a health topic,(44; 66) and reducing beliefs in misinformation.(44) In Guatemala, after adjusting by age, community, sex and language, people from Indigenous Maya communities who watched a misinformation campaign (videos) had 1.78 times the odds (95% CI 1.14 to 2.77) of getting vaccinated compared with those who did not see the videos.(37) In Canada, after completing a misinformation campaign (videos), South Asian youth participants from the Greater Toronto and Hamilton Area, reported an increase in their self-reported knowledge regarding the COVID-19 vaccine from 73.3% to 100.0% (p=0.005), and their self-reported confidence to have a conversation about the vaccine with their unvaccinated community members increased from 63.6% to 100.0% (p=0.002).(44) One study in the U.S. used media monitoring to work with Hispanic social media influencers, volunteers, and celebrities to spread pro-vaccine messaging online, with radio use reaching 26.9 million people, and op-eds reaching 2.9 million people.(42) In a survey of a sample of people in Zimbabwe who received a

misinformation campaign (messages to WhatsApp), a 0.26 standard deviation increase in knowledge about COVID-19 was found.(66)

Strategies found to have limited or no evidence for reducing misinformation

Narrative

We identified two experimental randomized studies,(83; 84) which were conducted in China and the U.S., and both addressing COVID-19.

One study in China found a mediating or suppressing effect of follower count (in social media) in the relationship between a debunker's identity (celebrity, media, or government) and sharing behaviour. However, the debunker's identity did not have a positive effect on the sharing of debunking information when controlling for mediating variables.(84) The other study in the U.S. used advertisements for Facebook providing video testimonials from peer role models promoting vaccination, and found that ads featuring peer modelling with psychological inoculation yielded a significantly higher rate of positive responses than Centers for Disease Controls (CDC) ads (30.5 versus 14.9/1000 people reached in English and 49.7 versus 31.5/1000 in Spanish; $P < 0.001$ for both English and Spanish rate comparisons).(83)

Legislative and other policy

We identified two studies,(20; 49) one conducted in Hong Kong and the other not linked to a geographic region. One study addressed COVID-19 and the other addressed several health topics. One study was an experimental randomized design, and the other an interrupted time series.

The experimental randomized study in Hong Kong conducted an online experiment to test the role of financial incentives and legislation in disseminating online healthcare misinformation. The study found that legislation may deter the sharing of healthcare information that users perceive as true, but cannot deter them from sharing the healthcare misinformation they perceive as fake.(20) The interrupted time series study retrieved all posts published by eligible pages six months before and after a Facebook policy to restrict anti-vaccine posting, and found that although the effect of Facebook's vaccine misinformation policy was statistically significant, the effect size was relatively small, after scaling for the number of subscribers and the volume of anti-vaccine posts remained steady after the policy.(49)

Economic

We identified two experimental randomized studies,(20; 69) which were conducted in Hong Kong and the U.K. and addressed different health topics.

Overall, both studies found that financial incentives might not have a beneficial effect in reducing the willingness to share misinformation.(20; 69) One study in Hong Kong found that financial incentives have a stronger impact on attracting readers to share healthcare misinformation that they perceive to be fake.(20) The study also reported that the power of financial incentives may demonstrate a marginal diminishing effect, while a small financial incentive may help foster healthcare information dissemination, and increasing the size of financial incentives may not foster the same level of additional dissemination effect.(20) One study in the U.K. found that paying participants to be accurate increased an accuracy score but not the proportion of participants correctly guessing the scientific validity of the posts. In contrast, the presence of the pop-up seemed not to directly affect any indicator of accuracy, but increased the reasoning techniques, suggesting an indirect effect of a pop-up.(69)

Curatorial

No evidence identified.

Investigative

No evidence identified.

Conclusions

The evidence indicates that educational strategies, technical and algorithmic strategies aimed to identify misinformation, monitoring and fact-checking, credibility labelling, and counter-misinformation campaigns have a beneficial effect in reducing misinformation. Some strategies like education and monitoring and fact-checking were found to be effective in changing the beliefs of people exposed to misinformation and even stimulating intention to take protective actions. Other strategies like credibility labelling and counter-misinformation campaigns were effective in helping people to identify misinformation or in improving their ability to evaluate a given message critically. Studies that evaluated strategies for debunking misinformation found that fact-checking is more effective when the source is provided, that the format of a correction does not make a considerable difference in effectiveness, and that it might be safe to repeat misinformation when correcting it, even when the audience is unfamiliar with the misinformation. We found limited evidence about narrative, economic and legislative and other policy strategies and evidence that was identified pointed to negligible beneficial effects related to changing the willingness of people to share misinformation. However, no conclusions can be drawn from the limited evidence available about these strategies. Lastly, this version did not identify evidence addressing curatorial and investigative strategies. Given the limited evidence available about narrative, economic and legislative and other policy strategies, as well as the lack of evidence about curatorial and investigative strategies, these are essential areas for further primary studies to be conducted to evaluate their effectiveness.

Next steps

We are currently conducting a new LES focused on assessing strategies to address misinformation related to political institutions that will complement the findings from this LES. In addition, we plan to incorporate the following enhancements to this LES in the next version that we will produce in summer 2024:

- add an enhanced plain-language summary of findings that is co-produced with citizen partners
- update searches to identify new studies that meet our inclusion criteria, which will be incorporated in the findings
- extract and incorporate insights about tactics (e.g., formats, humour, how and what types of evidence is presented, who produced the evidence) that can be used (e.g., using social-media platforms) for strategies identified in the LES (e.g., for monitoring and fact checking, counter-misinformation campaigns and credibility labelling)
- finalize and add risk of bias assessments to summary and interpretation of findings
- conduct GRADE profiles and incorporate in the summary and interpretation of findings.

Table 2: Characteristics of included studies

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Condition studied	Gender/sex analysis
Yang 2023 (36)	Online in several languages	Natural Language Processing chatbot	Technical and algorithmic	Natural Language Processing-based Artificial Intelligence	COVID-19	No
Ma 2023 (76)	China	Behavioural research (experimental randomized study)	Educational	Inoculation theory	COVID-19	No
Abascal 2022 (37)	Guatemala	Before and after study	Counter-misinformation campaigns	Linguistically and culturally tailored social media ad campaign	COVID-19	Yes
Kim 2022 (74)	Online	Behavioural research (experimental randomized study)	Educational	Message-framing tactics	Human papillomavirus (HPV)	No
Kandasamy 2022 (44)	Canada	Cross-sectional and one-group pretest-post-test design	Counter-misinformation campaigns	Public health programme to mobilise and empower (campaign)	COVID-19	No
Xue 2022 (48)	Online in English	Observational study	Monitoring and fact-checking	Natural Language Processing-based Artificial Intelligence	COVID-19	No
Vraga 2022 (64)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Debunking	Sunscreen and skin cancer	No
Folkvord 2022 (67)	The Netherlands	Behavioural research (experimental randomized study)	Credibility labelling	Protective message	COVID-19	No
Winters 2021 (65)	Sierra Leone	Randomized controlled trial	Monitoring and fact-checking Educational	Debunking	Malaria and typhoid	No
Zhang 2021 (70)	US	Behavioural research (experimental randomized study)	Credibility labelling	Fact-checking labelling	Vaccines	No
Stekelenburg 2021 (39)	US	Behavioural research (quasi-experimental study)	Educational	Intervention aimed at increasing belief accuracy	COVID-19	No
Kim 2021 (54)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Message attention and credibility	HPV	No
Du 2021 (31)	Online (Reddit)	Machine Learning–Based Approaches	Technical and algorithmic	Machine learning based methods	HPV	No
Vandormael 2021 (81)	U.S., Mexico, the U.K., Germany, and Spain	Randomized controlled trial	Educational	Video for prevention	COVID-19	No
Bowles 2020 (66)	Zimbabwe	Experimental randomized study	Counter-misinformation campaigns	Dissemination of messages aimed at targeting misinformation	COVID-19	Yes
Gesser-Edelsburg 2018 (27)	Israel	Behavioural research (mixed methods including an experimental randomized design	Monitoring and fact-checking	Information correction	Measles	No

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Condition studied	Gender/sex analysis
		and a descriptive qualitative design)				
Panizza 2022 (69)	UK	Behavioural research (experimental randomized study)	Credibility labelling *not sure if is better classified as fact checking) Economic	Pop-ups meant to advise participants to fact-check and other intervention based on monetary incentives	Climate change Eating chocolate Vaccines for COVID-19	No
Duarte 2022 (38)	Brazil	Before and after study	Educational	Intervention to increase literacy	Coconut oil intake	Yes
Gu 2022 (49)	Online	Interrupted time series	Legislative and other policy	Facebook policy (2019) on user endorsements of vaccine content on its platform	Vaccines	No
Khan 2021 (33)	Online	Machine Learning–Based Approaches	Technical and algorithmic	Algorithm to classify misinformation posts	COVID-19	No
Vijaykumar 2021 (45)	Brazil	Cross-sectional	Educational	Social correction behaviours in WhatsApp	COVID-19	Yes
Kirkpatrick 2021 (75)	US	Behavioural research (experimental randomized study)	Educational	Prospect Theory, Loss-Framing, and Perceived Severity (Youtube)	-Measles, Mumps, Rubella (MMR) -COVID-19	No
Featherstone 2020 (52)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Refutational messages	Vaccines	No
Moore 2016 (28)	US	Before and after study	Educational	Conferences	Escherichia coli O157:H7 in cattle	No
Vraga 2018 (29)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Social correction (Facebook and Twitter)	Zika	No
Vraga 2019 (30)	Online	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Inoculation and observational correction	HPV vaccination	No
Ecker 2020 (51)	Online	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction and backfire effect Fact-checking	HIV among other issues	No
van der Meer 2020 (63)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Corrective information type and source (narrative, educational)	Hypothetical infectious disease outbreak	No
Trevors 2020 (62)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Positive and negative emotional text content in refutational texts	Vaccines	No
Thacker 2020 (61)	US, Australia, Canada	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Refutational messages	Genetically modified food	No

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Condition studied	Gender/sex analysis
Tully 2020 (80)	US	Behavioural research (experimental randomized study)	Educational	News literacy	-Genetically modified food -Seasonal flu vaccine	No
Chao 2021 (84)	China	Behavioural research (experimental randomized study)	Narrative	Debunker identity	COVID-19	No
Tseng 2021 (79)	US	Randomized controlled trial	Educational	Cultivating a critical awareness of flawed scientific claims	Science	No
Steffens 2021 (58)	Australia	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Debunking strategies	Vaccines	No
Swire-Thompson 2021 (60)	US	Behavioural research (experimental study)	Monitoring and fact-checking Educational	Correction	Vaccines and climate change	No
Roozenbeek 2021 (77)	US	Behavioural research (experimental study)	Educational	Asking people to think about the accuracy of a single headline improves "truth discernment" of intentions to share news headlines about COVID-19	COVID-19	No
Meppelink 2021 (34)	The Netherlands	Machine Learning–Based Approaches	Technical and algorithmic Credibility labelling	Supervised machine learning (SML) to classify health-related webpages as 'reliable' or 'unreliable'	-Vaccination in kids -HPV	No
MacFarlane 2021 (55)	Australia	Behavioural research (experimental study)	Monitoring and fact-checking Educational	Refuting	Vitamin E for COVID-19	No
Freeze 2021 (68)	US	Behavioural research (experimental randomized study)	Credibility labelling	Warnings	Affordable care act	No
Ramirez 2022 (83)	US	Behavioural research (experimental pilot study)	Narrative	Psychological inoculation	COVID-19 vaccination	No
Hayawi 2022 (32)	Online	Machine Learning–Based Approaches	Technical and algorithmic Credibility labelling	Machine learning detection framework	COVID-19 vaccination	No
Jiang 2022 (53)	Hong Kong	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Inoculation	COVID-19 vaccination	No
Wang 2022 (35)	US	Machine Learning–Based Approaches	Technical and algorithmic Educational	Factual information vs misinformation (Twitter)	COVID-19	No
Gavin 2022 (73)	Kyrgyzstan, India, and the U.S.	Behavioural research (experimental randomized study)	Educational	Accuracy of nudge intervention	COVID-19	Yes

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Condition studied	Gender/sex analysis
Vlasceanu 2023 (82)	US	Behavioural research (experimental randomized study)	Educational	Belief change	- Child's untreated wandering eye, - Abortion	No
Berlotti 2023 (72)	Italy	Behavioural research (experimental randomized study)	Educational	Prebunking-counterfactual	COVID-19	No
Blomberg 2023 (50)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction	Vitamin C COVID-19	No
Altay 2023 (71)	France	Behavioural research (experimental randomized study)	Technical and algorithmic Educational	Chatbot	COVID-19	No
Mourali 2022 (56)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction and debunking	COVID-19 (masking)	No
Silesky 2023 (42)	US	Implementation research	Counter-misinformation campaigns Monitoring and fact-checking	Media monitoring findings for developing campaigns	COVID-19 vaccination	No
Talabi 2022 (40)	Nigeria	Behavioural research (quasi-experimental study)	Educational	Counselling	COVID-19 vaccination	No
Song 2022 (78)	Hong Kong	Behavioural research (experimental randomized study)	Educational	Evidence types and presentation mode on individuals' responses to corrective messages about COVID-19 on social media	COVID-19	No
Yang 2022 (46)	China	Qualitative research (content analysis)	Monitoring and fact-checking	Rumour debunking	COVID-19	No
Lohiniva 2022 (41)	Ghana	Implementation research	Monitoring and fact-checking	The infodemic management system	COVID-19 vaccination	No
Verduci 2021 (43)	Italy	Implementation research	Monitoring and fact-checking	Chatbot Nutripedia	Nutrition during Pregnancy and Early Life	No
Au 2021 (20)	Hong Kong	Behavioural research (experimental randomized study)	Economic Legislative and other	Financial incentives and legislation	Different health topics	Yes
Sun 2021 (59)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction	COVID-19	Yes
Yoon 2022 (47)	Korea	Qualitative research (content analysis)	Counter-misinformation campaigns	Using network logic of YouTube	Cancer	No
Pennycook 2020 (57)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Nudging	COVID-19	No

Table 3: Summary of findings according to the type of response/strategy

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
Monitoring and fact-checking	**Conditions**: COVID-19 (n=8) Vaccines (n=6) Other (n=10) **Jurisdictions**: Australia (n=2), China (n=1), Ghana (n=1), Hong Kong (n=1), Italy (n=1), Sierra Leone (n=1), the U.S. (n=13), multi-country in Australia, Canada and the U.S. (n=1), NA (n=3) **Sample:** 12.664 participants (5,439 women) (27; 29; 30; 41-43; 46; 48; 50-65) 354 cases of health-related rumours (46) 212,700,000 messages (42) 12,553 Facebook posts and their associated comments (n=122,362) (48)	Experimental randomized (n=19)	• Overall, eleven studies conducted exclusively in the U.S. found that: o if corrective information is present rather than absent, incorrect beliefs based on misinformation are debunked and the exposure to factual elaboration, compared to simple rebuttal, stimulates intentions to take protective actions (63) o when the misinformation is corrected and a source is provided, misperceptions are reduced compared to not providing correction; social corrections without sources are not effective in reducing misperceptions compared to the control (29) o refutational messages increased pro-vaccination attitudes in comparison to misinformation messages without refutation (52) o all refutation texts (with or without positive or negative emotional content) improved knowledge revision (62) o if the key ingredients of a correction are presented, the format of correction does not make a considerable difference (60) o corrections that exposed participants to novel misinformation did not lead to stronger misconceptions compared to a control group never exposed to false claims or corrections; suggesting that it is safe to repeat misinformation when correcting it, even when the audience might be unfamiliar with the misinformation (51; 60) o humorous corrections produce more attention to the misinformation text than non-humorous corrections, in contrast, non-humorous corrections receive higher credibility ratings than humorous corrections, which suggest that credibility and attention to the corrections are not fully aligned, which explain the lack of direct effect of correction strategy on the credibility of the misinformation (54) o both logic-based and humour-based corrections were effective in leading individuals to report greater agreement with expert consensus (30) o crafting positively framed misinformation corrections for the bolstering of message credibility within typically incongruent ideological groups is effective (50) o extended exposure to false claims and debunking attempts weakens the belief that there is an objectively correct answer to how people ought to behave in a situation, which leads to less positive reactions toward the prescribed behaviour (56) o people's perceptions of the severity of the influence of misinformation on others engendered anticipated guilt, which, in turn, strengthened their intentions to correct misinformation related to COVID-19 (59) o people shared false claims about COVID-19 partly because they simply failed to think sufficiently about whether or not the content is accurate when deciding what to share (57) o real-time user corrections were successful in reducing the effects of a misinformation video about sunscreen on people's beliefs, but were not effective in changing their intentions to use it (64) • One multi-country study found that refutation texts supplemented with persuasive information have the potential to substantially impact both readers' final attitudes and knowledge toward a subject (61) • One study in Sierra Leone comprised a three-arm experiment (two interventions + control), one intervention explicitly discussed misinformation and explained why it was incorrect and then provided the scientifically correct information, the other intervention only focused on providing correct information, without directly discussing related misinformation; the study found that both interventions substantially reduced belief in misinformation compared

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
			with the control group, estimates from these analyses suggested that direct debunking may be more effective at countering misinformation (65) • One study in Israel found that both the average satisfaction and reliability level attributed to a theory-based correction intervention were significantly higher than the average satisfaction and reliability level with a common information correction intervention (27) • One study in Australia found no evidence that repeating myths increased agreement with myths compared with the other debunking strategies or the control (58) • One study in Australia investigated the impact of misinformation on hypothetical demand (i.e., willingness-to-pay) for an unproven treatment and the propensity to share misinformation online; the study found that both tentative and enhanced refutations reduced demand for the treatment (18% and 25%, respectively) and misinformation promotion (29% and 55%) (55) • One study in Hong Kong found that participants who received inoculation messages reported higher vaccine attitudes and vaccine intention than those in the conventional health advocacy group, both attitudinal threat and counterarguing moderated the relationships between the experimental conditions and the outcome variables (53)
		Implementation research (n=3)	• In the U.S., media monitoring was used to work with Hispanic social media influencers, volunteers, and celebrities to spread pro-vaccine messaging online, the radio reached 26.9 million people, and the op-eds reached 2.9 million people (42) • In Ghana, a process that identifies misinformation was implemented; the process rated the risk of identified misinformation posts and developed proposed responses to address them (41) • In Italy, a mobile campaign (Nutripedia) was developed specifically to promote correct information for the general population (Nutripedia website) and to address individual doubts and questions from parents (Nutripedia app) (43)
		Qualitative research (n=1)	• One study in China found that since rumours in public health crises often involve different objects, rumour refutation requires various information sources; therefore, different rumour-debunking models apply, those socialized rumour-debunking models could be divided into the following five categories: the government-led model, the media-led model, the scientific community-led model, the rumour-debunking platform-led model, and the multi-agent collaborative model (46)
Counter-misinformation campaigns	**Conditions**: COVID-19 (n=4) Cancer (n=1) **Jurisdictions**: Canada (n=1), Guatemala (n=1), U.S. (n=1), Zimbabwe (n=1), Korea (n=1)	Experimental randomized (n=1)	• After adjusting by age, community, sex and language, people from indigenous Maya communities in Guatemala who watched a misinformation campaign (videos) had 1.78 times the odds (95% CI 1.14 to 2.77) of getting vaccinated compared with those who did not see the videos (37)
		Quasi experimental (n=1)	• After completing a misinformation campaign (videos), South Asian youth participants from the Greater Toronto and Hamilton Area, reported an increase in their self-reported knowledge regarding the COVID-19 vaccine from 73.3% to 100.0% (p=0.005), and their self-reported confidence to have a conversation about the vaccine with their unvaccinated community members increased from 63.6% to 100.0% (p=0.002) (44)
		Cross-sectional (n=1)	• In a survey of a sample of people in Zimbabwe who received a misinformation campaign (messages to WhatsApp), it was found a 0.26 sigma increase in knowledge about COVID-19 (66)
		Implementation research (n=1)	• The study used media monitoring to work with Hispanic social media influencers, volunteers, and celebrities to spread pro-vaccine messaging online, the radio reached 26.9 million people, and the op-eds reached 2.9 million people (42)

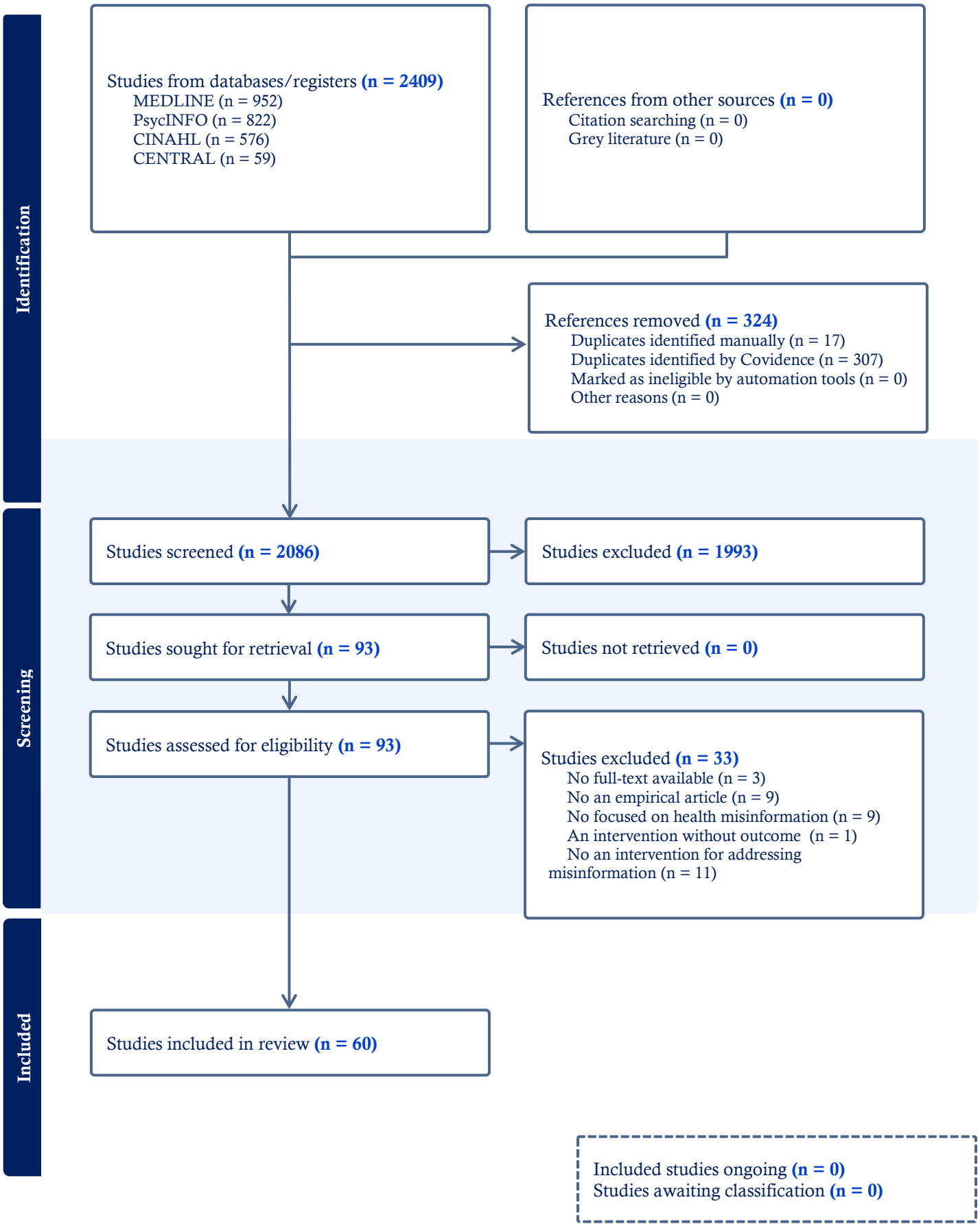
Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	Sample: 2,470 participants (1,411 women) (42; 44; 66) 573 videos (47)	Qualitative research (n=1)	• In Korea, despite government warnings about the risks and dangers of fenbendazole self-administration, this study found that YouTube has reinforced their use, and therefore recommends to health authorities three strategies to fight against social media cancer misinformation; 1) to upload a variety of pertinent information through multiple channels; 2) to consider YouTube's recommendation system, current viewing habits, and information flow network between patients and caregivers; 3) to take an active role in resolving social media misinformation (47)
Credibility labelling	**Conditions**: COVID-19 (n=2) Vaccines (n=2) Other (n=2) **Jurisdictions**: China (n=1), the Netherlands (n=2), U.S. (n=2), U.K. (n=1), NA (n=1) **Sample**: 8,040 participants (4,737 women) (67-70) 15,465,687 tweets (32) 468 Dutch webpages (34)	Experimental randomized (n=4)	• One study in the Netherlands showed that including a protective message in a video with misinformation did not significantly affect the critical evaluation of the message (67) • One study in the U.K. found that pop-ups reminding credibility of the source (lateral reading) seemed not to directly affect any indicator of accuracy in identifying misinformation, but increased the Civic Online Reasoning techniques, suggesting an indirect effect (69) • One study in the U.S. found evidence that valid retrospective warnings of misleading news can help individuals discard erroneous information, although the corrections were weak; however, when informative news is wrongly labelled as inaccurate, these false warnings reduce the news' credibility (68) • Another study in the U.S. reported that fact-checking labels attached to misinformation posts made vaccine attitudes more positive than the misinformation control condition, especially when the labelling was performed by universities and health institutions (70)
		Machine Learning–Based Approaches (n=2)	• In the Netherlands, one study found that the best-performing machine learning model was successful in identifying reliable information, even in terms of out-of-sample prediction, tested on a dataset about HPV vaccination; however, the model is better used to classify reliable information compared to unreliable information (34) • One study performed online found a superior performance of credibility labelling when using the deep learning models compared with XGBoost for a relatively larger training set; the study recommended BERT because it was able to predict most of the misinformation (32)
Educational	**Conditions**: COVID-19 (n=13) Vaccines (n=5) Other (n=9) **Jurisdictions**: Australia (n=2), Brazil (n=2), China (n=1), Hong Kong (n=2), France (n=1), Italy (n=1), Nigeria (n=1), Sierra Leone (n=1), the U.S. (n=11), multi-country (n=3), NA (n=1) **Sample**: 39,782 participants (19,956 women) (28; 35; 38-40;	Experimental randomized (n=20)	• Overall, eight studies conducted exclusively in the U.S. found that: o while speaker expertise did moderate the interaction between framing and evidence, messages were more persuasive when delivered by a non-expert, an expert speaker increased the persuasiveness of videos only when the evidence provided was statistical (75) o all refutation texts (with or without positive or negative emotional content) improve knowledge revision (62) o news literacy messages can alter misinformation perceptions and beliefs, but not with a single message (80) o when testing the efficacy of a structured reading support intervention for evaluation and critique on cultivating a critical awareness of flawed scientific claims in an online setting, there was no difference with people not receiving the intervention (79) o asking people to think about the accuracy of a single headline does not improve "truth discernment" of intentions to share news headlines about COVID-19 (77) o real-time user corrections reduced the misinformation effects of a video about sunscreen on people's beliefs but were not effective in changing their intentions to use it (64) o if the key ingredients of a correction are presented, the format of the correction does not make a considerable difference (60) o changing beliefs trigger corresponding changes in behaviours, in both political and nonpolitical contexts, suggesting that targeting beliefs might be a viable strategy of behavioural change (82) • One multi-country study found that refutation texts supplemented with persuasive information have the potential to substantially impact both readers' final attitudes and knowledge toward a subject (61)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	45; 53; 55; 58; 60-62; 64; 65; 71-82) 22,111,831 English tweets (35)		• One multi-country study found that a video intervention increased prevention knowledge by 7.6% and 5.3% relative to the do-nothing and placebo arms, respectively, it was also found that the video intervention improved behavioural intent toward COVID-19 prevention when compared with the placebo and do-nothing arms (81) • One multi-country study found that the accuracy nudge's effectiveness in reducing the spread of misinformation appeared to depend on location and information type; in India, decreased the willingness to share false general COVID-19 information but did not decrease willingness to share vaccine information, while in the U.S., the nudge decreased willingness to share false information related to the COVID-19 vaccine but not information related to COVID-19 generally (73) • One study in China found that online interventions based on the inoculation theory are effective in enhancing misinformation discrimination, and one of the underlying mechanisms of this effect lies in its promotion of persuasion knowledge (76) • One study in Sierra Leone comprised a three-arm experiment (two interventions + control), one intervention explicitly discussed misinformation and explained why it was incorrect and then provided the scientifically correct information, the other intervention only focused on providing correct information, without directly discussing related misinformation; the study found that both interventions substantially reduced belief in misinformation compared with the control group, estimates from these analyses suggested that direct debunking may be more effective at countering misinformation (65) • One study in Australia found no evidence that repeating myths increased agreement with myths compared with the other debunking strategies or the control (58) • One study in Australia investigated the impact of misinformation on hypothetical demand (i.e., willingness-to-pay) for an unproven treatment and the propensity to share misinformation online; the study found that both tentative and enhanced refutations reduced demand for the treatment (18% and 25%, respectively) and misinformation promotion (29% and 55%) (55) • One study in Hong Kong found that participants who received inoculation messages reported higher vaccine attitudes and vaccine intention than those in the conventional health advocacy group, both attitudinal threat and counterarguing moderated the relationships between the experimental conditions and the outcome variables (53) • One study in Italy reported that among participants with higher levels of conspiracy mentality, those exposed to counterfactual prebunking rated the fake news headline less plausible than those in the control condition and than those exposed to another type of prebunking, that is, forewarning of the existence of misinformation (72) • One Study in France found that interacting with a chatbot for a few minutes significantly increased people's intentions to get vaccinated (s = 0.12) and positively impacted their attitudes toward COVID-19 vaccination (s = 0.23) (71) • One study in Hong Kong found that the presence of statistical evidence in assertions reduced message elaboration, which in turn reduced the effects of the message in correcting misperceptions, decreased perceived message believability and lowered social media users' intentions to further engage with and disseminate the corrective message (78) • One study found that evidence-based messages directly countering misinformation and promoting HPV vaccination in social media environments positively influenced parents in the experimental group compared with those in the control group, which was associated with increased intention to vaccinate among parents of unvaccinated children aged 9 to 14 years (74)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
		Quasi experimental (n=4)	• One study in the U.S. found that an intervention (infographic) aimed at increasing belief accuracy was not effective (39) • One study in the U.S. that exposed participants to four conferences about Escherichia coli O157:H7 in cattle found that all participants agreed that they better understood pre-harvest control, how food safety policy was made, and were confident they could create an effective message about STEC pre-harvest-control (28) • One study in Brazil found that participants after being exposed to the conclusions of a meta-analysis showing that coconut oil does not show superior health benefits when compared to other oils and fats, 73.5% of those who considered coconut oil healthy did not change their opinion (38) • One study in Nigeria reported that social media users who received counselling intervention on the COVID-19 vaccine reported more positive intention to make themselves available for vaccination than their counterparts who were not exposed to such an intervention (40)
		Cros-sectional (n=1)	• One study in Brazil that examined the extent to which WhatsApp users might be willing to correct their peers who might share COVID-19 misinformation, found a pattern of how different demographics influenced the three types of social correction behaviours, younger participants exhibited greater passivity in engaging with social correction; higher educational attainment was associated with providing correction to the original sender; and male participants were more likely to send the correction to the entire group (45)
		Machine Learning–Based Approaches (n=1)	• One study in the U.S. found that providing factual information on Twitter leads to a decrease in misinformation (i.e., suppression) with a time lag (35)
Curatorial	No evidence found		
Narrative	**Conditions**: COVID-19 (n=2) **Jurisdictions**: China (n=1), U.S. (n=1) **Sample**: 1,196 observations (84)	Experimental randomized (n=2)	• One study in China found a mediating or suppressing effect of follower count (in social media) in the relationship between a debunker's identity (celebrity, media, or government) and sharing behaviour; however, the debunker's identity did not have a positive effect on the sharing of debunking information when controlling for mediating variables (84) • One study in the U.S. used advertisements for Facebook providing video testimonials from peer role models promoting vaccination; ads featuring peer modelling with psychological inoculation yielded a significantly higher rate of positive responses than CDC ads (30.5 versus 14.9/1000 people reached in English and 49.7 versus 31.5/1000 in Spanish; $P < 0.001$ for both English and Spanish rate comparisons) (83)
Technical and algorithmic	**Conditions**: COVID-19 (n=5) VPH (n=2) **Jurisdictions**:	Machine Learning–Based Approaches (n=6)	• One study developed a chatbot named DR-COVID with an ensemble Natural Language Processing (NLP) model on the Telegram platform, then evaluated various performance metrics and multi-lingual text-to-text translation to Chinese, Malay, Tamil, Filipino, Thai, Japanese, French, Spanish, and Portuguese; the chatbot responded accurately to open-ended, COVID-19 related questions, achieving an overall accuracy of 0.838 [95% CI: 0.826-0.851] (36) • A machine learning-based approach was effective in the identification and classification of HPV vaccine misinformation on Reddit and may be generalizable to other social media platforms (31) • Another study trained machine learning algorithms to identify COVID-19-related misinformation and found a better performance when trained with extracted features from a COVID-19 fake news dataset (33)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	France (n=1), the Netherlands (n=1), U.S. (n=1), NA (n=4) **Sample**: 701 participants (291 women) (71) 37,577,518 tweets (32; 35) 468 Dutch webpages (34)		• One study found a superior performance of credibility labelling when using the deep learning models compared with XG Boost for a relatively larger training set; the study recommended BERT because was able to predict most of the misinformation (32) • In the Netherlands, one study found that the best-performing machine learning model was successful in identifying reliable information, even in terms of out-of-sample prediction, tested on a dataset about HPV vaccination; however, the model is better used to classify reliable information compared to unreliable information (34) • One study in the U.S. found that providing factual information on Twitter leads to a decrease in misinformation (i.e., suppression) with a time lag (35)
		Experimental randomized (n=1)	• One Study in France found that interacting with a chatbot for a few minutes significantly increased people's intentions to get vaccinated (s = 0.12) and positively impacted their attitudes toward COVID-19 vaccination (s = 0.23) (71)
Economic	**Conditions**: Different health topics (n=2) **Jurisdictions**: Hong Kong (n=1), the U.K. (n=1) **Sample**: 5,750 participants (3,479 women) (20; 69)	Experimental randomized (n=2)	• One study in Hong Kong found that financial incentives have a stronger impact on attracting readers to share healthcare misinformation that they perceive to be fake; perceived believability and financial incentives may increase the likelihood of sharing healthcare information (20) ○ The power of financial incentives may demonstrate a marginal diminishing effect, while a small financial incentive may help foster healthcare information dissemination, increasing the size of financial incentives may not foster the same level of additional dissemination effect (20) • One study in the U.K. found that paying participants to be accurate increased an accuracy score but not the proportion of participants correctly guessing the scientific validity of the posts, by contrast, the presence of the pop-up seemed not to affect directly any indicator of accuracy, but increased the Civic Online Reasoning techniques, suggesting an indirect effect of the pop-up (69)
Legislative and other policy	**Conditions**: Vaccines (n=1) Other (n=1) **Jurisdictions**: Hong Kong (n=1), NA (n=1) Sample: 363 participants (137 women) (20) 172 anti- and pro-vaccine Facebook pages (49)	Experimental randomized (n=1)	• One study in Hong Kong conducted an online experiment to test the role of financial incentives and legislation in disseminating online healthcare misinformation, the study found that legislation may deter the sharing of healthcare information that users perceive as true but cannot deter them from sharing the healthcare misinformation they perceive as fake (20)
		Interrupted-time series (n=1)	• One study retrieved all posts published by eligible pages six months before and after a Facebook policy to restrict anti-vaccine posting; the study found that although the effect of Facebook's vaccine misinformation policy was statistically significant, the effect size was relatively small after scaling for the number of subscribers and the volume of anti-vaccine posts remained steady after the policy (49)
Investigative	No evidence		

Figure 3: Prisma Chart



References

1. General OotUSS. Confronting health misinformation. The U.S. Surgeon General's Advisory on Building a Healthy Information Environment; 2021.
2. Janmohamed K, Walter N, Nyhan K, et al. Interventions to Mitigate COVID-19 Misinformation: A Systematic Review and Meta-Analysis. *Journal of Health Communication* 2021;26(12): 846-857.
3. do Nascimento I, Pizarro A, Almeida J, et al. Infodemics and health misinformation: A systematic review of reviews. 2022.
4. Linkins L. Critical appraisal process for assessment of public health measures for COVID-19 cohort studies. Hamilton, Canada: Health Information Research Unit, McMaster University; 2023.
5. McMaster Health Forum. Suite of living evidence syntheses about COVID-19 public health and social measures. 2023. https://www.mcmasterforum.org/spark-action/suite-of-living-evidence-syntheses-about-covid-19-public-health-and-social-measures (accessed 26 September 2023.
6. Calo R, Coward C, Spiro ES, Starbird K, West JD. How do you solve a problem like misinformation? *Science Advances* 2021;7(December): 8-10.
7. Alan Turing Institute. Understanding vulnerability to online health misinformation. *Brief* 2021(March).
8. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: Implications for public health practice. *Journal of Public Health Management and Practice* 2017;23(4): 380-387.
9. Jaiswal J, LoSchiavo C, Perlman DC. Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. *AIDS and Behavior* 2020;24(10): 2776-2780.
10. Xiaoli N, 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.
11. Plan International. The Truth Gap: The State of the World´s Girls. 2021. https://plan-international.org/uploads/2022/02/sotwgr2021-commsreport-en.pdf (accessed 12 September 2022.
12. Grimes K, Matlow A, Tholl B, Dickson G, Taylor D, Chan M-K. Leaders supporting leaders: Leaders' role in building resilience and psychologically healthy workplaces during the pandemic and beyond. *Healthcare Management Forum* 2022;35(4): 213-217.
13. Global Commission on Evidence to Address Societal Challenges. The Evidence Commission report: A wake-up call and path forward for decision- makers, evidence intermediaries, and impact-oriented evidence producers. Hamilton, Ontario: McMaster Health Forum; 2022. 144-144 p.
14. Broadband Commission research report on 'Freedom of Expression Addressing Disinformation on the Internet'. Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression2020. 348-348 p.
15. Swire-Thompson B, Lazer D. Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health* 2019;41: 433-451.
16. Broniatowski DA, Jamison AA, Qi S, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health* 2018;108(10): 1378-1384.
17. Guidry JPD, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine* 2015;33(39): 5051-5056.
18. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science (New York, NY)* 2018;359(6380): 1146-1151.
19. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research* 2021;23(1): 1-17.
20. Au CH, Ho KKW, Chiu DKW. Stopping healthcare misinformation: The effect of financial incentives and legislation. *Health Policy* 2021;125(5): 627-633.
21. Ceretti E, Covolo L, Cappellini F, et al. Evaluating the Effectiveness of Internet-Based Communication for Public Health: Systematic Review. *J Med Internet Res* 2022;24(9): e38541.

22. Chou WS, Gaysynsky A, Vanderpool RC. The COVID-19 Misinfodemic: Moving Beyond Fact-Checking. *Health Educ Behav* 2021;48(1): 9-13.

23. Czerniak K, Pillai R, Parmar A, Ramnath K, Krocker J, Myneni S. A scoping review of digital health interventions for combating COVID-19 misinformation and disinformation. *J Am Med Inform Assoc* 2023;30(4): 752-760.

24. Walter N, Murphy ST. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs* 2018;85(3): 423-441.

25. Akhtar P, Ghouri AM, Khan HUR, et al. Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Ann Oper Res* 2022: 1-25.

26. Pomeranz JL, Schwid AR. Governmental actions to address COVID-19 misinformation. *Journal of Public Health Policy* 2021;42(2): 201-210.

27. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch GS. Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. *PLoS ONE* 2018;13(12): e0209505.

28. Moore DA, Smith DR, Sischo WM, Heaton K, Besser TE. Escherichia coli O157:H7--Discerning Facts from Fiction: An Integrated Research and Extension Project for Multiple Audiences. *Zoonoses and public health* 2016;63(1): 72-81.

29. Vraga EK, Bode L. I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society* 2018;21(10): 1337-1353.

30. Vraga EK, Kim SC, Cook J. Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *Journal of Broadcasting & Electronic Media* 2019;63(3): 393-414.

31. Du J, Preston S, Sun H, et al. Using machine learningbased approaches for the detection and classification of human papillomavirus vaccine misinformation: Infodemiology study of reddit discussions. *Journal of Medical Internet Research* 2021;23(8): e26478.

32. Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS. ANTi-Vax: A novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* 2022;203(Abdelminaam, D.S., Ismail, F.H., Taha, M., Taha, A., Houssein, E.H., & Nabil, A. (2021). CoAID-DEEP: an optimized intelligent framework for automated detecting COVID-19 misleading information on Twitter. IEEE Access, 9, 27840-27867. https://pubmed.ncbi.nl): 23-30.

33. Khan S, Hakak S, Deepa N, Prabadevi B, Dev K, Trelova S. Detecting COVID-19-Related Fake News Using Feature Extraction. *Frontiers in public health* 2021;9(101616579): 788074.

34. Meppelink CS, Hendriks H, Trilling D, van Weert JCM, Shao A, Smit ES. Reliable or not? An automated classification of webpages about early childhood vaccination using supervised machine learning. *Patient Education and Counseling* 2021;104(6): 1460-1466.

35. Wang Y, Gao S, Gao W. Investigating dynamic relations between factual information and misinformation: Empirical studies of tweets related to prevention measures during COVID-19. *Journal of Contingencies and Crisis Management* 2022;30(4): 427-439.

36. Yang LWY, Ng WY, Lei X, et al. Development and testing of a multi-lingual Natural Language Processing-based deep learning system in 10 languages for COVID-19 pandemic crisis: A multi-center study. *Frontiers in public health* 2023;11((Yang) Ministry of Health Holdings, Singapore(Ng, Tan, Lim, Gunasekeran, Wong, Ting) Singapore National Eye Center, Singapore Eye Research Institute, Singapore(Ng, Wang, Gunasekeran, Ho, Wong, Ting) Duke-National University of Singapore Medical School, Na): 1063466.

37. Abascal Miguel L, Lopez E, Sanders K, et al. Evaluating the impact of a linguistically and culturally tailored social media ad campaign on COVID-19 vaccine uptake among indigenous populations in Guatemala: a pre/post design intervention study. *BMJ Open* 2022;12(12): e066365.

38. Duarte AC, Spiazzi BF, Merello EN, et al. Misinformation in nutrition through the case of coconut oil: An online before-and-after study. *Nutrition, metabolism, and cardiovascular diseases : NMCD* 2022;32(6): 1375-1384.

39. Stekelenburg AV, Schaap G, Veling H, Buijzen M. Investigating and Improving the Accuracy of US Citizens' Beliefs about the COVID-19 Pandemic: Longitudinal Survey Study. *Journal of Medical Internet Research* 2021;23(1): 24069.

40. Talabi FO, Ugbor IP, Talabi MJ, et al. Effect of a social media-based counselling intervention in countering fake news on COVID-19 vaccine in Nigeria. *Health Promotion International* 2022;37(2): 1-10.

41. Lohiniva A-L, Nurzhynska A, Hudi A-h, Anim B, Aboagye DC. Infodemic Management Using Digital Information and Knowledge Cocreation to Address COVID-19 Vaccine Hesitancy: Case Study From Ghana. *JMIR Infodemiology* 2022;2(2): 234-243.

42. Silesky MD, Panchal D, Fields M, et al. A Multifaceted Campaign to Combat COVID-19 Misinformation in the Hispanic Community. *Journal of Community Health* 2023;48(2): 286-294.

43. Verduci E, Vizzuso S, Frassinetti A, et al. Nutripedia: The Fight against the Fake News in Nutrition during Pregnancy and Early Life. *Nutrients* 2021;13(9): 2998-2998.

44. Kandasamy S, Ariyarajah A, Limbachia J, et al. South Asian Youth as Vaccine Agents of Change (SAY-VAC): evaluation of a public health programme to mobilise and empower South Asian youth to foster COVID-19 vaccine-related evidence-based dialogue in the Greater Toronto and Hamilton Area, Canada. *BMJ Open* 2022;12(9): e061619.

45. Vijaykumar S, Rogerson DT, Jin Y, de Oliveira Costa MS. Dynamics of social corrections to peers sharing COVID-19 misinformation on WhatsApp in Brazil. *Journal of the American Medical Informatics Association : JAMIA* 2021;29(1): 33-42.

46. Yang F, Ren Y, Wang S, Zhang X. Health-Related Rumor Control through Social Collaboration Models: Lessons from Cases in China during the COVID-19 Pandemic. *Healthcare (2227-9032)* 2022;10(8): 1475-N.PAG.

47. Yoon HY, You KH, Kwon JH, et al. Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study. *Journal of Medical Internet Research* 2022;24(11): N.PAG-N.PAG.

48. Xue H, Gong X, Stevens H. COVID-19 Vaccine Fact-Checking Posts on Facebook: Observational Study. *Journal of Medical Internet Research* 2022;24(6): e38423.

49. Gu J, Dor A, Li K, et al. The impact of Facebook's vaccine misinformation policy on user endorsements of vaccine content: An interrupted time series analysis. *Vaccine* 2022;40(14): 2209-2214.

50. Blomberg M. Communicating veracity: The impact of emotion and multimodal resources for correcting online misinformation in older adults. *Dissertation Abstracts International Section A: Humanities and Social Sciences* 2023;84(3-A): No-Specified.

51. Ecker UKH, Lewandowsky S, Chadwick M. Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications* 2020;5(Arkes, H.R., Boehm, L.E., and Xu, G. (1991). Determinants of judged validity. Journal of Experimental Social Psychology, 27, 576-605 https://dx.doi.org/10.1016/0022-1031(91)90026-3 1992-11516-001.Ayers, M.S., and Reder, L.M. (1998). A theoretical review o).

52. Featherstone JD, Zhang J. Feeling angry: the effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. *Journal of Health Communication* 2020;25(9): 692-702.

53. Jiang LC, Sun M, Chu TH, Chia SC. Inoculation works and health advocacy backfires: Building resistance to COVID-19 vaccine misinformation in a low political trust context. *Frontiers in Psychology* 2022;13(Agley, J., & Xiao, Y. (2021). Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. BMC Public Health 21:89 https://pubmed.ncbi.nlm.nih.gov/33413219 https://dx.doi.org/10.1186/s12889-020-1).

54. Kim SC, Vraga EK, Cook J. An Eye Tracking Approach to Understanding Misinformation and Correction Strategies on Social Media: The Mediating Role of Attention and Credibility to Reduce HPV Vaccine Misperceptions. *Health Communication* 2021;36(13): 1687-1696.

55. MacFarlane D, Tay LQ, Hurlstone MJ, Ecker UKH. Refuting spurious COVID-19 treatment claims reduces demand and misinformation sharing. *Journal of Applied Research in Memory and Cognition* 2021;10(2): 248-258.

56. Mourali M, Drake C. Debunking Health Misinformation on Social Media: the Challenge of Dynamic Conversations. *Journal of Medical Internet Research* 2022;24(3): e34831.

57. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science* 2020;31(7): 770-780.

58. Steffens MS, Dunn AG, Marques MD, Danchin M, Witteman HO, Leask J. Addressing myths and vaccine hesitancy: A randomized trial. *Pediatrics* 2021;148(5).

59. Sun Y, Oktavianus J, Wang S, Lu F. The Role of Influence of Presumed Influence and Anticipated Guilt in Evoking Social Correction of COVID-19 Misinformation. *Health Communication* 2022;37(11): 1368-1377.

60. Swire-Thompson B, Cook J, Butler LH, Sanderson JA, Lewandowsky S, Ecker UKH. Correction format has a limited role when debunking misinformation. *Cognitive Research: Principles and Implications* 2021;6(Appelt, K.C., Hardisty, D.J., and Weber, E.U. (2011). Asymmetric discounting of gains and losses: A query theory account. Journal of Risk and Uncertainty, 43, 107-126 https://dx.doi.org/10.1007/s11166-011-9125-1Baddeley, A.D., and Hitch, G.J. (1993). The).

61. Thacker I, Sinatra GM, Muis KR, et al. Using persuasive refutation texts to prompt attitudinal and conceptual change. *Journal of Educational Psychology* 2020;112(6): 1085-1099.

62. Trevors G, Kendeou P. The effects of positive and negative emotional text content on knowledge revision. *The Quarterly Journal of Experimental Psychology* 2020;73(9): 1326-1339.

63. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication* 2020;35(5): 560-575.

64. Vraga EK, Bode L, Tully M. The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer. *Health Communication* 2022;37(13): 1622-1630.

65. Winters M, Oppenheim B, Sengeh P, et al. Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: Evidence from a randomised controlled trial in Sierra Leone. *BMJ Global Health* 2021;6(11): e006954.

66. Bowles J, Larreguy H, Liu S. Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PLoS ONE* 2020;15(10 October): e0240005.

67. Folkvord F, Snelting F, Anschutz D, et al. Effect of Source Type and Protective Message on the Critical Evaluation of News Messages on Facebook: Randomized Controlled Trial in the Netherlands. *Journal of Medical Internet Research* 2022;24(3): e27945.

68. Freeze M, Baumgartner M, Bruno P, et al. Fake claims of fake news: Political misinformation, warnings, and the tainted truth effect. *Political Behavior* 2021;43(4): 1433-1465.

69. Panizza F, Ronzani P, Martini C, Mattavelli S, Morisseau T, Motterlini M. Lateral reading and monetary incentives to spot disinformation about science. *Scientific reports* 2022;12(1): 5678.

70. Zhang J, Featherstone JD, Calabrese C, Wojcieszak M. Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine* 2021;145((Zhang, Featherstone, Calabrese, Wojcieszak) Department of Communication, University of California Davis, One Shields Ave, Davis, CA 95616, United States(Zhang) Department of Public Health Sciences, University of California Davis, One Shields Ave, Davis): 106408.

71. Altay S, Hacquin A-S, Chevallier C, Mercier H. Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. *Journal of Experimental Psychology: Applied* 2023;29(1): 52-62.

72. Bertolotti M, Catellani P. Counterfactual thinking as a prebunking strategy to contrast misinformation on COVID-19. *Journal of Experimental Social Psychology* 2023;104(Anthony, A., & Moulding, R. (2019). Breaking the news: Belief in fake news and conspiracist beliefs. Australian Journal of Psychology, 71(2), 154-162 https://dx.doi.org/10.1111/ajpy.12233 https://dx.doi.org/10.1111/ajpy.12233 2018-62379-001.Bago, B., Rand): 1-10.

73. Gavin L, McChesney J, Tong A, Sherlock J, Foster L, Tomsa S. Fighting the spread of COVID-19 misinformation in Kyrgyzstan, India, and the United States: How replicable are accuracy nudge interventions? *Technology, Mind, and Behavior* 2022;3(3): No-Specified.

74. Kim SJ, Schiffelbein JE, Imset I, Olson AL. Countering Antivax Misinformation via Social Media: Message-Testing Randomized Experiment for Human Papillomavirus Vaccination Uptake. *Journal of Medical Internet Research* 2022;24(11): e37559.

75. Kirkpatrick AW, Park M, Domgaard S, Zhao W, Steinberg C, Hsu Y. Vaccine Videos and Information Sharing: The Effects of Framing, Evidence Type, and Speaker Expertise. *Journal of Health Communication* 2021;26(9): 608-617.

76. Ma J, Chen Y, Zhu H, Gan Y. Fighting COVID-19 Misinformation through an Online Game Based on the Inoculation Theory: Analyzing the Mediating Effects of Perceived Threat and Persuasion Knowledge. *International Journal of Environmental Research and Public Health* 2023;20(2): 980.

77. Roozenbeek J, Freeman ALJ, van der Linden S. How accurate are accuracy-nudge interventions? A preregistered direct replication of Pennycook et al. (2020). *Psychological Science* 2021;32(7): 1169-1178.

78. Song Y, Wang S, Xu Q. Fighting misinformation on social media: effects of evidence type and presentation mode. *Health Education Research* 2022;37(3): 185-198.

79. Tseng AS, Bonilla S, MacPherson A. Fighting "bad science" in the information age: The effects of an intervention to stimulate evaluation and critique of false scientific claims. *Journal of Research in Science Teaching* 2021;58(8): 1152-1178.

80. Tully M, Vraga EK, Bode L. Designing and testing news literacy messages for social media. *Mass Communication & Society* 2020;23(1): 22-46.

81. Vandormael A, Adam M, Greuel M, et al. The Effect of a Wordless, Animated, Social Media Video Intervention on COVID-19 Prevention: Online Randomized Controlled Trial. *JMIR public health and surveillance* 2021;7(7): e29060.

82. Vlasceanu M, McMahon CE, Van Bavel JJ, Coman A. Political and nonpolitical belief change elicits behavioral change. *Journal of Experimental Psychology: Applied* 2023(Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. British Journal of Social Psychology, 29(2): No-Specified.

83. Ramirez AG, Despres C, Chalela P, et al. Pilot study of peer modeling with psychological inoculation to promote coronavirus vaccination. *Health Education Research* 2022;37(1): 1-6.

84. Chao F, Wang X, Yu G. The influence of the debunker's identity and emotional expression on the sharing behavior of debunking information. *Frontiers in Psychology* 2021;12(Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. Sci. Adv. 6:eaay3539 https://pubmed.ncbi.nlm.nih.gov/32284969 https://dx.doi.org/10.1126/sciadv.aay3539A).

85. Ecker UKH, O'Reilly Z, Reid JS, Chang EP. The effectiveness of short-format refutational fact-checks. *British Journal of Psychology* 2020;111(1): 36-54.

86. van der Linden S. Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine* 2022;28(3): 460-467.

Vélez CM, Wilson MG, Patiño-Lugo DF, Cura J, Smith M, Brehaut J, Gretton J, Nicklin W, Lavis JN, Misinformation Living Evidence Synthesis Team*. Living evidence synthesis 22.1: Impact of strategies to mitigate health-related misinformation in diverse settings and populations: Living Evidence Synthesis. Hamilton: McMaster Health Forum, 15 April 2024.

This evidence synthesis 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 living evidence synthesis 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.

*The Misinformation Living Evidence Synthesis Team includes collaborators on the overall project and includes Mpho Begin, Timothy Caulfield, Heather Devine, Graham Dickson, Jeremy Gretton, Kelly Grimes, M. Mustafa Hirji, Alfonso Iorio, Nina Jetha, Jennifer Kitts, Cynthia Lisée, Tamara Navarro, Justin Presseau, Thomas Piggott, Jude Porter, Gabrielle Plamondon and Bill Tholl.