McMaster University | Health Forum

Living Evidence Synthesis

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

25 February 2026

[MHF product code: LES 22.2]

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

*Note that updates and revisions from version 1 are highlighted in yellow

Context

- 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; 4)
- In addition, misinformation can affect some individuals and groups more than others, leading them to be more exposed to health threats, resulting in greater social and health inequities.(7)
  - For example, individuals with lower digital, numerical and health literacy and/or cognitive skills are more vulnerable to misinformation,(5; 6) and groups most vulnerable to health-related misinformation include younger people, those with lower educational attainment, racial minorities, social media users,(8) and women, trans, and nonbinary people.(9)
- 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 decisions.
- While there are many strategies that have been identified to address misinformation, there is a need to evaluate their effects.

Questions

1) What is the impact of strategies to mitigate misinformation in diverse settings, and across diverse populations?
2) Which behavioural science-informed change approaches, techniques, and barriers/enablers are targeted in trials of interventions to mitigate health-related misinformation?

High-level summary of key findings

Note that this second version of the living evidence synthesis has been methodologically enhanced from the first version by: 1) assessing the risk of bias of all included studies; and 2) conducting a behavioural science analysis of change techniques included in the identified evidence (by two behavioural scientists – JP and AL) that can be leveraged to mitigate health-related misinformation. Methodological details for these enhancements have been included below and findings updated. Note that while we enhanced the methods used in this version, we did not update the search (last was conducted on 4 May 2023).

Overview of evidence identified

- We identified 2,086 unique articles (after removing duplicates) and included 59 studies (mostly published in the last four years), which included 41 randomized controlled trials, six quasi-experimental studies, six studies that used machine learning approaches, three implementation research studies, two cross-sectional studies, and one other type of observational study.

- The risk of bias was assessed using the RoB2 tool (N=36 studies, n=79 experiments), the ROBINS-I tool (N=12 studies, n=14 outcomes), and the CASP tool (N=6 studies); five studies were not assessed because they come from machine learning-based approaches for which we do not have an appropriate quality appraisal tool.
- Using the RoB2 tool, the risk of bias was assessed as low in 11, some concerns in 27 and high in 41 experiments; using the ROBINS-I tool, the risk of bias was assessed as low in four outcomes, moderate in two, and serious in eight; using CASP, the methodological quality was appraised as moderate in all six studies.
- 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=23), 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 educational (n=26), monitoring and fact-checking (n=24), technical and algorithmic (n=7), credibility labelling (n= 6), counter-misinformation campaigns (n=5), narrative (n=2), 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 supporting 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.

Behavioural science-informed change techniques for addressing misinformation

- We identified n=37 studies within the main review that included behavioural science-related content.
- Thirteen behavioural science theories were identified, with inoculation theory being the most prevalent (n=7 studies); 16 implicit (i.e., not directly citing established behavioural science) approaches were also identified, with the usage of source credibility being the most prevalent (n=9 studies).
- Multiple behaviour change techniques were identified, including leveraging rewards (e.g., providing an incentive for an outcome produced) and comparing behaviours (e.g., comparing one's own performance to another's performance).

- The use of behaviour change techniques that involved providing a comparison of outcomes was especially effective, such as involving a credible source; and provision of strong pros and weak cons for doing the behaviour.
  - Studies tended to report using a relatively small number of behaviour change techniques (range 0-4 techniques identified across studies).
- Knowledge and skills, social influences, and intentions to do the behaviour, beliefs about the consequences, and emotions were all frequently targeted barriers and enablers across studies.

Background

Increasing digitalization and the use of social media are a double-edged sword.(10) It creates opportunities to rapidly communicate and disseminate information to address social challenges, making it an important tool for reaching individuals and communities.(2; 10) However, as emphasized by the United Nations (UN),(11) digital technologies and social media also have the potential to introduce misinformation to citizens.(10) 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; 4) 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)

It is therefore not surprising that misinformation and disinformation continue to be identified as leading global challenges that need to be the focus of concerted efforts to address them. For example, in October 2024, the World Economic Forum rated misinformation and disinformation as the top global risk over the next two years and the 5th highest global risk over the next 10 years.(12) Moreover, the other global risks identified can either be the focus of misinformation (e.g., climate change) or drivers of it (e.g., political polarization and adverse outcomes from the use of AI).(12; 13) In addition, a Lancet editorial recently made the following statement about misinformation that conveys the concern and efforts that should be afforded to addressing misinformation: "Health misinformation was weaponised as propaganda, exploiting fear, undermining public trust, and hindering collective action in critical moments. Today, misleading social media content pervades information on cancer prevention and treatment; can lead patients to abandon evidence-based treatments in favour of influencer-backed alternatives; downplays the seriousness of mental health conditions; and promotes unregulated supplements claiming to work for everything from weight loss to reversal of ageing."(14)

Adding more concern, the impacts of misinformation can affect some members of society more than others, including those with lower digital, numerical and health literacy and/or cognitive skills which make them more vulnerable to misinformation.(5; 6) As a result, these groups may be more exposed to health threats, leading to even greater social and health inequities.(7) In addition, 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.(8) 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 and their health.(9) 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 (15) and in the Global Commission on Evidence.(11) Individuals can engage with misinformation through different sources.(16) In particular, while social-media platforms are a key driver of misinformation,(3; 17) 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.(16) For example, a study that analyzed 800 vaccine-related Pinterest posts found that 74% were anti-vaccine in sentiment.(18; 19)

Although vaccines were the most common topic of misinformation before COVID-19,(20) other common topics for misinformation include reproductive health, substance use or smoking, non-communicable diseases, pandemics, eating disorders, and medical treatments.(20) Governments have employed a variety of strategies designed to debunk misinformation, including monitoring and fact-checking, economic incentives, and legislative policies.(21-24) 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.(25) 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.(26) 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.(15) The report provides a framework for ten potentially effective responses to misinformation and the possible intersections with freedom-of-expression rights,(15) 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.(27)

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") (15; 27)

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		

Response/ strategy	Description	Purpose of the strategy	Intersections with freedom-of-expression rights
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: 1) strategies to address health-related misinformation in different settings and across diverse populations; and 2) behavioural science-informed change techniques, and barriers/enablers to mitigate health-related misinformation.

Methods

This second version of the living evidence synthesis has been methodologically enhanced from the first version by: 1) assessing the risk of bias of all included studies; and 2) conducting a behavioural science analysis of change techniques and barriers/enablers that can be considered to mitigate health-related misinformation. Methodological details for these enhancements have been included below and findings updated.

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 the database inception until 4 May 2023. 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, observational, and qualitative studies for any populations, settings, and diseases (we did not limit to only COVID misinformation). Outcomes considered included change in attitudes/behaviour, health benefits, harms, and costs. Our research questions were not restricted to assessing the effectiveness of interventions. Rather, the aim was broader, including how they were designed, implemented, experienced, and understood across contexts. We excluded evidence syntheses but reviewed their references to identify additional studies to include. Additionally, we identified all studies that belong to the field of

behavioural science, specifically by recognizing those with hypotheses related to the study of human behaviour, focused on how individuals and groups act and make decisions.

Data extraction was conducted by one team member and checked by another for consistency. For quality appraisal, we used the Cochrane risk of bias tool (RoB 2) for any experimental studies and ROBINS-I for observational study designs, which was enhanced for assessment of cohort studies in a series of living evidence syntheses evaluating COVID-19 public health and social measures.(28; 29) For qualitative studies, we used CASP qualitative research checklist.(30; 31) For behavioural science studies, the risk of bias assessment was performed for each outcome assessed, given that each outcome was measured through a specific experiment with their particular methodological nuances.

We analyzed the evidence by presenting narrative evidence profiles across studies by outcome measure, and then by behavioural science-informed change theories, techniques, and barriers/enablers. The behavioural science analysis was conducted by two team members (JP and AL) using behavioural science taxonomies and frameworks. Specifically, we identified the behaviour change techniques (BCTs) used in the intervention conditions, the factors targeted for change as intended barriers or enablers, and the specific theories drawn upon to inform the interventions.

What we found

After removing duplicates, we screened 2,085 titles and abstracts; and selected 93 as eligible. After full-text review, 59 studies were included in this LES (see Figure 3 for the PRISMA chart, and Appendix 2 for the list of studies excluded). Most studies were published in 2020 and after, only four studies were published before 2020.(32-35) 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=37) came from behavioural sciences. The specific study designs were experimental randomized studies (n=41),(21; 32; 34-76) machine learning–based approaches (n=6),(59; 61; 63; 77-79) quasi experimental (n=5),(33; 43; 76; 80; 81) implementation research (n=3),(69; 82; 83) cross-sectional (n=2),(84; 85) qualitative research (n=2),(86; 87) observational study (n=2).(88) (89)

Thirty-six studies were appraised using the RoB2 tool; 32 of them were behavioural science studies, which comprised a total of 75 experiments.(32; 34-72) The risk of bias was assessed as low in 11, some concerns in 24 and high in 40 experiments. The other four studies were unique experiments: three with some concerns,(40; 44; 45) and one with a high risk of bias.(55) Twelve studies were assessed using the ROBINS-I tool, with five being behavioural science studies,(21; 73-76) comprising seven different outcomes: four with a low risk of bias and three with a serious risk of bias. Seven studies measured only one outcome: two with a moderate risk of bias (80; 89) and five with a serious risk of bias.(33; 81; 84; 85; 88) The methodological quality of six studies was appraised using the CASP qualitative research checklist; all six were of moderate quality.(69; 79; 82; 83; 86; 87) Five studies were not assessed because they come from machine learning-based approaches for which we do not have an appropriate quality appraisal tool.(59; 61; 63; 77; 78)

The interventions addressed in the included studies (some studies addressed more than one strategy) were monitoring and fact-checking (n=24),(32; 34; 35; 39; 40; 42; 47; 50-52; 54; 57; 58; 62; 67-69; 71; 72; 82; 83; 86; 88; 90) counter-misinformation campaigns (n=5),(44; 69; 80; 84; 87) credibility labelling (n= 6),(38; 41; 46; 59-61) educational (n=26),(33; 36; 37; 39; 40; 43; 45; 48; 50; 52-58; 62-66; 70; 75; 76; 81; 85) narrative (n=2),(73; 74) technical and algorithmic (n=7),(59; 61; 63; 66; 77-79) economic (n=2),(21; 46) and legislative and other policy (n=2).(21; 89)

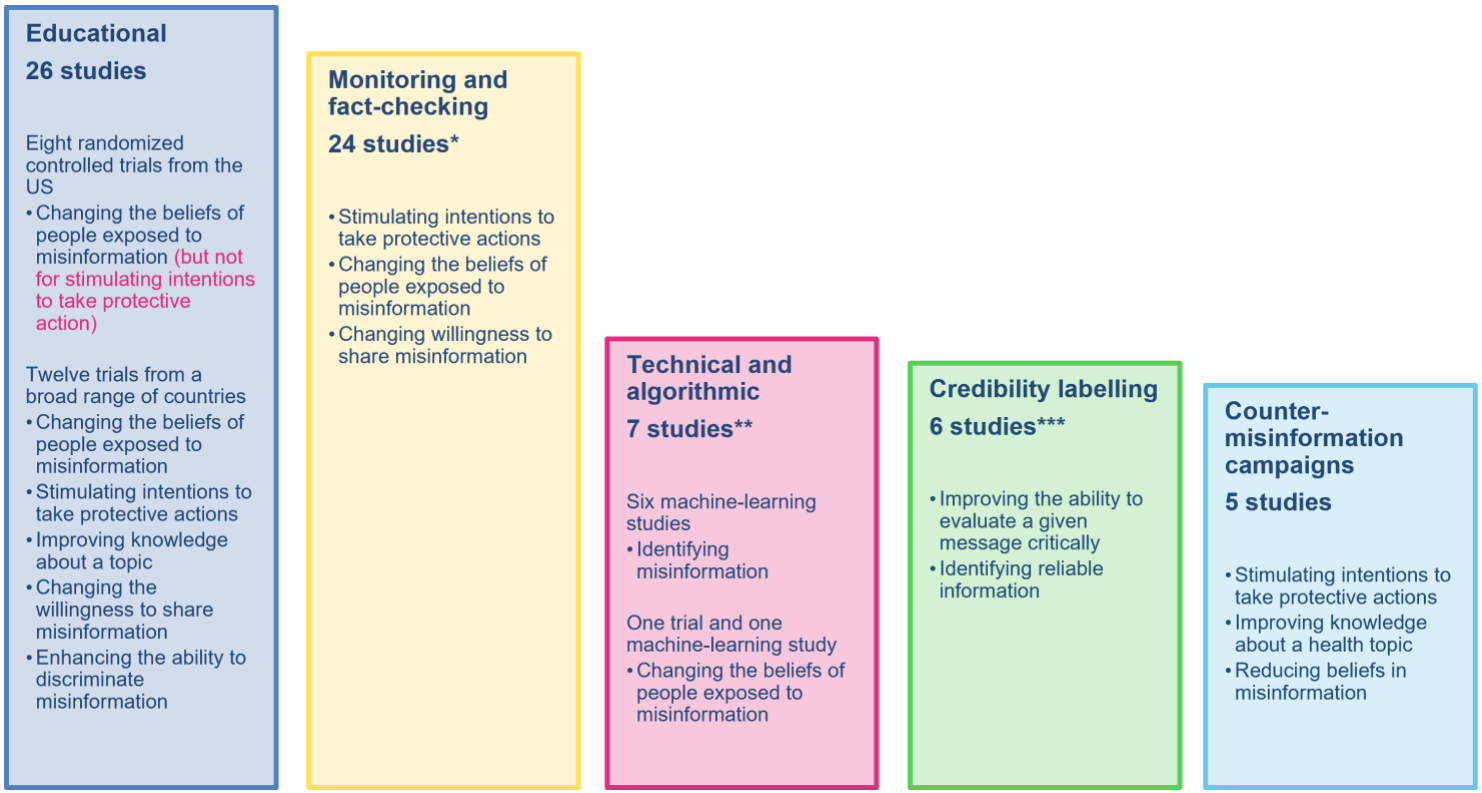
The topics addressed (some studies addressed more than one topic) were COVID-19 (n=31),(37; 38; 43-45; 56; 58; 61-74; 76; 78-80; 82; 84-86; 88) vaccination (excluding for COVID-19) (n=14),(35; 41; 47; 50; 54; 57; 59; 61; 62; 69; 74; 76; 82; 89) human papilloma virus (n=4),(35; 36; 42; 77) and other health topics (n=18).(21; 32-34; 39; 40; 46; 48; 49; 51-53; 55; 60; 75; 81; 83; 87)

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. No evidence was 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 full list of included studies in **Table 2,** a detailed description of the principal findings for each response/strategy in **Table 3** and a summary of behaviour-change techniques by established or implicit behavioural science theories tested in **Tables 4 and 5**. The findings are presented in two categories: those that we found to be effective in reducing misinformation, and those for which we found limited or no evidence of reducing misinformation. Studies excluded at the final stages of review listed in **Appendix 2** and a detailed summary of findings by study included, and by type of response/strategy, are provided in **Appendix 3 and 4**. In addition, a summary of behaviour change techniques by established or implicit behavioural science theories tested is provided in **Appendix 5** and article-level behavioural science theories/approaches and behaviour change techniques reported are provided in **Appendix 6**.

Figure 1a: 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 1b: 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,(33; 36; 37; 39; 40; 43; 45; 48; 50; 52-58; 62-66; 70; 75; 76; 81; 85) 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 risk of bias was assessed with the RoB2 tool in 20 studies, compromising 41 experiments (high n=19, some concerns n=12, low n=10), with the ROBINS-I tool in five studies, compromising seven outcomes (serious n=5, low n=2), and not assessed in one study. 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,(39; 43; 50; 53; 57) (RoB low n=1, some concerns n=3, high n=1) but not effective for stimulating intentions to take protective actions when compared to not providing education (RoB low n=1, high n=2).(39; 55; 56) 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.(55) Additionally, one study reported that news literacy messages can alter misinformation perceptions and beliefs, but not with a single message.(53) 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.(48)

The other eleven experimental randomized studies were conducted in different countries (e.g., Australia, China, France, Germany, Hong Kong, Italy, Mexico, Sierra Leone, 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,(36; 52; 62; 66) (RoB low n=1, high n=3) changing the beliefs of people exposed to misinformation (RoB low n=1, some concerns n=3, high n=1),(40; 45; 52; 66; 70) improving knowledge about a topic (RoB low=2, high n=1),(36; 45; 58) changing the willingness to share misinformation (RoB some concerns n=2),(65; 70) and enhancing the ability to discriminate misinformation (RoB high).(37) Additionally, one study in Australia found no evidence that repeating myths increased agreement with myths compared with other debunking strategies (RoB high).(54) 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 (RoB high).(64)

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 (RoB serious),(76) and for improving knowledge about a topic (RoB serious).(33) The other two studies found educational interventions not to be effective for increasing belief accuracy (RoB low n=1, serious n=1).(43; 81)

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 (RoB serious).(85) One machine learning study in the U.S. found that providing factual information on Twitter leads to a decrease in misinformation.(63)

Monitoring and fact-checking

We identified 24 studies,(32; 34; 35; 39; 40; 42; 47; 50-52; 54; 57; 58; 62; 67-69; 71; 72; 82; 83; 86; 88; 90) 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). The risk of bias was assessed with the RoB2 tool in 19 studies, compromising 44 experiments (high n=27, some concerns n=12, low n=5), with the ROBINS-I tool in one study (serious), and with CASP tool in four studies (moderate n=4). 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, twelve 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 (RoB some concerns n=2, high n=2),(34; 47; 51; 68) and that those strategies might change the beliefs of people exposed to misinformation (some concerns=2, high=2),(35; 39; 68; 71) as well as their willingness to share misinformation (RoB high).(72) Studies also reported that the positive effect of monitoring and fact-checking is stronger if the source of the fact-checking is provided (RoB high),(34) that the format of the correction does not make a considerable difference in effectiveness (RoB high),(57) and that humorous corrections might produce more attention to the misinformation text than non-humorous corrections, but not improve credibility of the correction (RoB high n=2).(35; 42) 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 (RoB low n=1, high n=1).(57; 90)

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 (RoB low n=1, high n=1),(52; 62) improving knowledge about a health topic (RoB low),(52) and reducing beliefs in misinformation (RoB some concerns n=1, high n=2).(32; 40; 54) 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%) (RoB high).(58)

The three implementation research studies described programs for monitoring and fact-checking in the U.S.,(69) Ghana,(82) and Italy.(83) The effectiveness of those programs was not evaluated; however, they reported that the targeted population was reached in each country (RoB moderate in all three).

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 (RoB moderate).(86) 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".(86)

Technical and algorithmic

We identified seven studies,(59; 61; 63; 66; 77-79) 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 risk of bias was assessed with the RoB2 tool in one study (high), with the CASP tool in one study (moderate), and not assessed in five studies. 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.(59; 61; 63; 78; 79) 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 (RoB moderate).(79) This chatbot responded accurately to open-ended, COVID-19 related questions, achieving an overall accuracy of 0.838 [95% CI: 0.826-0.851].(79) Two studies reported effectiveness in the identification and classification of HPV vaccine misinformation on Reddit (RoB not assessed).(59; 77) 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 (RoB not assessed).(61)

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 (RoB high),(66) and that providing factual information on Twitter leads to a decrease in misinformation (i.e., suppression) with a time lag (RoB not assessed).(63)

Credibility labelling

We identified six studies,(38; 41; 46; 59-61) 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 risk of bias was assessed with the RoB2 tool in four studies, compromising 11 experiments (high n=6, some concerns n=5), and not assessed in two studies. 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 (RoB some concerns),(38) or accuracy in identifying misinformation (RoB high n=2).(46; 60) 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 (RoB some concerns).(41)

Two studies of machine learning–based approaches found those strategies successful for classifying reliable information compared to unreliable information (RoB not assessed).(59; 61) 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.(61)

Counter-misinformation campaigns

We identified five studies,(44; 69; 80; 84; 87) which were conducted in Canada (n=1), Guatemala (n=1), U.S. (n=1), Zimbabwe (n=1), and Korea (n=1). The risk of bias was assessed with the RoB2 tool in one study (some concerns), with the ROBINS-I tool in two studies (moderate n=1, serious n=1), and with CASP tool in two studies (moderate n=2). 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 (RoB moderate),(80) improving knowledge about a health topic (RoB some concerns n=1, serious n=1),(44; 84) and reducing beliefs in misinformation (RoB serious).(84) 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 (RoB moderate).(80) 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) (Rob Serious).(84) 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 (RoB moderate).(69) 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 (RoB some concerns).(44)

Strategies found to have limited or no evidence for reducing misinformation

Narrative

We identified two experimental randomized studies,(73; 74) which were conducted in China and the U.S., and both addressing COVID-19. The risk of bias was assessed with the ROBINS-I tool in both studies (low 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.(73) 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).(74)

Legislative and other policy

We identified two studies,(21; 89) one conducted in Hong Kong and the other not linked to a geographic region. The risk of bias was assessed with the ROBINS-I tool in both studies (moderate n=1, serious n=1). 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 (RoB serious).(21) 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 (RoB moderate).(89)

Economic

We identified two experimental randomized studies,(21; 46) which were conducted in Hong Kong and the U.K. and addressed different health topics. The risk of bias was assessed with the RoB2 tool in one study (high), and with ROBINS-I tool in the other study (serious).

Overall, both studies found that financial incentives might not have a beneficial effect in reducing the willingness to share misinformation.(21; 46) 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 (RoB serious).(21) 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.(21) 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 (RoB high).(46)

Curatorial

No evidence identified.

Investigative

No evidence identified.

Behavioural science-informed change techniques for addressing misinformation

Out of the 59 articles in the main review, we identified 37 as behavioural science related.(21; 32; 34-39; 41-43; 46-54; 56-58; 60; 62; 64-68; 70-76) Drawing upon behavioural science taxonomies and frameworks, we sought to identify which behaviour change techniques (BCTs) were reported in intervention conditions, (91) which factors were the intended barriers/enablers targeted for change, (92; 93) and which specific theories were drawn upon to inform the interventions (see Tables 4 and 5 for detailed findings). The risk of bias was assessed with the RoB2 tool in 32 studies, compromising 76 experiments (low n=12, some concerns n=24, high n=40), and with ROBINS-I tool in five studies, compromising seven outcomes (low n=4, serious n=3).

Settings of intervention

As reported above, most studies to date have been conducted in an online survey-based environment, with a few conducted within a smartphone app, or an in-person physical space.

Outcomes and processes measured

A wide variety of factors was measured across studies, with relatively little consistency among them. Despite misinformation affecting what people ultimately decide and do, actual behaviour was seldom measured across studies.

Types of behavioural processes and outcomes targeted (mechanisms of action)

Drawing upon the Capability-Opportunity-Motivation model within the Behaviour Change Wheel and the Theoretical Domains Framework, we sought to clarify which behavioural factors the interventions seem to be targeting. Within *Capability* factors, many studies unsurprisingly targeted knowledge (n=31) and skills (n=13), while only two focused on memory/attention/decision processes. *Opportunity*-wise, many studies targeted social opportunity via social influences (n=21) while comparatively fewer focused on addressing physical opportunity in terms of the environmental context and

resources (n=2); the online nature of these evaluations likely have a role in the relative under-representation of those factors. In terms of targeting *Motivation*, many studies focused on the reflective side of motivation, namely targeting intentions (n=17), beliefs about consequences (n=11), and beliefs about capabilities (n=1). Some studies also targeted the more automatic side of motivation, namely addressing emotion (n=12), and a few also targeted reinforcement (n=3).

Behaviour-change techniques used

Drawing upon the BCT Taxonomy version 1, we identified the 'active ingredients' reported in intervention conditions within health misinformation-targeting interventions. We were able to identify 16 of 93 possible BCTs reported in at least one study. Here, we report the BCTs using labels from the BCT taxonomy v1. The most common BCTs reported included Credible source (n=9), Prompts/cues (n=8), Pros and cons (n=7), Information about health consequences (n=7), Reduce negative emotions (n=7), Future punishment (n=5), Instruction on how to perform the behaviour (n=4), Salience of consequences (n=3), and Reward (outcome) (n=2). A number of techniques were represented in only one study: Feedback on outcome(s) of behaviour, Social comparison, Incentive (outcome), Demonstration of the behaviour, Social support (practical), Information about others' approval, and Material incentive (behaviour). Findings suggest that most studies to date gravitate to a familiar and limited set of techniques, indicating substantial room for further consideration of the broader set of possible strategies that could be leveraged, especially given the breadth of barriers/enablers targeted for change.

We also categorized studies by whether they drew upon existing theories from behavioural science and/or implicit theories/approaches from broader literature and sought to describe the BCTs and targeted factors within each of those types of theories/approaches.

Inoculation theory was drawn upon in seven studies. Inoculation theory proposes that presenting messages containing weakened arguments against an adaptive attitude (e.g., vaccines are useful) early and prior to exposure to misinformation, could enable individuals to develop resistance against stronger misinformation efforts that they may encounter in the future. Typically, a warning that misinformation exposure is presented, then weakened misinformation is presented, and information is provided to discredit the false information. Overall, identified studies suggests that inoculation messaging is effective for influencing behavioural intentions, attitude, reducing reactance to credible information, and recognizing true information. While inoculation theory could in principle involve operationalizing a number of BCTs, we found that most commonly the BCT Pros and cons was used in intervention conditions (5/7 articles). Familiarity/backfire effects were tested in three studies to evaluate whether exposure to misinformation can sabotage a refutation by making the misinformation more accessible in memory. Overall, to date such backfire effects were not supported in included studies, adding indirect support for inoculation theory-informed approaches. No relevant BCTs from the taxonomy could be coded to describe such interventions.

A category of behavioural science theories known as social cognition models (e.g., theory of planned behaviour; health belief model) was drawn upon in three studies. These models suggest that the path to behaviour and behaviour change operates through changing specific types of beliefs (including attitudes, social beliefs, and/or capability beliefs) and behavioural intentions. Overall, the results in these studies show change in attitude, behavioural intentions, and recognition of true/false information. While multiple BCTs could be operationalized to address these types of beliefs in principle, a smaller subset were actually operationalised in included studies, with the BCT Information about health consequences most used (3/3 studies).

Source credibility effects were tested in nine studies, based on the premise that perceived credibility of reliable sources itself should reduce misinformation effects or enhance true information effects. Findings from studies suggest addressing source credibility can affect certain behaviours (e.g., encourage vaccine search behaviour online) as well as perceptions of severity of not following recommendations, perceptions of sources in a favourable direction, and recognition of true/false information. The BCT Credible source was used in interventions conditions in 9/9 articles.

Accuracy nudges were tested in three studies. An accuracy nudge involve asking or reminding people to be more critical about the information they are exposed to prior to being exposed to any information. These nudges appear to effectively encourage online sharing of accurate information, and they help message recipients distinguish between true and false content. The BCT Prompts/cues was used in intervention conditions in 3/3 articles.

The provision of a protective warning message that people will be exposed to misinformation was tested in 4 articles. Overall, this strategy seems to be ineffective at helping people discern both true and false information. Additional approaches that build upon this technique (e.g., inoculation messaging; see above) may therefore be necessary. The BCT Prompts/cues was used in intervention conditions in 3/3 articles.

Conclusions

The evidence suggests that educational strategies, technical and algorithmic approaches aimed at identifying 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 evaluating strategies for debunking misinformation found that fact-checking is more effective when the source is provided, that the format of a correction does not significantly impact effectiveness, and that it may be safe to repeat misinformation when correcting it, even when the audience is unfamiliar with the misinformation. We found limited evidence regarding narrative, economic, legislative, and other policy strategies. The evidence identified pointed to negligible beneficial effects related to changing people's willingness 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 on narrative, economic, legislative, and other policy strategies, as well as the lack of evidence on curatorial and investigative strategies, these are essential areas for further primary studies to be conducted to evaluate their effectiveness.

The behavioural science analysis indicates that interventions based on inoculation theory (i.e., provision of a misinformation warning, weakening misinformation, and then providing information to discredit it) are effective in influencing behavioural intentions and attitudes, reducing reactance to credible information, and recognizing true information. Furthermore, the use of accuracy nudges (i.e., reminding people to be more critical of information before information exposure) seems to encourage online sharing behaviour of accurate information effectively, helping message recipients discern between true and false content. Finally, the provision of a protective warning message that one will or may be exposed to misinformation appears to be an ineffective strategy for helping people discern both true and false information.

The behavioural science analysis also found corroborating support for some results pertaining to objective 1 above, while highlighting behaviour change techniques to enhance the actionability and reproducibility of these strategies. First, leveraging source credibility may be an effective strategy for influencing certain behaviours (e.g., encouraging vaccine search behaviour online) or other outcomes such as the recognition of true/false information. Overall, BCTs related to the comparison of outcomes (e.g., use of source credibility; provision of strong pros and weak cons, as recommended by inoculation theory) appear to be effective strategies. Additionally, changing beliefs about a behaviour (e.g., through education) may indeed alter attitudes toward the behaviour, behavioural intentions, and recognition of true/false information; however, the path to changing behaviour with such approaches remains unclear. Finally, familiarity/backfire effects (i.e., referring to any misinformation can sabotage a refutation because it makes the misinformation more accessible in memory) or the repeating of misinformation does not seem to affect misinformation countering efforts negatively.

The behavioural science analysis also underscored the relative lack of measurement of actual misinformation-related behaviours across studies; a move to more consistent measurement of beliefs, perceptions, and especially behaviours is a priority for building a cumulative evidence base. As the real-world ramifications of decisions and actions guided by

misinformation become increasingly documented, more research should be done in real-world settings to build upon the online and hypothetical nature of many studies to date. Given only a small subset of well-established BCTs have been reported to date in this literature, there is considerable opportunity to draw from a broader set of behavioural science perspectives and their accompanying BCTs to continue to evaluate innovative ways of addressing misinformation.

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 (timing to be determined):

- 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
- 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	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Abascal 2022 (80)	Guatemala	Before and after study	Counter-misinformation campaigns	Linguistically and culturally tailored social media ad campaign	No	COVID-19	Yes	ROBINS-I: Moderate
Altay 2023 (66)	France	Behavioural research (experimental randomized study)	Technical and algorithmic Educational	Chatbot	Yes	COVID-19	No	RoB2: High for intentions High for attitude
Au 2021 (21)	Hong Kong	Behavioural research (experimental randomized study)	Economic Legislative and other	Financial incentives and legislation	Yes	Different health topics	Yes	ROBINS-I: Serious for likelihood to share
Berlotti 2023 (64)	Italy	Behavioural research (experimental randomized study)	Educational	Prebunking-counterfactual	Yes	COVID-19	No	RoB2: High for plausibility High for veridicality
Blomberg 2023 (67)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction	Yes	Vitamin C COVID-19	No	RoB2: High for credibility High for hope High for information processing
Bowles 2020 (44)	Zimbabwe	Experimental randomized study	Counter-misinformation campaigns	Dissemination of messages aimed at targeting misinformation	No	COVID-19	Yes	RoB2: Some concerns
Chao 2021 (73)	China	Behavioural research (experimental randomized study)	Narrative	Debunker identity	Yes	COVID-19	No	ROBINS-I: Low for sharing of debunking information
Du 2021 (77)	Online (Reddit)	Machine Learning–Based Approaches	Technical and algorithmic	Machine learning based methods	No	HPV	No	Not assessed
Duarte 2022 (81)	Brazil	Before and after study	Educational	Intervention to increase literacy	No	Coconut oil intake	Yes	ROBINS-I: Serious
Ecker 2020 (90)	Online	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction and backfire effect Fact-checking	Yes	HIV among other issues	No	RoB2: Low for indirect believability Low for direct believability
Featherstone 2020 (47)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Refutational messages	Yes	Vaccines	No	RoB2: Some concerns for pro-vaccination attitude

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Folkvord 2022 (38)	The Netherlands	Behavioural research (experimental randomized study)	Credibility labelling	Protective message	Yes	COVID-19	No	RoB2: Some concerns for critical evaluation Some concerns for accuracy
Freeze 2021 (60)	US	Behavioural research (experimental randomized study)	Credibility labelling	Warnings	Yes	Affordable Care Act	No	RoB2: High for memory score High for misinformation score High for uncertainty score
Gavin 2022 (65)	Kyrgyzstan, India, and the U.S.	Behavioural research (experimental randomized study)	Educational	Accuracy of nudge intervention	Yes	COVID-19	Yes	RoB2: Low for likelihood to share true and false information about COVID-19 or COVID-19 vaccines
Gesser-Edelsburg 2018 (32)	Israel	Behavioural research (mixed methods including an experimental randomized design and a descriptive qualitative design)	Monitoring and fact-checking	Information correction	Yes	Measles	No	RoB2: High for reliability High for satisfaction High for intentions of sending child to kindergarten, getting child vaccinated
Gu 2022 (89)	Online	Interrupted time series	Legislative and other policy	Facebook policy (2019) on user endorsements of vaccine content on its platform	No	Vaccines	No	ROBINS-I: Moderate
Hayawi 2022 (61)	Online	Machine Learning–Based Approaches	Technical and algorithmic Credibility labelling	Machine learning detection framework	No	COVID-19 vaccination	No	Not assessed
Jiang 2022 (62)	Hong Kong	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Inoculation	Yes	COVID-19 vaccination	No	RoB2: High for vaccine attitude High for vaccine intentions
Kandasamy 2022 (84)	Canada	Cross-sectional and one-group pretest-post-test design	Counter-misinformation campaigns	Public health programme to mobilise and empower (campaign)	No	COVID-19	No	ROBINS-I: Serious
Khan 2021 (78)	Online	Machine Learning–Based Approaches	Technical and algorithmic	Algorithm to classify misinformation posts	No	COVID-19	No	Not assessed

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Kim 2021 (42)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Message attention and credibility	Yes	HPV	No	RoB2: High for HPV misperceptions High for attention High for credibility
Kim 2022 (36)	Online	Behavioural research (experimental randomized study)	Educational	Message-framing tactics	Yes	Human papillomavirus (HPV)	No	RoB2: Low for HPV vaccination attitude High for HPV vaccination intentions Low for message perceptions
Kirkpatrick 2021 (48)	US	Behavioural research (experimental randomized study)	Educational	Prospect Theory, Loss-Framing, and Perceived Severity (Youtube)	Yes	-Measles, Mumps, Rubella (MMR) -COVID-19	No	RoB2: Low for intention to share information Low for perceived severity
Lohiniva 2022 (82)	Ghana	Implementation research	Monitoring and fact-checking	The infodemic management system	No	COVID-19 vaccination	No	CASP: Moderate
Ma 2023 (37)	China	Behavioural research (experimental randomized study)	Educational	Inoculation theory	Yes	COVID-19	No	RoB2: High for misinformation recognition Some concerns for persuasive intent Some concerns for persuasion knowledge and perceived threat
MacFarlane 2021 (58)	Australia	Behavioural research (experimental study)	Monitoring and fact-checking Educational	Refuting	Yes	Vitamin E for COVID-19	No	RoB2: High for willingness-to-pay High for propensity to promote
Meppelink 2021 (59)	The Netherlands	Machine Learning–Based Approaches	Technical and algorithmic Credibility labelling	Supervised machine learning (SML) to classify health-related webpages as 'reliable' or 'unreliable'	No	-Vaccination in kids -HPV	No	Not assessed
Moore 2016 (33)	US	Before and after study	Educational	Conferences	No	Escherichia coli O157:H7 in cattle	No	ROBINS-I: Serious

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Mourali 2022 (68)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction and debunking	Yes	COVID-19 (masking)	No	RoB2: High for disposition toward masking (attitude toward masking plus intentions toward masking) High for sharing social media thread Some concerns for perceived objectivity of truth and perceived argument strength
Panizza 2022 (46)	UK	Behavioural research (experimental randomized study)	Credibility labelling	Pop-ups meant to advise participants to fact-check and other intervention based on monetary incentives	Yes	Climate change Eating chocolate Vaccines for COVID-19	No	RoB2: High for accuracy-correct guessing and accuracy score High for use of fact-checking strategies High for response times and search behaviour
Pennycook 2020 (72)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Nudging	Yes	COVID-19	No	RoB2: High for social media sharing intentions High for discerning between true and false content
Ramirez 2022 (74)	US	Behavioural research (experimental pilot study)	Narrative	Psychological inoculation	Yes	COVID-19 vaccination	No	ROBINS-I: Low for vaccine search behaviour
Roozenbeek 2021 (56)	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	Yes	COVID-19	No	RoB2: Low for intentions to share news headlines
Silesky 2023 (69)	US	Implementation research	Counter-misinformation campaigns Monitoring and fact-checking	Media monitoring findings for developing campaigns	No	COVID-19 vaccination	No	CASP: Moderate

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Song 2022 (70)	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	Yes	COVID-19	No	RoB2: Some concerns for COVID-19 perceptions/misperceptions) Some concerns for messaging believability) Some concerns for user engagement intentions
Steffens 2021 (54)	Australia	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Debunking strategies	Yes	Vaccines	No	RoB2: High for intention to vaccinate children High for change in agreement with vaccination myths High for vaccine confidence
Stekelenburg 2021 (43)	US	Behavioural research (quasi-experimental study)	Educational	Intervention aimed at increasing belief accuracy	Yes	COVID-19	No	RoB2: Low for belief accuracy High for coronavirus-related behaviour
Sun 2021 (71)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Correction	Yes	COVID-19	Yes	RoB2: Some concerns for treating appraisals (susceptibility and severity) of the influence of the misinformation on others Some concerns for anticipated guilt/anger Some concerns for intentions to correct misinformation
Swire-Thompson 2021 (57)	US	Behavioural research (experimental study)	Monitoring and fact-checking Educational	Correction	Yes	Vaccines and climate change	No	RoB2: High for climate perceptions High for inference questions
Talabi 2022 (76)	Nigeria	Behavioural research (quasi-experimental study)	Educational	Counselling	Yes	COVID-19 vaccination	No	ROBINS-I: Serious for vaccine perceptions Serious for vaccine intentions
Thacker 2020 (52)	US, Australia, Canada	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Refutational messages	Yes	Genetically modified food	No	RoB2: Low for knowledge Low for attitude Low for emotions

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Trevors 2020 (50)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Positive and negative emotional text content in refutational texts	Yes	Vaccines	No	RoB2: Some concerns for knowledge revision Some concerns for reading times
Tseng 2021 (55)	US	Randomized controlled trial	Educational	Cultivating a critical awareness of flawed scientific claims	No	Science	No	RoB2: High
Tully 2020 (53)	US	Behavioural research (experimental randomized study)	Educational	News literacy	Yes	- Genetically modified food -Seasonal flu vaccine	No	RoB2: Some concerns for information credibility Some concerns for news literacy beliefs Some concerns for misperceptions
van der Meer 2020 (51)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Corrective information type and source (narrative, educational)	Yes	Hypothetical infectious disease outbreak	No	RoB2: Some concerns for intentions to take protective action Some concerns for crisis perceived threat Some concerns for crisis emotions
Vandormael 2021 (45)	U.S., Mexico, the U.K., Germany, and Spain	Randomized controlled trial	Educational	Video for prevention	No	COVID-19	No	RoB2: Some concerns
Verduci 2021 (83)	Italy	Implementation research	Monitoring and fact-checking	Chatbot Nutripedia	No	Nutrition during Pregnancy and Early Life	No	CASP: Moderate
Vijaykumar 2021 (85)	Brazil	Cross-sectional	Educational	Social correction behaviours in WhatsApp	No	COVID-19	Yes	ROBINS-I: Serious
Vlasceanu 2023 (75)	US	Behavioural research (experimental randomized study)	Educational	Belief change	Yes	- Child's untreated wandering eye, - Abortion	No	ROBINS-I: Low for accuracy ratings Low for monetary support

Reference	Jurisdiction	Study design	Type of response/ strategy	Detail of intervention	Included in BCT analysis	Condition studied	Gender/ sex analysis	Risk of bias
Vraga 2018 (34)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Social correction (Facebook and Twitter)	Yes	Zika	No	RoB2: High for Zika misperceptions High for evaluations of correction response
Vraga 2019 (35)	Online	Behavioural research (experimental randomized study)	Monitoring and fact-checking	Inoculation and observational correction	Yes	HPV vaccination	No	RoB2: Some concerns for tweet credibility High for change in issue beliefs
Vraga 2022 (39)	US	Behavioural research (experimental randomized study)	Monitoring and fact-checking Educational	Debunking	Yes	Sunscreen and skin cancer	No	RoB2: High for intention to wear sunscreen High for sunscreen beliefs
Wang 2022 (63)	US	Machine Learning–Based Approaches	Technical and algorithmic Educational	Factual information vs misinformation (Twitter)	No	COVID-19	No	Not assessed
Winters 2021 (40)	Sierra Leone	Randomized controlled trial	Monitoring and fact-checking Educational	Debunking	No	Malaria and typhoid	No	RoB2: Some concerns
Xue 2022 (88)	Online in English	Observational study	Monitoring and fact-checking	Natural Language Processing-based Artificial Intelligence	No	COVID-19	No	ROBINS-I: Serious
Yang 2022 (86)	China	Qualitative research (content analysis)	Monitoring and fact-checking	Rumour debunking	No	COVID-19	No	CASP: Moderate
Yang 2023 (79)	Online in several languages	Natural Language Processing chatbot	Technical and algorithmic	Natural Language Processing-based Artificial Intelligence	No	COVID-19	No	CASP: Moderate
Yoon 2022 (87)	Korea	Qualitative research (content analysis)	Counter-misinformation campaigns	Using network logic of YouTube	No	Cancer	No	CASP: Moderate
Zhang 2021 (41)	US	Behavioural research (experimental randomized study)	Credibility labelling	Fact-checking labelling	Yes	Vaccines	No	RoB2: Some concerns for vaccine attitude Some concerns for perceived expertise of source Some concerns for perceived trustworthiness of source

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) (32; 34; 35; 39; 40; 42; 47; 50-52; 54; 57; 58; 62; 67-69; 71; 72; 82; 83; 86; 88; 90) 354 cases of health-related rumours (86) 212,700,000 messages (69)	Experimental randomized studies (n=19) Experiments (n=44) Risk of bias: Low n=5 Some concerns n=12 High n=27	• Overall, eleven studies conducted exclusively in the U.S. found that: ○ 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 (51) ○ 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 (34) ○ refutational messages increased pro-vaccination attitudes in comparison to misinformation messages without refutation (47) ○ all refutation texts (with or without positive or negative emotional content) improved knowledge revision (50) ○ if the key ingredients of a correction are presented, the format of correction does not make a considerable difference (57) ○ 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 (57; 90) ○ 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 (42) ○ both logic-based and humour-based corrections were effective in leading individuals to report greater agreement with expert consensus (35) ○ crafting positively framed misinformation corrections for the bolstering of message credibility within typically incongruent ideological groups is effective (67) ○ 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 (68) ○ 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 (71) ○ 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 (72) ○ 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 (39) • 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 (52)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	12,553 Facebook posts and their associated comments (n=122,362) (88)		• 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 (40) • 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 (32) • One study in Australia found no evidence that repeating myths increased agreement with myths compared with the other debunking strategies or the control (54) • 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%) (58) • 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 (62)
		Implementation research (n=3) Risk of bias: Moderate 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 (69) • 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 (82) • 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) (83)
		Qualitative research (n=1) Risk of bias: moderate	• 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 (86)
Counter-misinformation campaigns	**Conditions**: COVID-19 (n=4) Cancer (n=1) **Jurisdictions**:	Experimental randomized (n=1) Risk of bias: some concerns	• 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 (80)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	Canada (n=1), Guatemala (n=1), U.S. (n=1), Zimbabwe (n=1), Korea (n=1) **Sample:** 2,470 participants (1,411 women) (44; 69; 84) 573 videos (87)	Quasi experimental (n=1) Risk of bias: moderate	• 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) (84)
		Cross-sectional (n=1) Risk of bias: serious	• 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 (44)
		Implementation research (n=1) Risk of bias: moderate	• 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 (69)
		Qualitative research (n=1) Risk of bias: moderate	• 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 (87)
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) (38; 41; 46; 60)	Experimental randomized (n=4) Experiments (n=11) Risk of bias: Some concerns n=5 High n=6	• 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 (38) • 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 (46) • 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 (60) • 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 (41)
		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 (59)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	15,465,687 tweets (61) 468 Dutch webpages (59)	Risk of bias not assessed	• 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 (61)
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) (33; 36; 37; 39; 40; 43; 45; 48; 50; 52-58; 62-66; 70; 75; 76; 81; 85) 22,111,831 English tweets (63)	Experimental randomized (n=20) Experiments (n=41) Risk of bias: Low n=10 Some concerns n=12 High n=19	• Overall, eight studies conducted exclusively in the U.S. found that: ○ 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 (48) ○ all refutation texts (with or without positive or negative emotional content) improve knowledge revision (50) ○ news literacy messages can alter misinformation perceptions and beliefs, but not with a single message (53) ○ 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 (55) ○ 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 (56) ○ 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 (39) ○ if the key ingredients of a correction are presented, the format of the correction does not make a considerable difference (57) ○ 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 (75) • 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 (52) • 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 (45) • 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 (65) • 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 (37)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
			• 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 (40) • One study in Australia found no evidence that repeating myths increased agreement with myths compared with the other debunking strategies or the control (54) • 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%) (58) • 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 (62) • 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 (64) • 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) (66) • 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 (70) • 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 (36)
		Quasi experimental (n=4) Outcomes (n=6) Risk of bias: Low n=2	• One study in the U.S. found that an intervention (infographic) aimed at increasing belief accuracy was not effective (43) • 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 (33) • 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 (81)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
		Serious n=4	• 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 (76)
		Cros-sectional (n=1) Risk of bias: serious	• 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 (85)
		Machine Learning–Based Approaches (n=1) Risk of bias not assessed	• 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 (63)
Curatorial	No evidence found		
Narrative	**Conditions**: COVID-19 (n=2) **Jurisdictions**: China (n=1), U.S. (n=1) **Sample**: 1,196 observations (73)	Experimental randomized (n=2) Risk of bias: moderate 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 (73) • 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) (74)
Technical and algorithmic	**Conditions**: COVID-19 (n=5) VPH (n=2) **Jurisdictions**: France (n=1), the Netherlands	Machine Learning–Based Approaches (n=6) Risk of bias: Moderate n=1 Not assessed n=5	• 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] (79) • 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 (77) • 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 (78)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	(n=1), U.S. (n=1), NA (n=4) **Sample**: 701 participants (291 women) (66) 37,577,518 tweets (61; 63) 468 Dutch webpages (59)		• 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 (61) • 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 (59) • 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 (63)
		Experimental randomized (n=1) Experiments (n=2) Risk of bias: high n=2	• 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) (66)
Economic	**Conditions**: Different health topics (n=2) **Jurisdictions**: Hong Kong (n=1), the U.K. (n=1) **Sample**: 5,750 participants (3,479 women) (21; 46)	Experimental randomized (n=2) Experiments (n=4) Risk of bias: high n=4	• 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 (21) ○ 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 (21) • 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 (46)
Legislative and other policy	**Conditions**: Vaccines (n=1) Other (n=1)	Experimental randomized (n=1) Risk of bias: serious	• 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 (21)

Response/ Strategy	Conditions, jurisdictions, and sample	Study design and quality appraisal	Findings
	Jurisdictions: Hong Kong (n=1), NA (n=1) Sample: 363 participants (137 women) (21) 172 anti- and pro-vaccine Facebook pages (89)	Interrupted-time series (n=1) Risk of bias: moderate	• 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 (89)
Investigative	No evidence		

Table 4: Summary of behaviour change techniques by established behavioural science theories tested

Established behavioural science theories tested (N of studies)	Behaviour change techniques identified in intervention conditions	BeSci-consistent processes and behaviours measured across studies
Inoculation theory (n=7) Theory conceptual summary: messages containing weakened arguments against an established attitude could enable individuals to develop resistance against stronger attacks that they may encounter to that attitude in the future.(94)	2.7. Feedback on outcome(s) of behaviour 4.1. Instruction on how to perform the behaviour 5.1. Information about health consequences 6.1. Demonstration of the behaviour 6.2. Social comparison 7.1. Prompts/cues 9.1 Credible source 9.2. Pros and cons 10.8. Incentive (outcome) 10.10. Reward (outcome)	*Capability* • Misinformation recognition • Memory score (good truth discernment) • Misinformation score (bad truth discernment) • Uncertainty score • Plausibility of fake headline • Veridicality of fake headline • Persuasion knowledge • Misinformation perceptions *Opportunity* • Debunk credibility *Motivation* • Reduced anger • Reduced fear • Reactance • Perceived threat • Attitude *Intention* • Intentions *Behaviour* • Link click (search behaviour)
Elaboration likelihood model (n=4) Theory conceptual summary: dual process models of information processing and attitude/behaviour change.(95)	5.2. Salience of consequences 9.2. Pros and cons 11.2. Reduce negative emotions	*Capability* • Knowledge • Processing type (i.e., heuristic vs central processing) *Opportunity* • Credibility • Believability *Motivation* • Humourous correction on attention to stimuli • Humourous correction on HPV misperceptions • Humourous correction on misinformation • Humourous correction on correction credibility • Attitude • Emotions (e.g., surprise) • *Intention* • Intentions *Behaviour* • Engagement
Social cognition models (n=4)	3.2. Social support (practical) 5.1. Information about health consequences;	*Capability* • Message perceptions • Accuracy ratings

Established behavioural science theories tested (N of studies)	Behaviour change techniques identified in intervention conditions	BeSci-consistent processes and behaviours measured across studies
Theory conceptual summary: particular cognitions, perceptions and social influence that combine to influence motivation and action; eg Theory of Planned Behaviour; Health Belief Model; Social Cognitive Theory.(96-98)	6.1. Demonstration of the behaviour 9.2. Pros and cons;	*Opportunity* • n/a *Motivation* • Attitude *Intention* • Intentions *Behaviour* • Supporting a charity • Link click (search behaviour) • Monetary support for different charities
Influence of presumed influence model (n=1) Theory conceptual summary: people's own exposure to media content functions as the foundation of making inferences about others' exposure to the same content (presumed exposure); the presumed exposure by others promotes the presumption that the media content will influence others (presumed influence); the presumed influence on others prompts the alignment of the attitude of self to the presumed attitude of others (influence of presumed influence.(99)	5.2. Salience of consequences 10.11. Future punishment;	*Capability* • n/a *Opportunity* • n/a *Motivation* • Threat appraisals (susceptibility and severity) • Anticipated guilt/anger *Intention* • Intentions to correct misinformation *Behaviour* • n/a
Cognitive appraisal theory (n=1) Theory conceptual summary: emotions arise from subjective interpretation and evaluation of events, rather than being a direct result of stimuli.(100)	5.2. Salience of consequences 10.11. Future punishment;	*Capability* • n/a *Opportunity* • n/a *Motivation* • Threat appraisals (susceptibility and severity) • Anticipated guilt/anger *Intention* • Intentions to correct misinformation *Behaviour* • n/a
Extended parallel process model (n=1) Theory conceptual summary: perceived threat leads to defensive reactions or not depending on levels of perceived efficacy.(101)	5.2. Salience of consequences 10.11. Future punishment;	*Capability* • n/a *Opportunity* • n/a *Motivation*

Established behavioural science theories tested (N of studies)	Behaviour change techniques identified in intervention conditions	BeSci-consistent processes and behaviours measured across studies
		• Threat appraisals (susceptibility and severity) • Anticipated guilt/anger *Intention* • Intentions to correct misinformation *Behaviour* • n/a
Cognitive dissonance theory (n=1) Theory conceptual summary: an internal/personal tension is created by having two contradicting values, beliefs or behaviours, and people become motivated to reduce that tension – or dissonance – by changing one of the conflicting values, beliefs or behaviours.(102)	3.2. Social support (practical) 5.1. Information about health consequences;	*Capability* • n/a *Opportunity* • n/a *Motivation* • Vaccine negative perceptions • Vaccine positive perceptions *Intention* • Vaccine intentions *Behaviour* • n/a
Social contagion theory (n=1) Theory conceptual summary: Individuals' emotions and behaviours can be influenced by other people's words, texts, expressions, gestures, and other messages.(73)	10.11 Future punishment; 11.2 Reduce negative emotions	*Capability* • n/a *Opportunity* • Number of followers *Motivation* • n/a *Intention* • n/a *Behaviour* • Sharing of debunking information
Knowledge revision components (KReC) framework (n=1) Theory conceptual summary: KReC consists of five principles that account for knowledge revision as readers engage with a refutation text. The first two principles consist of core assumptions (encoding and passive activation), and the remaining three principles describe conditions and the mechanism of knowledge revision (coactivation, integration, and competing activation.(103)	11.2 Reduce negative emotions	*Capability* • Knowledge revision • Reading times *Opportunity* • n/a *Motivation* • n/a *Intention* • n/a *Behaviour*

Established behavioural science theories tested (N of studies)	Behaviour change techniques identified in intervention conditions	BeSci-consistent processes and behaviours measured across studies
		• n/a
Limited capacity model of mediated message processing (n=1) Theory conceptual summary: this model has five major assumptions—the first about the nature of cognition, the second about the nature of motivation, the third about the nature of media, the fourth about the nature of time, and the fifth about the nature of communication; processing messages involves three major subprocesses: encoding, storage, and retrieval.(104)	11.2. Reduce negative emotions	*Capability* • Attention to the correction image *Opportunity* • n/a *Motivation* • Humourous correction on attention to stimuli • Humourous correction on HPV misperceptions • Humourous correction on misinformation credibility • Humourous correction on correction credibility *Intention* • n/a *Behaviour* • n/a
Truth-default theory (n=1) Theory conceptual summary: the central most idea of TDT is that people default to the truth, or at least truth as they know it; generally, humans communicate honestly unless there is a reason(s) to deceive others, and humans passively accept incoming communication as honest unless they have a reason(s) to suspect otherwise.(105)	7.1. Prompts/cues	*Capability* • Accuracy of beliefs *Opportunity* • Credibility evaluation of information *Motivation* • n/a *Intention* • n/a *Behaviour* • n/a
Loss vs. gain-framed messages (Prospect Theory) (n=1) Theory conceptual summary: depending on how otherwise equivalent options are described – or framed - individuals tend to evaluate potential gains and losses differently. People tend to be averse to losses, i.e. feel losses more strongly than the equivalent gain and in particular tend to be loss averse.(106)	10.11 Future punishment; 11.2. Reduce negative emotions	*Capability* • n/a *Opportunity* • n/a *Motivation* • Perceived severity *Intention* • Intentions *Behaviour* • n/a

Table 5: Summary of behaviour change techniques by implicit behavioural science theories tested

Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
Source credibility (n=9)	4.1. Instruction on how to perform the behaviour; 5.1. Information about health consequences; 5.2. Salience of consequences 6.3. Information about others' approval; 9.1. Credible source; 11.2. Reduce negative emotions;	*Capability* • Fact beliefs • Incorrect beliefs *Opportunity* • Credibility evaluation of information • Source expertise • Source trustworthiness *Motivation* • Attitude • Perceived severity • Crisis emotions *Intention* • Intentions • Propensity to promote (share, like, or flag) *Behaviour* • Coronavirus-related behaviour • Link click (vaccine search behaviour)
Emotional messaging (n=4)	9.1 Credible source; 10.11 Future punishment 11.2. Reduce negative emotions;	*Capability* • Reliability of information • Satisfaction with information • Knowledge revision *Opportunity* • Credibility *Motivation* • Hope *Intention* • Intentions *Behaviour* • Reading times • Sharing of debunking information
Protective warning message that people will be exposed to misinformation (n=4)	7.1. Prompts/cues; 9.1 Credible source	*Capability* • Credibility evaluation of information • Accuracy of beliefs • Memory score (good truth discernment) • Misinformation score (bad truth discernment) • Uncertainty score (selected none are correct for the answers) • Plausibility of fake headline • Veridicality of fake headline *Opportunity*

Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
		• n/a *Motivation* • Reactance • Attitude *Intention* • n/a *Behaviour* • n/a
Familiarity/backfire effect (the idea that even referring to the misinformation can sabotage a refutation) (n=3)	No BCTs identified	*Capability* • Change in myth agreement • Inference questions *Opportunity* • Indirect believability • Direct believability *Motivation* • n/a *Intention* • Intentions *Behaviour* • n/a
Accuracy nudge (asking people to be more critical about the information they are exposed to) (n=3)	7.1. Prompts/cues	*Capability* • Discerning between true and false content *Opportunity* • n/a *Motivation* • n/a *Intention* • Intentions to share *Behaviour* • n/a
Simple rebuttal vs factual elaboration vs control (n=2)	5.2. Salience of consequences 9.1 Credible source;	*Capability* • n/a *Opportunity* • n/a *Motivation* • Crisis perceived threat • Crisis emotions (i.e., hope, lower confusion, anxiety, and fear)

Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
		Intention • Intentions • Propensity to promote (share, like, or flag) *Behaviour* • n/a
Humour-based messaging (n=2)	11.2. Reduce negative emotions	*Capability* • n/a *Opportunity* • n/a *Motivation* • Humourous correction on attention to stimuli • Humourous correction on HPV misperceptions • Humourous correction on misinformation credibility • Humourous correction on correction credibility *Intention* • n/a *Behaviour* • n/a
Presentation mode (e.g., visual vs. text) (n=2)	No relevant BCTs	*Capability* • Processing type (heuristic vs. central) • Elaboration *Opportunity* • Credibility • Believability *Motivation* • Hope *Intention* • Intentions *Behaviour* • Engagement
Threat of legislation to punish spread of misinformation, and financial incentives (n=2)	10.1 Material incentive (behaviour); 10.10. Reward (outcome) 10.11 future punishment;	*Capability* • Accuracy-correct guessing • Accuracy score • Use of fact-checking strategies • Response times *Opportunity* • n/a *Motivation* • n/a

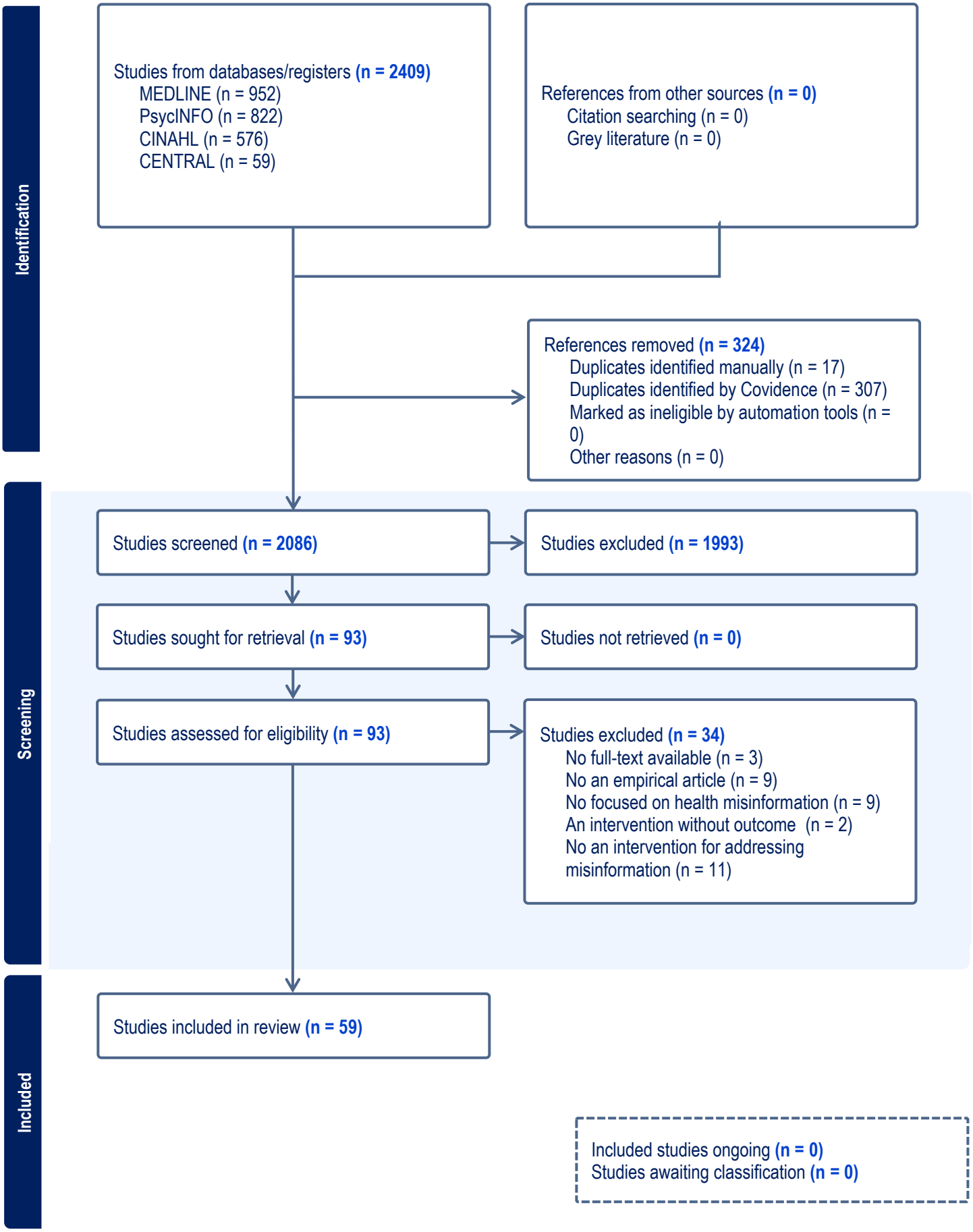
Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
		Intention • Likelihood to share (mis)information *Behaviour* • Search behaviour
News literacy boosting messages (n=2)	4.1. Instruction on how to perform the behaviour 7.1. Prompts/cues;	*Capability* • Fact beliefs • Incorrect beliefs • News literacy beliefs *Opportunity* • Information credibility *Motivation* • n/a *Intention* • Intentions *Behaviour* • n/a
News literacy boosting messages (n=2)	4.1. Instruction on how to perform the behaviour 7.1. Prompts/cues;	*Capability* • Fact beliefs • Incorrect beliefs • News literacy beliefs *Opportunity* • Information credibility *Motivation* • n/a *Intention* • Intentions *Behaviour* • n/a
Debunking in prolonged social media debates (n=1)	5.1. Information about health consequences; 9.2. Pros and cons	*Capability* • Perceived objectivity of truth • Perceived argument strength *Opportunity* • n/a *Motivation* • Attitude *Intention* • Intentions

Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
		Behaviour • Sharing social media thread
With versus without statistical evidence (n=1)	5.2. Salience of consequences	*Capability* • Elaboration *Opportunity* • Believability *Motivation* • n/a *Intention* • Intentions *Behaviour* • Engagement
Social media-based counselling intervention (n=1)	3.2. Social support (practical) 5.1. Information about health consequences;	*Capability* • n/a *Opportunity* • n/a *Motivation* • Vaccine negative perceptions • Vaccine positive perceptions *Intention* • Vaccine intentions *Behaviour* • n/a
Civic online reasoning (looking for information on other websites (lateral reading) and looking beyond the first results suggested by search engines (click restraint)(46) (n=1)	4.1. Instruction on how to perform the behaviour;	*Capability* • Belief accuracy • Using fact checking strategies • Response times *Opportunity* • n/a *Motivation* • n/a *Intention* • n/a *Behaviour* • Online search behaviour
Gamified environment with rewards and rankings (n=1)	2.7. Feedback on outcome(s) of behaviour; 6.2. Social comparison; 10.8. Incentive (outcome);	*Capability* • Misinformation recognition • Persuasion knowledge

Implicit approach/theory (n of studies)	Behaviour change techniques used in main intervention conditions	BeSci-consistent processes and behaviours measured across studies
	10.10. Reward (outcome)	*Opportunity* • n/a *Motivation* • Perceived threat *Intention* • Intentions *Behaviour* • n/a
Debunking in scrolling comments (n=1)	4.1. Instruction on how to perform the behaviour	*Capability* • Fact beliefs • Incorrect beliefs *Opportunity* • n/a *Motivation* • n/a *Intention* • Intentions *Behaviour* • n/a

Note. Numeric prefixes in BCTs reflect coding structure from the BCT Taxonomy version 1.

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, Sangngam N, et al. Interventions to mitigate vaping misinformation: A meta-analysis. *Journal of Health Communication* 2022;27(2): 84-92.
3. Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. *Bull World Health Organ* 2022;100(9): 544-561.
4. Calo R, Coward C, Spiro ES, Starbird K, West JD. How do you solve a problem like misinformation? *Sci Adv* 2021;7(50): eabn0481.
5. Vidgen B, Taylor H, Pantazi M, Anastasiou Z, Inkster B, Margetts H. Understanding vulnerability to online misinformation. Alan Turing Institute; 2021.
6. 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. *J Public Health Manag Pract* 2017;23(4): 380-387.
7. Jaiswal J, LoSchiavo C, Perlman DC. Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. *AIDS Behav* 2020;24(10): 2776-2780.
8. Nan X, Wang Y, Thier K. Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. *Social science & medicine (1982)* 2022;314(ut9, 8303205): 115398.
9. International P. The truth gap: the state of the world's girls. 2021. 2021.
10. Grimes K, Matlow A, Tholl B, Dickson G, Taylor D, Chan MK. Leaders supporting leaders: Leaders' role in building resilience and psychologically healthy workplaces during the pandemic and beyond. *Healthc Manage Forum* 2022;35(4): 213-217.
11. Challenges GCoEtAS. 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.
12. Forum WE. Global Risks Report 2024. Geneva, Switzerland: World Economic Forum; 2024.
13. Van Bavel JJ, Gadarian SK, Knowles E, Ruggeri K. Political polarization and health. *Nature Medicine* 2024;30(11): 3085-3093.
14. The L. Health in the age of disinformation. *The Lancet* 2025;405(10474): 173.
15. Development BCfS, UNESCO. Broadband Commission research report on 'freedom of expression addressing Disinformation on the Internet'. balancing act: countering Digital Disinformation while respecting freedom of expression. Broadband Commission for Sustainable Development

UNESCO; 2020.

16. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. *Annu Rev Public Health* 2020;41: 433-451.
17. Broniatowski DA, Jamison AM, Qi S, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *Am J Public Health* 2018;108(10): 1378-1384.
18. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine* 2015;33(39): 5051-6.
19. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science* 2018;359(6380): 1146-1151.
20. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. *J Med Internet Res* 2021;23(1): e17187.
21. Au CH, Ho KKW, Chiu DKW. Stopping healthcare misinformation: The effect of financial incentives and legislation. *Health Policy* 2021;125(5): 627-633.
22. 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.

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

24. 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.

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

26. 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.

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

28. 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.

29. Forum MH. Suite of living evidence syntheses about COVID-19 public health and social measures. Hamilton, Canada: McMaster Health Forum; 2023.

30. Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences* 2020;1(1): 31-42.

31. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *Int J Evid Based Healthc* 2015;13(3): 179-87.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

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

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

55. 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.

56. 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.

57. 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).

58. 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.

59. 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.

60. 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.

61. 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.

62. 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).

63. 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.
64. 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.
65. 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.
66. 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.
67. 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.
68. Mourali M, Drake C. Debunking Health Misinformation on Social Media: the Challenge of Dynamic Conversations. *Journal of Medical Internet Research* 2022;24(3): e34831.
69. 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.
70. 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.
71. 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.
72. 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.
73. 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).
74. 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.
75. 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.
76. 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.
77. 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.
78. 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.
79. 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.

80. 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.
81. 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.
82. 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.
83. 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.
84. 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.
85. 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.
86. 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.
87. 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.
88. 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.
89. 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.
90. 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).
91. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013;46(1): 81-95.
92. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005;14(1): 26-33.
93. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011;6: 42.
94. McGuire WJ. The Effectiveness of Supportive and Refutational Defenses in Immunizing and Restoring Beliefs Against Persuasion. *Sociometry* 1961;24(2): 184-197.
95. Petty RE, Briñol P. The elaboration likelihood model. Handbook of theories of social psychology, Vol 1: Sage Publications Ltd; 2012. p. 224-245.
96. Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 1991;50(2): 179-211.
97. Bandura A. Health promotion from the perspective of social cognitive theory. *Psychology & Health* 1998;13(4): 623-649.
98. Janz NK, Becker MH. The Health Belief Model: a decade later. *Health Educ Q* 1984;11(1): 1-47.
99. Cho H, Shen L, Peng L. Examining and Extending the Influence of Presumed Influence Hypothesis in Social Media. *Media Psychol* 2021;24(3): 413-435.

100. Campbell TS, Johnson JA, Zernicke KA. Cognitive Appraisal. In Gellman MD (editor). Encyclopedia of Behavioral Medicine. Cham: Springer International Publishing; 2020. p. 486-487.

101. Popova L. The extended parallel process model: illuminating the gaps in research. *Health Educ Behav* 2012;39(4): 455-73.

102. Festinger L. A theory of cognitive dissonance: Stanford University Press; 1957. xi, 291-xi, 291 p.

103. Butterfuss R, Kendeou P. KReC-MD: Knowledge revision with multiple documents. *Educational Psychology Review* 2021;33(4): 1475-1497.

104. Lang A. Using the Limited Capacity Model of Motivated Mediated Message Processing to Design Effective Cancer Communication Messages. *Journal of Communication* 2006;56(Suppl 1): S57-S80.

105. Levine TR. Truth-default theory and the psychology of lying and deception detection. *Curr Opin Psychol* 2022;47: 101380.

106. Tversky A, Kahneman D. The Framing of Decisions and the Psychology of Choice. *Science* 1981;211(4481): 453-458.

Vélez CM, Wilson MG, Lithopoulos A, Presseau J, Wu N, Patiño-Lugo DF, Cura J, Smith M, Brehaut J, Gretton J, Nicklin W, Lavis JN, Misinformation Living Evidence Synthesis Team*. Living evidence synthesis 22.2: Impact of strategies to mitigate health-related misinformation in diverse settings and populations: Living Evidence Synthesis. Hamilton: McMaster Health Forum, 25 February 2026.

This living 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, 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.