World Health Organization

European Region

Managing false information in health emergencies: an operational toolkit

2024

ABSTRACT

The proposed toolkit provides procedures for the detection and handling of false information, following a five-step process involving signal detection, verification, risk assessment, response design and outreach. A valuable resource for authorities and other stakeholders, this toolkit helps facilitate active infodemic management, promoting accurate information dissemination and informed public health decisions.

KEYWORDS

EMERGENCIES
EMERGENCY PREPAREDNESS
HEALTH COMMUNICATION
COMMUNITY PARTICIPATION
INFODEMIC
SIGNAL DETECTION

Document number: WHO/EURO:2024-8271-48043-71198

Art-direction and layout: www.freightdesign.co.uk

World Health Organization

European Region

Managing false information in health emergencies: an operational toolkit

2024


© WHO
Infodemic
Management Training
Chișinău, Republic of Moldova
6-7 March 2023
Infodemic Management Training
televic

Contents

Acknowledgements

The WHO Regional Office would like to acknowledge the many stakeholders who reviewed and shared feedback on this document, including members of the Technical Advisory Group for Risk Communication, Community Engagement and Infodemic Management (RCCE-IM) in the WHO European Region, especially Dr. Seema Yasmin, Stanford University, United States; and Robert Steiner, University of Toronto, Canada.

This document has been developed with contributions from experts from the RCCE-IM team at the WHO Regional Office for Europe. The main author is Ravi Sreenath, and Stefan Voinea and Kimberly Rambaud made substantial technical contributions to the content of this implementation tool. Cristiana Salvi supervised the process, provided guidance and made significant contributions to the technical review. Additional staff members and consultants within the WHO Regional Office for Europe who offered technical review and guidance included Altug Akin, Philippe Borremans, Ben Duncan, Olha Izhyk, and Leonardo Palumbo.

The WHO Regional Office also expresses gratitude to Dr Gerald Rockenschaub (former Regional Emergency Director, WHO Regional Office for Europe) for his support and review of this tool.

Abbreviations

CSOs	civil society organizations
HSE	Health Service Executive (Ireland)
IM	infodemic management
MMR	measles, mumps and rubella
RCCE-IM	risk communication, community engagement and infodemic management

This toolkit provides operational support to stakeholders engaged in infodemic management in the WHO European Region, in the context of health emergencies.

What is the operational toolkit for detecting and addressing false information in health emergencies?

Introduction

As a result of the COVID-19 pandemic, there has been an alarming acceleration in the creation and distribution of "information disorders" such as misinformation, disinformation and malinformation, across both digital and physical spaces *(1)*. This scenario was aggravated by the current information environment where factual information is sometimes disregarded, and conspiracy theories find fertile ground. These conditions have consistently eroded trust in authorities, undermined advances in public health, complicated health decision-making and put lives at risk.

WHO defines this information disorder, or infodemic – as an overabundance of information, including false or misleading information, in digital and physical environments during an emergency. Infodemic management (IM) is the systematic use of risk- and evidence-based analysis and approaches to manage harmful information and reduce its impact on health behaviours during health emergencies. While offline information tracking and data collection is a valuable source of information, this toolkit is mainly focused on online data monitoring.

IM is a crucial part of an integrated emergency public health intervention that includes Risk Communication, Community Engagement and Infodemic Management (RCCE-IM). During emergency response, the primary role of IM is to detect, prevent and address various forms of health information disorders, contributing to an improved health information ecosystem.

By effectively managing an infodemic, accurate information dissemination can be promoted and communities can be aided to make well-informed decisions for health protection. This in turn contributes to establishing structures, systems and skills for sustained IM. During the COVID-19 pandemic in the European Region, there has been a marked increase in requests by Member States to the WHO Regional Office for Europe for IM capacity-building support. This surge in demand has been accompanied by calls for assistance in implementing strategies to mitigate and manage the spread of false information.

Purpose of the toolkit

The purpose of this toolkit is to provide operational support to national authorities, partners, civil society organizations (CSOs) and other stakeholders engaged in IM in the WHO European Region, in the context of health emergency preparedness and response.

As part of comprehensive IM, this toolkit focuses on actionable tips and steps that support the detection and evaluation of false health information to inform response. This in turn will prevent and mitigate the impact of harmful information on public health.

The toolkit complements IM initiatives to translate specific country needs into action, in two major ways:

- **Operationalizing IM:** The toolkit is an extension of *Advancing infodemic management in risk communication and community engagement in the WHO European Region: implementation guidance (2)* and provides actionable tools in the execution of IM tasks as well as examples and case studies. This can help reduce the risk of errors, improving overall quality and efficiency.

- **Reinforcing learning:** The toolkit can serve as a reinforcement tool for Member States and other relevant stakeholders who take part in IM workshops or trainings. It serves as a reference guide with key concepts and processes which enable participants to reinforce and apply IM knowledge in their day-to-day work.

Structure of the toolkit

The toolkit describes a five-phase process (Fig. 1) to deal with false information signals during a health emergency (Fig. 1). It can also be applied to managing other components of an infodemic such as public questions, concerns and information voids (where people seek accurate health information but cannot find it). The five phases: detection, verification, risk assessment, response design and outreach are briefly summarized below and then expanded on throughout the toolkit, with sections on each respective phase containing short, actionable guidance and, as needed, relevant checklists, algorithms and infographics.

By employing the toolkit, valuable knowledge can be acquired and consolidated into an infodemic insights report which can serve as a pivotal resource for guiding RCCE-IM interventions.

Phase 1: Signal detection

The first phase is monitoring and signal detection. This phase involves actively monitoring the information ecosystem to identify potential signals of false information or rumours related to public health. This requires the use of various online and offline tools and methods, such as social listening, media monitoring, community engagement and expert networks to identify signals in real-time. The outcome of this phase is a set of identified signals that need to be verified and further assessed in the subsequent phases of the process.

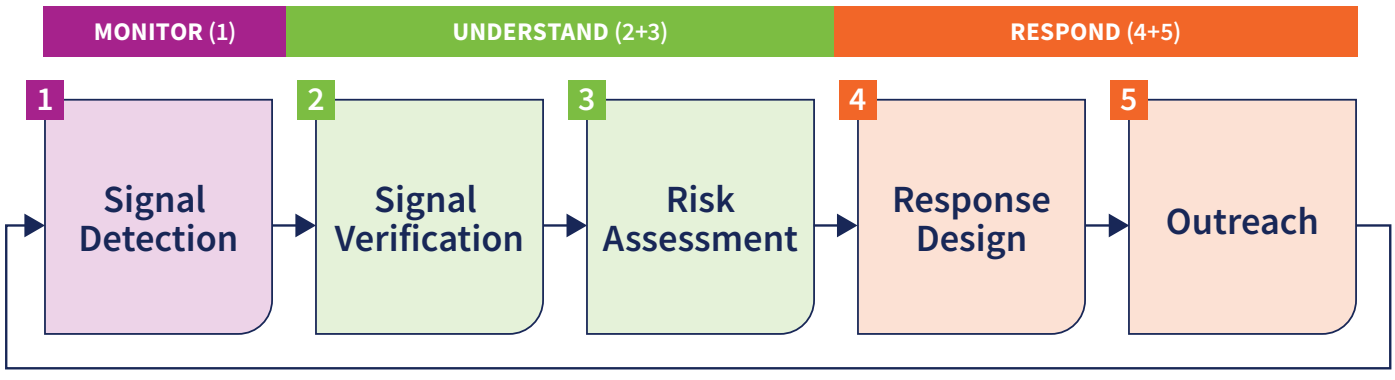
Phase 2: Signal verification

The signal verification phase involves determining whether a signal is true or false, and identifying the source of information. In this phase, information is gathered from various sources to validate or invalidate the signal. This can include fact-checking the information, analysing the credibility of the source, and assessing the accuracy and consistency of the information across multiple sources. The outcome of this phase is a determination of whether the signal is true, false or partially true, and the confidence level of that determination.

Phase 3: Risk assessment

The risk assessment phase involves performing an integrated analysis and evaluating its potential impact on public health, as well as assessing the level of risk associated with it. The outcome of this phase is a determination of the potential consequences of the false information signal on public health and to guide the development of action, including no action or a response strategy based on the following criteria: 1) source credibility; 2) spread of the false information; and 3) public health consequences.

Fig. 1. The key phases to detect and address false information signals during a health emergency



Phase 4: Response design

In case a response is needed, the response design phase involves developing effective interventions to counter false information and address the identified risks. It includes segmenting audiences; crafting accurate, clear and relevant messages; selecting appropriate communication channels; and determining the best timing and frequency for message targeting. This section also provides tactical guidance on designing "debunks" and "prebunks" as response interventions to counter false information. The outcome of this phase is the design of an effective response.

Phase 5: Outreach

In this final phase of the process, the response plan is implemented and key messages are targeted to segmented audiences, encouraging them to adopt the behaviour that supports the intended public health outcomes. The outreach phase can also link back to the first phase of signal detection through a feedback loop. The outcome of this phase is enabling people to take informed decisions to protect their health, through refuting false information and promoting accurate information and advice.

Operational challenges to detecting and addressing false information

When setting up a system for detecting harmful health information, a variety of operational factors come into play. Recognizing and understanding these aspects, especially in the context of health emergencies, is crucial for designing effective strategies to combat false information and mitigate its negative impacts. The operational challenges encountered should not be a reason for inaction, but rather a catalyst to innovate and implement robust, sustainable interventions.

1. **Addressing false information is considerably more resource-intensive than producing it.** The process often mirrors the act of trying to extinguish a fire with a single water droplet, particularly for resource-strapped public health authorities combating a steady stream of harmful health information. Advancements in artificial intelligence models such as ChatGPT have further simplified the creation of false information, driving its cost almost to zero. In contrast, a social listening system demands considerable resources, expertise and effort to monitor, analyse and respond effectively. To address this disparity, leveraging partnerships with technology companies, investing in automated detection systems and fostering international collaboration can create a more scalable and resource-efficient approach to combating false health information.

2. **Social listening data lacks offline context.** Online listening primarily captures the views and opinions of a specific demographic, leaving out substantial sections of the population. From an operational perspective, online data needs to be complemented with offline research using diverse sources, such as offline surveys and community engagement, for an inclusive and accurate understanding of signals.

3. **Human analysts are needed to make sense of automatically generated data.** Social media posts often lack context and are riddled with language complexities, making automated data interpretation impossible. It's crucial to employ human analysts to discern the true meaning of shared information, evaluate its accuracy and assess associated health risks *(3)*.

4. **Response strategies may not reach the same audience as the initial false information.** Due to the "filter bubble" effect *(4)*, outreach activities often fail to target the same audience that was initially exposed to misinformation. Algorithmic personalization on social media platforms tailors content to user preferences and viewpoints, thus creating a "bubble" that primarily exposes them to information reinforcing their existing

Ethical considerations

Signal detection is critical for IM and detecting and addressing false information during response, yet this process raises ethical considerations that must be considered and addressed. Here are some key challenges and considerations regarding social listening in the context of RCCE-IM:

1. **Privacy and data protection.** Privacy considerations are paramount when engaging in social listening to detect health misinformation. Focusing on themes and trends rather than specific individuals helps minimize the risk to individual privacy. Analysing anonymized and aggregated data can still provide valuable insights without exposing personally identifiable information *(5)*.

2. **Transparency and potential for overreach.** Setting up monitoring systems needs to be done with transparency about the purpose and scope of the monitoring. It is important to be transparent about the data being collected, how it will be used and who will have access to it. There is a risk of overreach if the data collected is used for other purposes beyond its intended scope.

3. **Algorithmic misrepresentation of intent.** Artificial intelligence-assisted social listening tools such as sentiment analysis algorithms may not always accurately determine the intent behind a post or message. This can lead to misinterpretation and potential harm if misinformation is incorrectly identified. Human oversight and intervention are essential to validate and interpret the results accurately.

4. **Bias, discrimination and cultural sensitivity.** Different cultures may express health-related information differently. It is crucial to ensure that the monitoring methods used do not perpetuate or reinforce biases, discrimination, or stigmatization of particular groups or individuals. It is also essential to consider cultural context and avoid misjudging or misrepresenting cultural expressions as misinformation.

5. **Validity and feasibility.** In the WHO European Region, many digital platforms do not allow researchers to access and analyse data, resulting in validity and feasibility challenges *(6)*. Policy-makers and platform operators should collaborate to devise secure and privacy-conscious mechanisms that allow entities with legitimate purposes to access relevant data for analysis.

The five phases of false information management

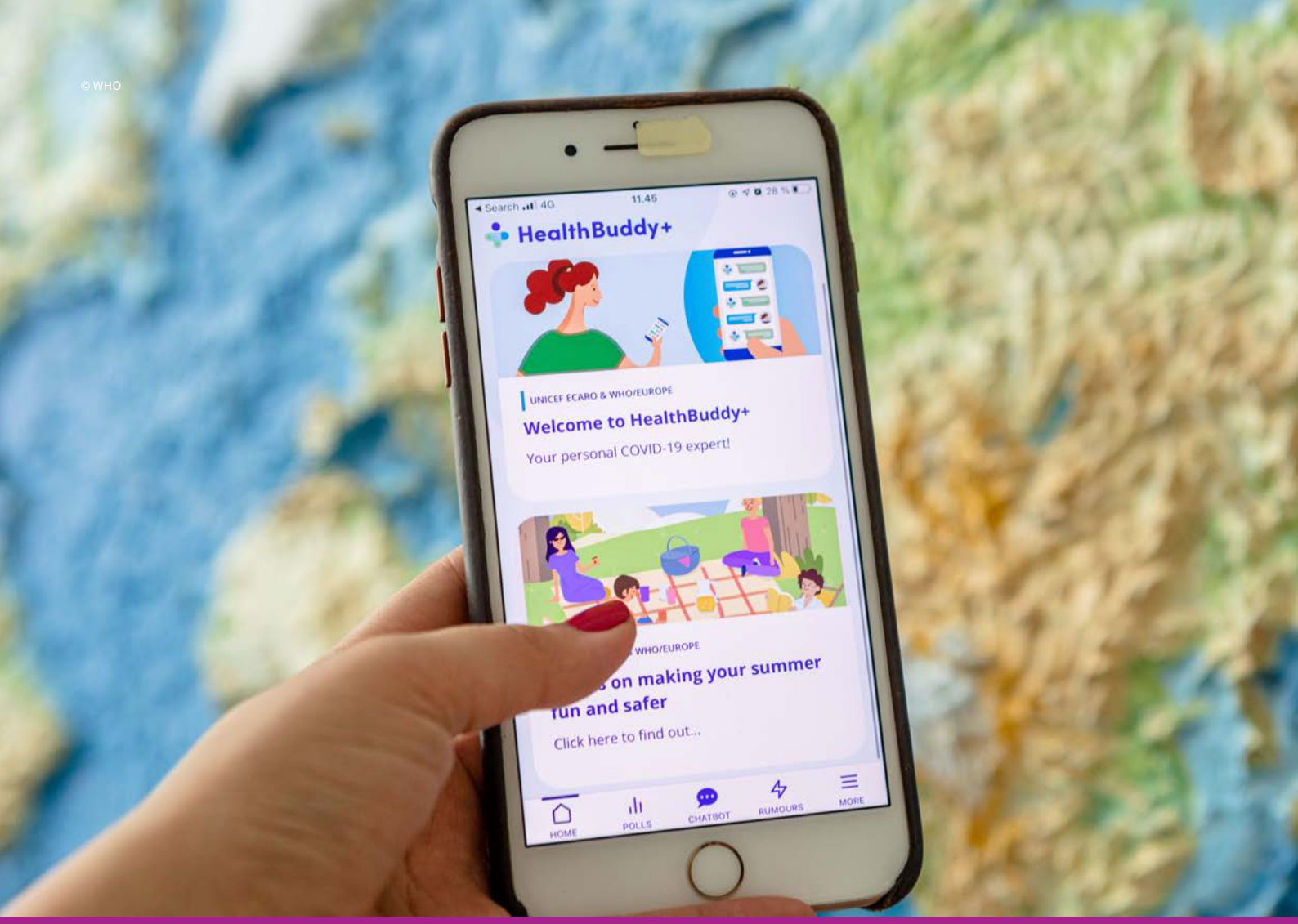
Signal detection

Signal verification

Risk assessment

Response design

Outreach



© WHO

The signal detection phase is critical to identifying potential outbreaks of false information and to help public health officials develop effective response strategies.

Phase 1: Signal detection

Description

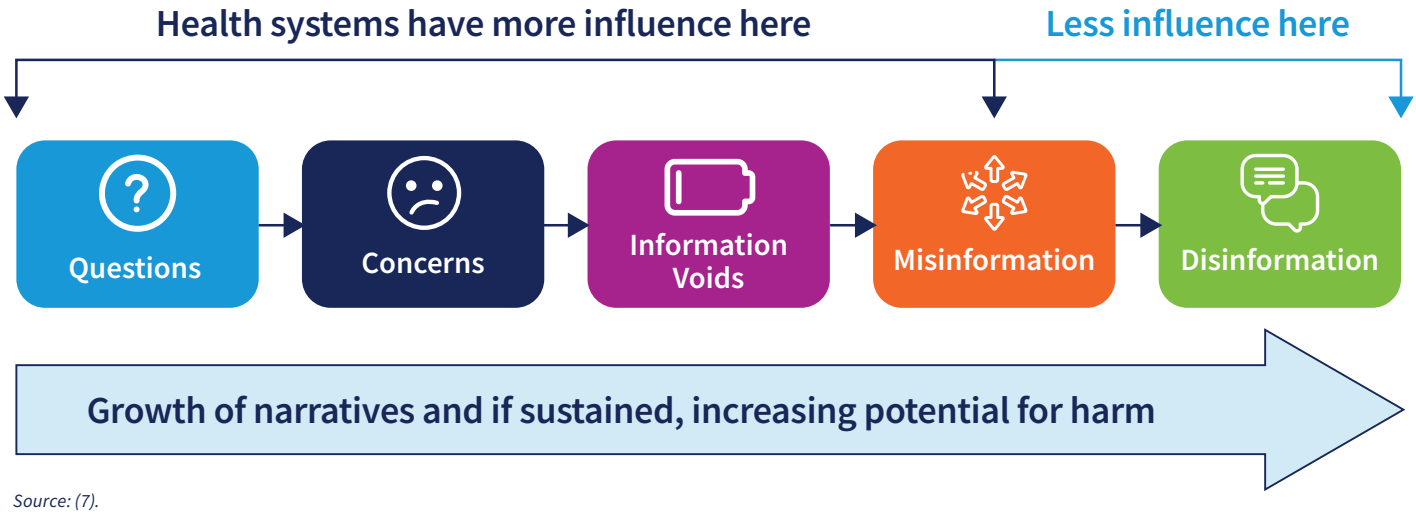
Signal detection involves establishing systematic processes to track information and data regarding a specific health topic, while also monitoring for potential emerging questions, concerns, rumours and false information. This can include monitoring social media, news outlets, online forums, call centre data and more.

The signal detection phase is critical to identifying potential outbreaks of false information and to help public health officials develop effective response strategies. By monitoring and analysing data effectively, public health officials can create risk communication strategies that are informed by users to inform people's decision making on their health protection. The goal of this phase is to identify false information during a health emergency. It lays the foundation for the subsequent phases of the process of verifying and addressing false information.

An infodemic is not solely composed of misinformation or disinformation; it also includes legitimate questions, concerns and information voids that people have. These signals serve as precursors to the spread of false information. Detecting the signals and addressing them – by filling information voids, answering questions, and alleviating concerns – is generally easier for health systems. Once misinformation and disinformation take root, their mitigation becomes more challenging (Fig. 2).

As introduced above, while offline information and data tracking is a valuable source of insights and will be mentioned in this toolkit, the focus will be online data monitoring. This is primarily due to the broader reach, immediacy and the dynamic nature of digital platforms, which provide real-time insights and facilitate prompt interventions in addressing false information.

Fig. 2. An infodemic is made up of more than misinformation



Source: (7).

Key steps in the signal detection phase

Step 1: Define the scope and objectives of the signal detection phase

The first step in the signal detection phase is to define the scope and objectives. This involves identifying the specific health-related topic or issue that requires monitoring, as well as the specific location or language, timeframe, goals and objectives of this exercise. Examples of scope setting include:

- What are the false information narratives surrounding mpox in the affected and vulnerable communities in country X?
- Who are the key generators of false narratives surrounding the avian influenza spreading to cats from birds in country Y?

The key aspects to be refined in this step include:

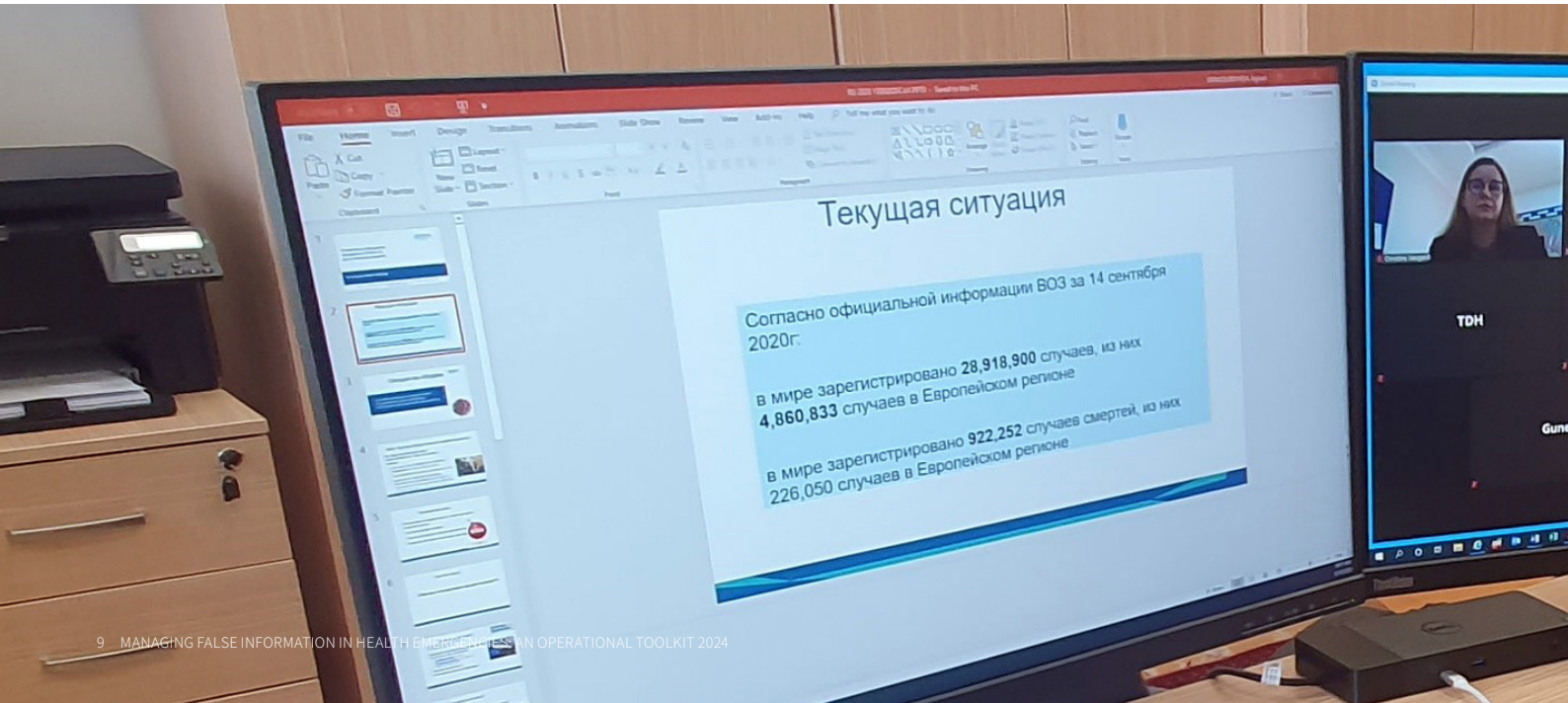
- **Selecting a specific topic that requires monitoring:** Determining the particular health topic or issue that needs close observation, such as a particular disease outbreak or a misleading health narrative.
- **Choosing analysis methods:** Selecting appropriate techniques for analysing the gathered data (see Step 3 below for available methods).
- **Defining boundaries of the analysis:** Establishing the time period for monitoring, target geographic areas to focus on, languages to consider and specific platforms where the information may be spreading.

Step 2: Develop a data collection plan and start collecting data

The next step in the signal detection phase is to identify and select data sources that will be used to monitor the spread of health-related false information. Data sources can include publicly available social media platforms, online news sources, online forums, blogs, and other relevant sources of information. The selection of data sources considers both the geographical region of interest and the health concern under examination. This targeted strategy ensures that monitoring is finely tuned to the locations and themes key to the study. For example, in some parts of the WHO European Region, Telegram and VKontakte are the most popular social media messaging channels, while in others Facebook or TikTok can be much more relevant.

Once data sources have been identified, the next step is to develop a data collection plan. Data collection is most often conducted through social listening. This involves determining the frequency and tools for data collection through social listening, as well as the specific data points that will be collected.

An essential aspect of the data collection plan is defining the relevant keywords and phrases to effectively monitor health-related false information. These keywords act as the search queries for social listening tools to gather specific data. The selection of keywords should be comprehensive and updated regularly to adapt to evolving trends and emerging false information.



Defining search terms

The quality and accuracy of risk signals and false-information narratives that are identified significantly depends on the search terms used for finding the signals. These search terms can be categorized into static and dynamic terms for clarity and ease of understanding.

- **Static search terms:** These are fixed terms used consistently for each report on a specific topic. Translations from the target languages of focus countries are also included.
- **Dynamic search terms:** These terms change based on current events, misinformation trends and the prevailing information-landscape. They are combined with the static search terms using appropriate Boolean operators (AND, OR, NEAR/5, etc.). These terms need to be updated regularly and translated into the local languages of focus countries.

In the COVID-19 example:

- **Covid-19 focused static search terms:**
  covid OR COVID-19 or couronne OR corona OR корона OR coronavirus OR коронавирус OR pandemic OR pandémie OR pandemie OR pandemia OR пандемия
- **Covid-19 focused dynamic search terms:** The search terms below are an extract from the dynamic terms list of one internal IM report (WHO Regional Office for Europe, unpublished, 2021) Please note that these terms below are for illustrative purpose only and need to be updated and refined for each production of an IM report based
- on current context.
  - **NEAR/5**
    (“cardiac-related death” OR “blood clotting” OR transmission OR booster* OR jab OR vaccin* OR “Swab test” OR “blood clots” OR “vaccine side effects” OR “prolonged symptoms” OR “vaccine passport” OR immunity OR lockdown OR masks OR “mask wearing” OR mask OR deoxygenation OR “insulin cost” OR Ivermectin OR “big pharma” OR bigpharma OR “lab-created” OR “US Labs” OR “United states labs” OR “wuhan lab” OR depopulation OR fertility OR “home remedy” OR “magical cure” OR herbs OR rash OR rashes OR DNA OR dna OR fertility OR pharma OR pharmac*)
  - **NOT**
    (Halloween OR wearables OR pele)

If there is a budget available for data collection, a diverse range of digital tools can be utilized to run searches and conduct sophisticated analyses in real time or periodically. These tools and platforms are primarily designed to cater to the needs of the private sector, serving purposes such as market research and brand reputation monitoring. Examples of widely used paid social listening tools and services include TalkWalker, CrowdTangle, Sprout Social and Brandwatch and an example list of free tools can be found in Box 1.[1] However, it's important to note that while these tools offer valuable functionalities, they may not be optimally suited for professionals seeking specific insights in the field of public health.

Box 1. Free social listening tools

1. **Google Alerts** *(8)* enables users to monitor the web for specific keywords and phrases. Users can set up alerts to receive notifications when new content is published that matches their search criteria.
2. **Google Trends** *(9)* allows users to see what people are searching for on Google. It provides insights into search interest over time and by location, as well as related topics and queries. Google Trends can be used to identify trending topics and keywords related to health-related misinformation, allowing users to monitor and analyse conversations and develop appropriate responses.
3. **Followerwonk** *(10)* provides free options that enable users to focus their social listening efforts on Twitter. It provides a range of analytics and insights, including follower demographics, social authority. and social influence.
4. **Hoaxy** *(11)* is a web-based tool that visualizes the spread of articles online. It searches for claims and fact-checking going back to 2016 and tracks the sharing of articles. It can be used to identify the spread of misinformation and disinformation.
5. **Bot Sentinel** *(12)* identifies and tracks bots and trolls on Twitter. It provides a range of analytics and insights, including bot scores, troll scores and sentiment analysis.
6. **Talkwalker Free Social Search** *(13)* is a social listening tool which provides a variety of metrics including top themes, influencers, engagement, sentiment and reach. The free version of Talkwalker is limited to 7 days of historical data.
7. RAND Corporation has compiled a repository of free tools that can work for specific use cases in addressing false information *(14).*

1 The mention of these tools and services (both paid and free) does not constitute an endorsement by the World Health Organization.

Best practices: online data sources commonly used during the COVID-19 pandemic

Given the large volume of content posted each day on social media, analysts need to take a targeted approach. Good practices demonstrated by national health authorities during the COVID-19 pandemic include:

- using Google Trends and other free tools to check what people in their country or region are searching for online;
- monitoring the online reaction (views, likes and shares) and comments produced in response to social media posts made by health authorities;
- conducting short opinion polls (one or two questions) using the polling feature on social media channels of the health authority or WHO Country Office;
- monitoring comments and reactions to news stories posted on social media by leading national, local and/or international news organizations;
- monitoring conversations in publicly available and open access key health and/or news discussion forums online; and
- monitoring social media posts made by societal leaders and influencers and the responses to these posts.

Data collected using the practices mentioned above has been used by national public health authorities to conduct a signal analysis and risk assessment of their findings. Typically, one or more members of an RCCE-IM team would review the signals and produce a report on how public sentiment was evolving, the main rumours and false information circulating, new narratives emerging and the latest trends in what people were searching for online. This was done daily by multiple health authorities during the most acute phases of the pandemic, or weekly, bi-weekly, or monthly during less acute phases.

Key resourcing issues to address when setting up a signal detection system

1. Define a realistic labour input from your team: e.g. 7 staff hours per week (1 hour per day, or 2 half days), 20 staff hours per week (one staff member working 50% or two working 25% on this task), 40 staff hours per week (one person working full time on monitoring).
2. Make a realistic assessment of how many social media accounts and other online sources can be read and analysed in the time available, then prioritize among accounts and other online sources to be monitored.
3. Create a monitoring plan linked to the signal detection objectives. This should define the sources to be monitored, how often content will be reviewed (e.g. daily, weekly or monthly) and what will be looked for.

Step 3: Analyse data and identify trends

Data analysis should be conducted to identify patterns and trends in the information being monitored. This can involve identifying the sources of false information, the types of false information being spread, and the rate and speed of dissemination. There are various methods that analysts can apply to the collected data to detect false information (Table 1). Most of these analytical methods are readily available in various social listening tools. Detailed guides on these methods are beyond the scope of this toolkit.

Step 4: Adjust signal detection plan

The signal detection plan should be regularly reviewed and adjusted based on the findings of the analysis. If the volume of social listening data is low or if the results are not relevant, data sources, search terms or data collection methods should be adjusted or new trends or areas of concern identified.

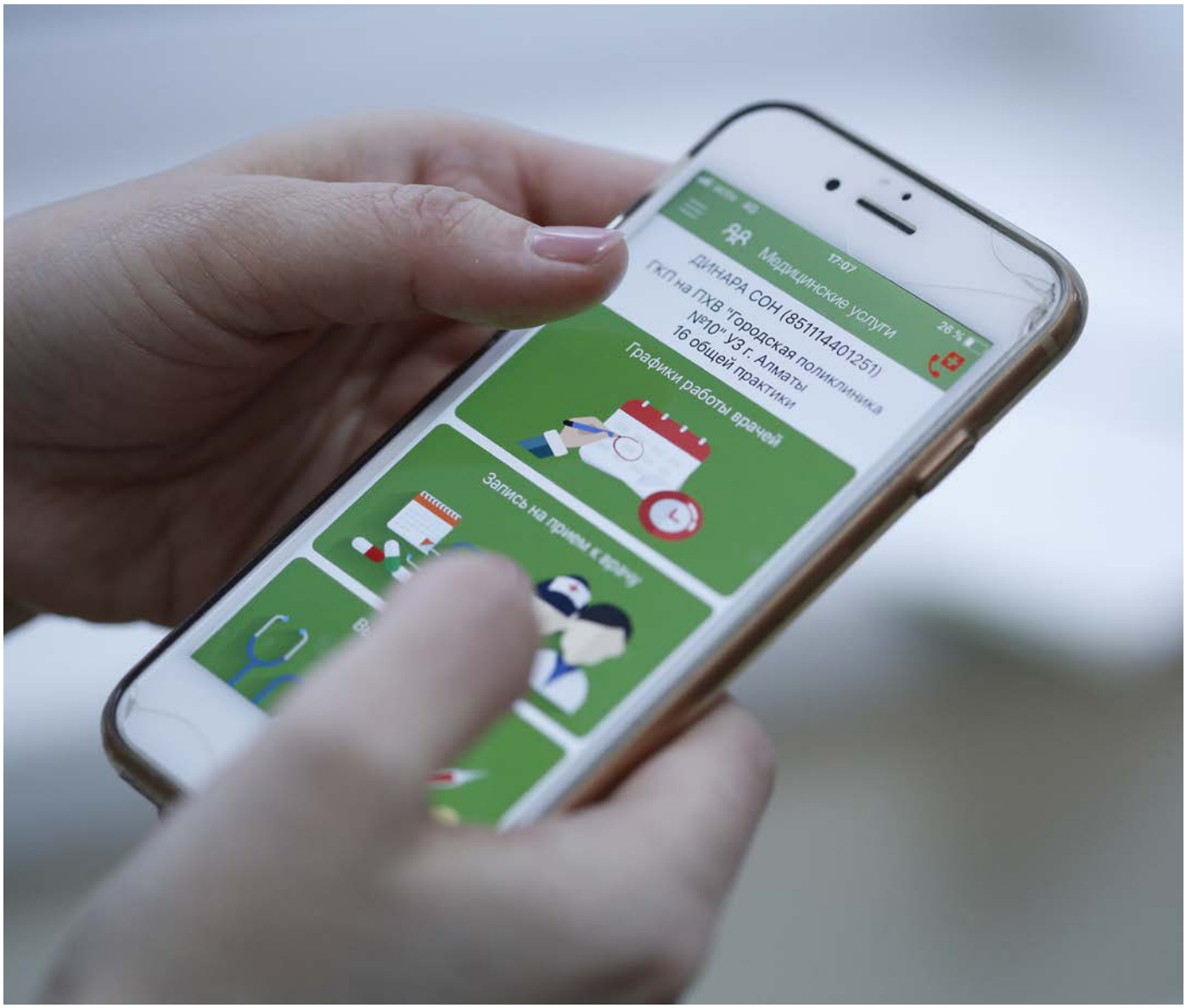
Table 1. Methods used by analysts to detect false information

Analysis type	Description
Text analysis	This involves analysing the language used in social media posts and online content to identify patterns and trends in how information is being shared. Text analysis can be used to identify the use of specific keywords or phrases that may be associated with false information, as well as to track changes in the language used to describe a particular health issue. • **Example:** In the case of an mpox outbreak in Country X, analysts use social listening tools to find media articles and social media posts and categorize them based on the WHO Public health taxonomy for social listening on mpox conversations *(15)*. Posts are assigned one of the following categories: cause, illness, treatment, interventions or meta-conversation.
Mention volume trend analysis	This involves tracking the volume of mentions of a particular health issue or topic over time to identify changes in how the issue is being discussed online. By analysing the content of these mentions, public health officials can identify whether false information is becoming viral and can take steps to counteract this by promoting accurate information and targeting messaging and outreach efforts to the areas where the false information is most prevalent. • **Example:** While monitoring the volume of mentions of mpox in Country X over time, analysts notice a sudden increase of false narratives, which is linked to an influencer making a misinformed claim on social media.
Sentiment analysis	This involves analysing the emotional tone of social media posts and online content to identify patterns in how people are reacting to a particular health issue. False messaging is engineered to go viral. Content that provokes strong negative emotions, such as hate, disgust and indignation, is more likely to spread quickly. Sentiment analysis can be used to identify the spread of false or misleading information that is generating strong emotional reactions among the public. • **Example:** During the mpox outbreak in Country X, analysts use social media analysis tools to automatically assess mpox conversations and assign them a defining sentiment: positive, negative or neutral. An unexpected rise in negative emotions is linked to false claims, allowing targeted risk communication to calm public anxiety.
Network analysis	This involves mapping the connections between individuals and groups who are sharing information about a particular health issue. Network analysis can be used to identify key influencers who are spreading false information and to track the spread of false information across different social media platforms. • **Example:** Mapping the network of connections between social media health influencers in Country X that are sharing information about mpox allows the analyst to understand that a network of only 10 people are responsible for more than 80% of the reach of the mpox-related social media posts. This leads the analyst to suggest (further) engaging those influencers in public outreach on the topic.
Geographic analysis	This involves analysing the geographic distribution of social media posts and online content to identify patterns in how information is being shared across different regions. Geographic analysis can be used to identify areas where false information is particularly prevalent and to target messaging and outreach efforts to these areas. • **Example:** Using Google Trends to look up which regions in the country are most interested in mpox and to identify key questions and concerns, the analyst can then propose region-specific RCCE-IM interventions.

Leverage community engagement for signal detection

Community engagement plays a crucial role in signal detection. Two methods can be employed to incorporate community engagement into signal detection.

1. **Conduct regular RCCE-IM surveys.** Identify communities at-risk or those who have been historically at-risk in previous health emergencies. This can be done through mapping community-based actors and structures or organizing community meetings in your country or area. Local emergency responders, such as national Red Cross and Red Crescent societies may have established community listening systems. Leverage those systems or develop mechanisms to regularly engage the communities to understand their concerns, questions and any rumours or false information that is circulated.
2. **Understand how the community interacts with health information.** Take into account the community context, ranging from preferred communications channels and style, to the main community influencers. This can be done through key informant interviews or focus groups with community members. Existing behavioural and cultural insights studies may provide insights into preferred channels and trusted sources of information.



© WHO

Comprehensive signal verification ensures that the subsequent risk assessment phase is based on verified and credible information.

Phase 2: Signal verification

Description

The goal of signal verification is to verify the accuracy of relevant signals identified during the signal detection phase. The focus is on gathering additional information and evidence to determine **whether the signals represent real and accurate information or if they are false or misleading**. Signal verification helps ensure that the subsequent risk assessment phase is based on verified and credible information.

Key steps in the signal verification phase

The main steps in the signal verification phase of the process include the following steps.

1. Identify the source of the signal to determine its credibility.
2. Check for supporting evidence that can corroborate the signal.
3. Check for contextual information, such as the timing and location of the signal.
4. Verify with experts or authorities to validate the information.
5. Summarize findings to inform the next phase (risk assessment).

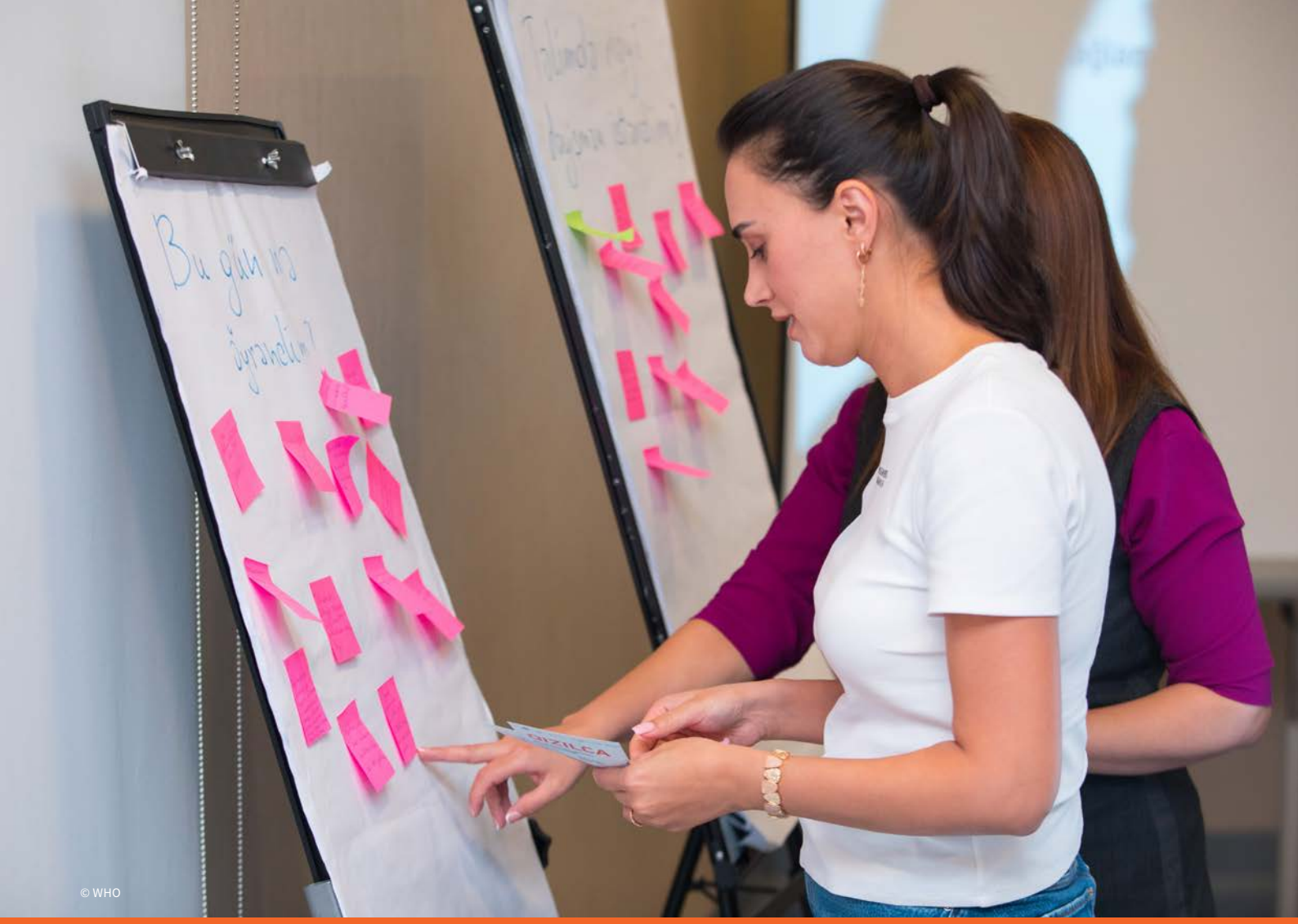
These activities help to ensure that the signals detected during the surveillance phase are validated, and that the information used in the subsequent risk assessment phase is accurate and reliable.

How to verify collected signals

Simple techniques and action-oriented steps for signal verification are described in Table 2.

Table 2. Signal verification activities

Activity	Description
Cross-checking	Cross-checking can be done by verifying the information from trusted sources to determine its accuracy. This can include conducting keyword searches on different search engines, reviewing social media platforms and checking news articles from various sources.
Source verification	Source verification involves determining the credibility and reliability of the sources that provided the information. This can be done by checking the background of the sources, their track record in providing accurate information and their affiliations.
Fact-checking	Fact-checking involves verifying the accuracy of the information by consulting reliable sources such as scientific research, government agencies, and reputable news organizations. Fact-checking resources such as the European Digital Media Observatory directory *(16)* can also be used.
Expert consultation	Expert consultation involves seeking the opinion of subject matter experts such as epidemiologists, clinicians and researchers to verify the accuracy and relevance of the information.
Documentation and reporting	Documentation and reporting involve keeping track of the sources of the information, of the verification process and of the results of the verification. This can help in identifying patterns and trends in false information and can be used for future reference.



© WHO

Not all signals need a response. Conducting a risk assessment allows responses to high-risk signals to be prioritized.

Phase 3: Risk assessment

Description

The goal of the risk assessment phase is to assess the public health threat posed by the signals identified during the signal detection and signal verification phases.

By performing an integrated analysis of social listening signals and other data sources, the outcome of this phase is a decision on **whether the identified signal warrants a response or not**. Not all signals need to be responded to and conducting a risk assessment allows responses to high-risk signals to be prioritized while also avoiding the amplification of low-risk rumours and false information.

The results of the risk assessment phase then inform the response design phase, in which appropriate responses are developed based on the identified risks. This phase also links to the outreach phase, as the risk assessment may identify specific audiences that are particularly vulnerable to the risks associated with the false information and who may require targeted outreach efforts.

Online social listening in combination with offline community engagement can be used to identify particularly vulnerable communities to false health information by tracking, analysing, and synthesizing community inputs both digital and offline. This process can help identify questions, queries, concerns, complaints and suggestions shared by communities, which can be integrated, categorized and analysed to produce actionable insights. By understanding the information needs of vulnerable communities, RCCE-IM interventions can be developed to address their specific concerns and promote accurate health information.

Key steps in the risk assessment phase

Three questions to guide the risk assessment:

1. How extensively has the signal spread among the target audience?
2. How influential is the source of the signal?
3. What level of risk does the signal pose to public health?

Risk assessment of infodemic signals is not a rigid formulaic process. While these guiding questions and risk assessment frameworks such as the one in Fig. 3 are helpful in reducing bias and noise in judgment *(17)*, they should be seen as tools that inform a human decision. The analyst's deep contextual understanding is crucial for accurate risk assessment.

Examples of risk assessment for specific signals are provided in the sections overleaf to illustrate this process at a tactical level.

Fig. 3. Example risk evaluation matrix.

Indicator	Low Risk	Medium Risk	High Risk
Risk to vaccine hesitancy and demand	Low risk to vaccine demand	Potential to trigger hesitancy to vaccinate	Potential to lead vaccine refusals
Reach and scope of misinformation	Limited potential reach or scope	Moderate potential reach or scope	Wide or cross-country reach or scope
Likelihood of issue spread or escalation	Unlikely to spread in community or online	Spreading in community and/or online	Spreading rapidly in community and online
Response capacity	Strong messaging and capacity in place	Limited existing messages and resources to manage crisis	Limited existing messages and capacity exceeded
General public trust	Remaining trust in government, health services, vaccines	Reduced trust in government, health services, vaccines	Outward displays of mistrust government, health services, vaccines
Response	Monitor closely, consider prebunking	Debunk, raise trusted voices	Debunk, raise trusted voices

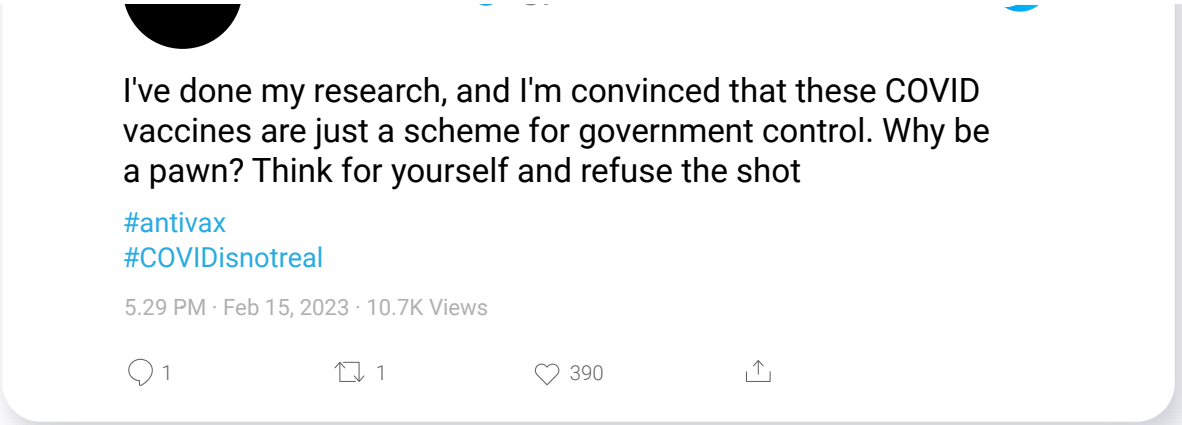
A closer look at each question

How extensively has the signal spread among the target audience?

The initial step in risk assessment involves determining the current and potential virality of the signal. By conducting signal detection and verification, statistical information about the signal can be gathered, including the following.

- **Reach:** The number of unique users who have seen a piece of content. This metric can be used to measure the size of the audience and the potential impact of this content.
- **Impressions:** The number of times a piece of content has been displayed. This metric can be used to measure the potential reach of this message.
- **Engagement rate:** The percentage of users who have engaged with a piece of content, such as likes, comments and shares. This metric can be used to measure the effectiveness of this content and the level of audience engagement.

Example post (content altered to protect privacy):



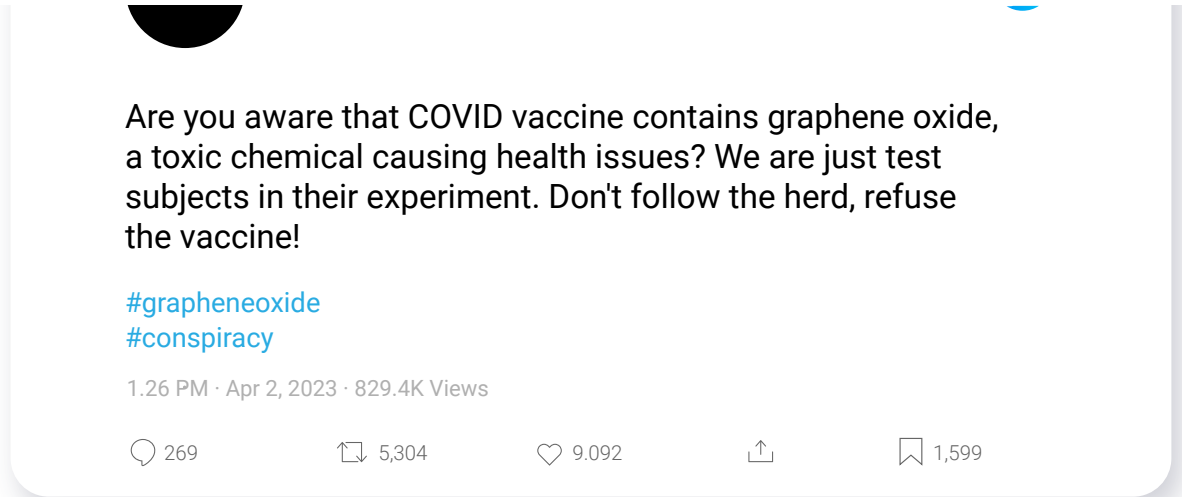
Reflection: This hypothetical tweet has minimal retweets and likes (engagement) even if it has garnered a significant number of views (more than 10 000).

Recommendation: Analysis shows that the signal has a low virality (not many similar posts have been shared after its posting). An analyst could recommend not to debunk or respond directly, but to simply keep monitoring for similar themes.

How influential is the source of the signal?

The second goal of the assessment is to determine the influence of the source regarding public health and scientific matters. Social media platforms provide a voice and the ability to connect with broad audiences. Often the influence does not depend on the level of expertise on the subject matter, but on the extent of the outreach. While it is concerning if a health professional or expert shares false information, in some cases it can be even more detrimental and harmful to trust if an athlete or musician with a large fanbase expresses skepticism towards COVID-19 vaccination.

Example post (content altered to protect privacy):



Reflection: In order to assess the author's influence, the following metrics and aspects can be considered.

Follower count: One of the most straightforward metrics is the number of followers, which generally indicates a larger audience and greater influence. While it's challenging to set a definitive threshold for a "large enough" follower count needing a response, a general guideline might be to prioritize sources with followers exceeding 100 000. Instead of a fixed number, an analyst could use a percentage-based system as a guideline: for example, if a source reaches or influences more than 1% of the target demographic or population, it necessitates a response.

Engagement: Beyond the raw follower count, the interaction the source has with its audience is crucial. A source with a smaller follower count but high engagement may have a more dedicated and active audience that is more likely to act on their message.

Domain credibility: While the source might be influential and credible in a different domain (like music or sports), some of that influence will remain in other domains and still make their messages impactful.

Past endorsements: If the source has previously been endorsed or amplified by other influential figures or entities, it can boost their influence and credibility among certain audiences.

Recommendation: Analysis reveals that the author has great influence on his large number of followers, even though the author is not a scientific organization or subject matter expert. The post has received significant engagement and almost a million views. An analyst could recommend a targeted debunking response to this signal.

What level of risk does the signal pose to public health?

When conducting our risk assessment, the last question calls for reflection on the threat to public health and safety due to the spread of false information. A few key points to consider are:

- **Severity:** what is the potential harm or impact of the information on public health?
- **Vulnerability:** which groups of people are particularly vulnerable to the health risk?

Example:

During the COVID-19 pandemic, an influential public figure suggested that injecting bleach or disinfectants could cure or prevent the disease *(19)*.

Reflection: This is a dangerous and potentially deadly suggestion as bleach and disinfectants are toxic substances that can cause serious harm to the human body.

Recommendation: This signal not only has high reach and comes from an influential figure, but the suggestion to inject bleach made by the public official also poses a major health risk. An analyst's recommendation could be to immediately issue a public health alert advising people not to inject bleach as it can cause serious harm and even death.

Reporting on the outcome of the risk assessment

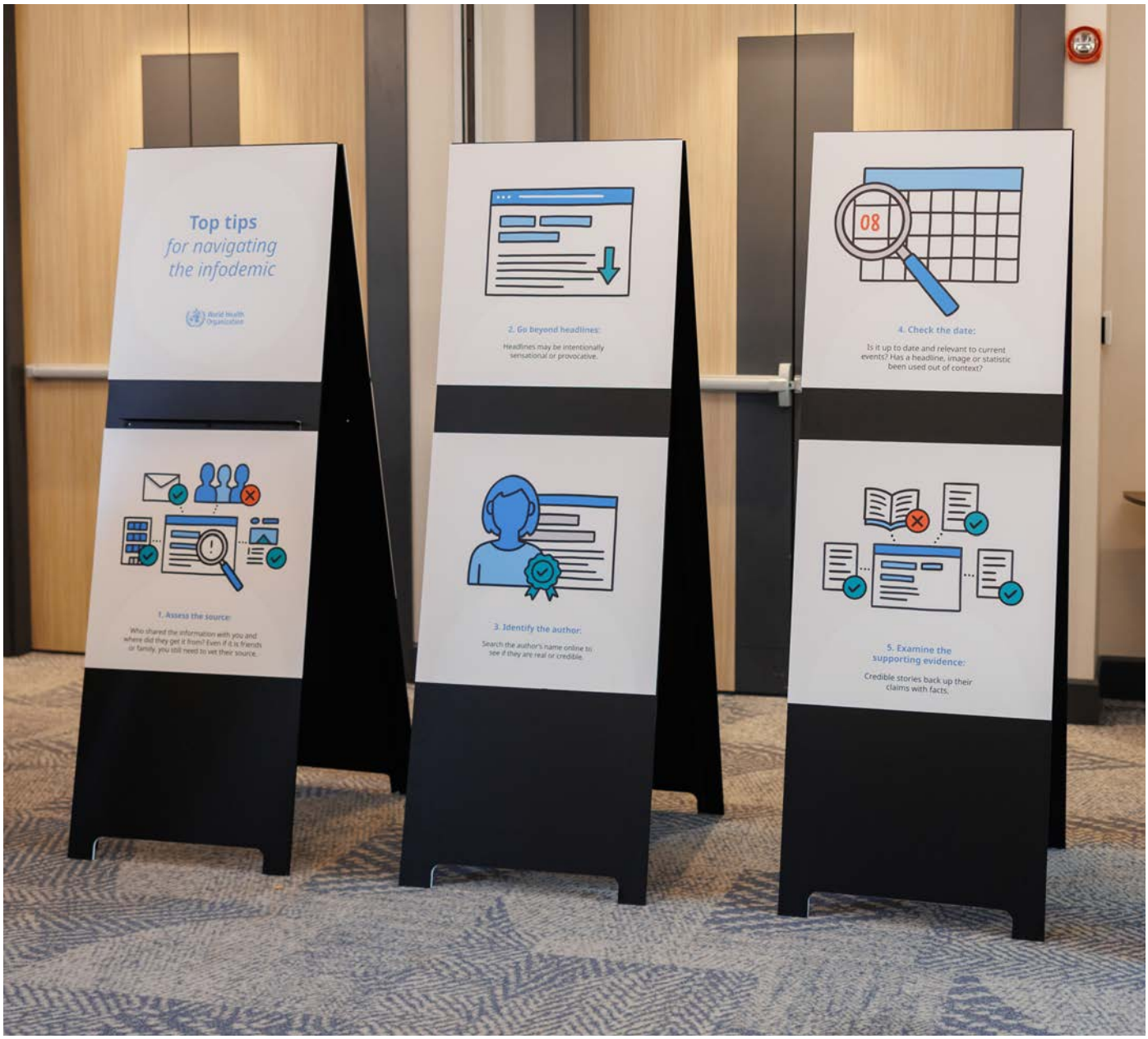
- To effectively communicate risk assessment results, it is crucial to generate regular reports that are readily shared with stakeholders. Stakeholders encompass a wide range of individuals and groups, including internal team members, other sector authorities and administration levels, CSOs, fact-checkers, international organizations and other pertinent parties.
- For more information on building an infodemic insights report, read the WHO and United Nations Children's Fund manual *How to build an infodemic insights report in 6 steps (20)*.

Leverage community engagement for risk assessment

Risk assessment becomes more accurate and relevant when community engagement is used in the process. Two approaches illustrate how community engagement can be used during risk assessment:

1. **Triangulate using other public engagement mechanisms:** if other engagement mechanisms are set up in the community, such as rumour reporting *(21)*, a signal to be assessed can be cross-checked to have a fuller understanding of whether the signal is also appearing in community settings. Local emergency responders might have community listening systems that can also be used to triangulate rumours.

2. **Tap into established relationships:** now is the time to engage with CSOs and community leaders. Tapping into established relationships enables a secondary step in risk assessment which allows for assessment of the reach of the signal and how the signal is perceived by a particular community. This step plays a pivotal role in fostering trust, which is essential for an effective RCCE-IM intervention.



Effective response interventions must recognize that knowledge alone does not guarantee action. People's behaviour is influenced by beliefs, cultural norms, emotions and social pressures.

Phase 4: Response design

Description

The response design phase is focused on designing an effective response plan as needed, including the development of key messages, selection of appropriate communication channels, and the creation of materials and resources.

The response design phase relies upon insights generated in the risk assessment phase. The high-priority signals identified during the risk assessment phase form the basis for designing response interventions. Response can be carried out through both online and offline channels. Interventions can include targeted messaging, engaging with key influencers, developing risk communication materials or other activities to address a specific false information narrative.

This phase also involves the ongoing monitoring of the situation to ensure that response interventions are effective. Adjustments need to be made based on new developments and feedback from stakeholders.

Effective response interventions must recognize that knowledge alone does not guarantee action *(22)*. People's behaviour is influenced by a complex interplay of beliefs, cultural norms, emotions and social pressures. For example, even when parents understand the importance of keeping a child with measles at home to prevent spreading the infection, they may still send them to school due to work commitments, social obligations or misunderstanding the severity of the situation. By integrating behavioral frameworks, interventions can target these underlying factors, resonating with the way people actually think and behave. This approach is more likely to lead to meaningful changes in behaviour, as it considers not just what people need to know, but also what motivates them to act.

Key steps in the response design phase

There are 10 main activities in the process of designing a response to false information.

1. **Identify the target audience(s)** for the response to determine who needs to receive the message and what their characteristics and communication preferences are.
    - Who should take action?
2. **Define the goals and objectives of the response** in a clear and measurable way to achieve the desired outcomes.
    - What do we want our target audience to do?
3. **Identify and engage stakeholders and partners** who can support the response effort and engage them in the process. Co-design response efforts where appropriate.
    - Who can help us achieve our desired outcome?
4. **Develop a rapid response outline** that outlines the tactics, timelines and resources needed to implement the response effectively.
    - How do we plan to achieve our desired outcomes?
5. **Develop response messages** by crafting clear, concise, and compelling messages that are tailored to the target audience and that address the specific concerns and false information being circulated.
    - What are the key actionable messages that help us achieve our desired outcomes?
6. **Determine the response channels** and select the most effective channels and trusted messengers to deliver the response messages to the target audience(s), such as social media, traditional media, through CSOs and through other offline channels.
    - What are the channels that our audience(s) use and the influencers they trust?

7. **Create response materials** and develop a range of materials to support the response, such as tweets, longer posts, social media tiles, fact sheets, infographics, videos and other social media content.
    - What are the materials that our audience(s) would be most engaged with?
8. **Test and refine the response** by conducting small-scale tests of the response messages and materials with members of the target audience(s) and refine them based on feedback.
    - How are our messages and materials received by target audience(s)?
9. **Monitor and evaluate the response** by regularly monitoring the effectiveness of the response, gathering feedback from the target audience(s), and evaluating the impact of the response on attitudes and behaviours via CSOs and other on-the-ground partners.
    - What are the key performance indicators we need to put in place to evaluate our intervention(s)?
10. **Learn and enhance the response** by incorporating feedback and adjusting key messages or messaging formats as necessary.
    - What are the main findings we have learned from monitoring and evaluation that need to be reflected into our plan?

How to develop effective response messages

The characteristics of the most effective response messages are presented in Table 3.

When considering response interventions, there can be two approaches.

1. **Developing corrective messaging:** Corrective messaging involves creating and disseminating accurate information to directly counteract the false information that has been spread. This messaging should be carefully crafted to speak to the specific needs of the target population and ensure that it effectively addresses the specific concerns and questions raised by the original false information. One example of corrective messaging is the debunking strategy.
2. **Developing counter messaging:** Counter-messaging involves creating messages that offer a different perspective or alternative explanation of the issues at hand without correcting it directly. This can be an effective strategy for addressing false information that is difficult to correct. One example of counter messaging is the prebunking strategy.

The following two sections will go into detail on these two response techniques, debunking and prebunking.

Table 3. Characteristics of effective response messages

Characteristics	Explanation
Timeliness	The response message should be delivered as quickly as possible to prevent the spread of false information or confusion.
Clarity	The message should be clear and easy to understand, use simple language and avoid jargon.
Accuracy	The response message should be based on accurate and reliable information from credible sources.
Specificity	The message should be specific to the topic or issue being addressed rather than general or vague.
Consistency	The message should be consistent with other messages from the same source and with information from other credible sources.
Actionability	The message should provide clear and actionable steps that the target audience can take to protect themselves or address the issue.
Empathy	The message should be delivered in a tone that is empathetic and understanding of the concerns, emotions and beliefs of the target audience.

Debunking as a reactive response technique

What is debunking?

Debunking is a method for exposing and correcting false or misleading information. This involves using evidence-based and logical arguments to challenge and disprove claims that are not supported by facts or scientific evidence. Debunking techniques can include fact-checking, source verification, expert opinions and the development of critical thinking skills. By using a debunking intervention, it is possible to reduce the spread of false information and promote accurate information.

How does debunking work?

When people are exposed to false information repeatedly, they may begin to accept it as true, even if it goes against their preexisting beliefs. However, when the false information is challenged and corrected, people may adjust their beliefs accordingly.

Table 4 describes the key factors that determine the effectiveness of the debunking technique.

Example: One example of debunking in a public health emergency is the WHO's Mythbusters webpage *(23)* that during the acute phase of the emergency provided accurate and reliable information about COVID-19 and addressed some of the common myths and misconceptions circulating in the media and social networks.

Table 4. Key factors determining effective debunking

Factor	Description
Timing	Debunking is most effective when it's done quickly and before the false information has a chance to spread widely and become entrenched in people's beliefs. *Example*: When a celebrity posts on Facebook a misleading fact about harms caused by vaccines, a health organization promptly replies with accurate information, preventing the misinformation from spreading widely.
Audience	The effectiveness of debunking can vary depending on the audience. Some people may be more resistant to changing their beliefs, particularly if those beliefs are deeply held and important to their identity. Once we have a clear understanding of our target audience, we can frame the accurate information in a way that is relevant and resonant with that specific audience. This could include using techniques like storytelling, personal anecdotes or emotional appeals to help our audience connect with the information on a deeper level. *Example:* A local public health entity recognizes that some older adults in their area are resistant to a new medical treatment, so they organize a town hall meeting with trusted local doctors to connect with the audience's values and experiences.

Table 4. Key factors determining effective debunking (continued)

Factor	Description
Message framing	The way the debunking message is framed can affect its effectiveness. Research has shown that debunking messages that focus on the correct information rather than the false information are more effective *(24)*. Also to be considered: Emotions: Our beliefs can be driven by emotions such as fear or anger. By acknowledging and empathizing with these emotions, trust and credibility can be built with the target audience, making them more receptive to the correct information. Context: False information is often misleading because it lacks important context. By providing added context, such as explaining the limitations of a study or the broader context of a news story, the effects of the false information can be countered. *Example:* A public health campaign seeks to debunk common misconceptions about the flu vaccine. Instead of solely focusing on the false information, the campaign emphasizes the correct information, using relatable stories and statistics. They also include a video with personal experiences from individuals who benefited from the vaccine. To connect with emotions, the campaign acknowledges common fears and concerns, providing reassurance through expert testimonials. It also provides context by explaining how vaccines are tested and approved.
Trustworthiness of the source	The credibility of the source delivering the debunking message is important. People are more likely to accept debunking information from sources they trust; these may or may not be public health officials and health-care workers, but the engagement of trusted influencers is key to establish this trust. *Example:* To counter misinformation about a public health crisis, a government agency collaborates with faith leaders, leveraging their credibility and trust within the community.
Clarity and simplicity of the message	Debunking messages that are simple, clear and easy to understand are most effective. Visuals can include tools such as graphs, charts and infographics which are effective at debunking false information, as they can help make complex information more accessible. *Example:* Instead of publishing a 50-page brochure, an environmental organization creates an easy-to-understand video debunking false information regarding the health impacts of climate change. They use simple language and clear visuals and make the information more accessible to the public.
Selecting channels	By using channels where the target audience is most active and engaged, the likelihood of the message being seen and engaged with is increased. Therefore, it is important to research and identify these channels, such as social media platforms, email newsletters or community groups, and utilize them for targeted messaging. *Example:* A nonprofit organization aimed at addressing mental health in war refugees researches the platforms that their demographic use to access health information. Based on the research, the nonprofit launches campaigns on those platforms to spread the message effectively.
Consensus	Debunking can be more effective when there is a consensus among experts or authoritative sources on the correct information. Working in partnership with other health agencies and relevant stakeholders can amplify the debunking messaging. *Example:* To debunk the myth that drinking cold water is unsafe during heat waves because "blood vessels would explode" *(25)*, CSOs and public health experts come together to issue a joint statement, demonstrating a unified agreement on the facts.

When – and why – does debunking not work?

While debunking can be an effective response technique, not all attempts to debunk false information are successful. In some cases, debunking efforts can backfire, leading target audiences to become even more entrenched in their false beliefs.

There are many psychological factors at play (see Table 5).

Table 5. Factors at play in unsuccessful debunking

Factor	Explanation
Backfire effect	Rarely, when people encounter information that contradicts their beliefs, they may become even more entrenched in their false beliefs. This is known as the backfire effect and it can occur when debunking is done in a way that challenges people's identity or core values *(26)*. *Example:* Some people may believe that ivermectin is a natural and safe alternative to vaccines, which they perceive as risky or harmful. If they are confronted with evidence that ivermectin is not effective or safe for COVID-19, they may feel threatened and defensive, and reject the correction. They may also rationalize their belief by finding flaws in the evidence or sources, or by seeking out more supportive information. *Mitigation:* Engage respectfully and empathetically, affirming the individual's values and identity before presenting contradictory evidence.
Familiarity effect	Repeated exposure to false information can make it seem more familiar and therefore more believable. Debunking may not be effective in correcting false information that has already become familiar to people *(27)*. *Example:* Individuals may have been exposed many times to the claim that ivermectin is effective against COVID-19, getting this message from social media, news outlets or from friends and family. If they are presented with a debunking message that contradicts this claim, they may not pay attention to it or remember it, because it is less familiar than the false information. *Mitigation:* Repeat the accurate health information frequently and through various channels, to build familiarity with the truth.
Overconfidence bias	People may believe that they are less susceptible to false information than others, which can make them resistant to correction. Such biases have been recorded at higher rates in people with higher educational attainment *(28,29)*. *Example:* Some people may think that they are well-informed about ivermectin and COVID-19, and that they can distinguish between true and false information better than others. If they are exposed to a debunking message that challenges their belief, they may dismiss it as irrelevant or inaccurate, because they trust their own judgment more than the external source. *Mitigation:* Frame the debunking information in a way that appeals to the individual's sense of intelligence and critical thinking.
Confirmation bias	People may seek out and believe information that confirms their preexisting beliefs, while discounting information that contradicts them. This can make them resistant to correction *(30)*. *Example:* Some may have a strong preference for ivermectin over vaccines for COVID-19, because of their personal values, experiences or emotions. If they encounter a debunking message that shows that ivermectin is not effective or safe for COVID-19, they may ignore it or reject it, because it does not fit with their worldview. They may also look for more information that supports their belief in ivermectin. *Mitigation:* Present information from sources that align with the target audience's worldview and create opportunities for active engagement.

Debunking techniques

There are several debunking techniques *(31)* that can be used to counter false information. Two of the most effective techniques – "truth sandwich" and the refutation technique – are presented below.

Truth sandwich

The "truth sandwich" is a technique used to refute health misinformation, which involves presenting the truth, briefly describing the falsehood, and then repeating the truth *(32).* This technique is designed to avoid further spreading misinformation while still addressing it.

Example: Truth sandwich

False information: "Measles, mumps and rubella (MMR) vaccines are not safe and can cause autism."

Applying the truth sandwich technique:

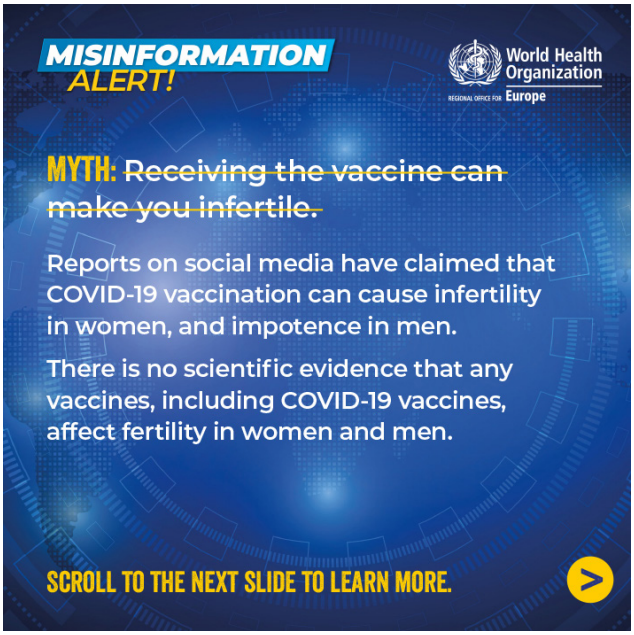
- Truth: MMR vaccines are safe and important vaccine for children.
- Falsehood: While some non-experts believe that vaccines can cause autism, there is no scientific evidence to support this claim.
- Truth: In fact, numerous studies have shown that vaccines are safe and do not cause autism. Getting vaccinated is the best way to protect yourself and others from serious diseases, talk to your doctor about vaccination."

Example: A WHO Regional Office for Europe campaign message debunking a common myth regarding COVID-19 vaccines and fertility (WHO Regional Office for Europe, unpublished, 2023)

By using the truth sandwich technique, false information can be corrected while still acknowledging and addressing it. This can help to build trust and credibility with the target audience and make them more receptive to the correct information.

Refutation technique

The refutation technique involves directly refuting false information with evidence and alternative information. It aims to correct false information by presenting accurate information in a clear and concise manner. While refutation can be a valuable tool, it should be complemented with other approaches, such as proactive communication, building trust and promoting accurate information, to effectively address the challenges posed by false information. In some cases, attempts to refute false information can backfire and reinforce people's beliefs in the false information, leading to further entrenchment *(33)*.



Example: Using the refutation technique to debunk a piece of false information:

False information: "The COVID-19 vaccine contains a microchip that the government will use to track your movements."

Refutation technique response:

- Identify the false information: The claim that the COVID-19 vaccine contains a microchip that will be used to track people is false.
- Present evidence: There is no evidence to support this claim. The vaccines have undergone rigorous testing and have been shown to be safe and effective at preventing COVID-19.
- Explain the evidence: The COVID-19 vaccines do not contain any tracking devices or microchips. They work by teaching the body how to recognize and fight the virus that causes COVID-19.
- Provide alternative information: The vaccines have been authorized for emergency use by the United States Food and Drug Administration, and have been administered to millions of people with few serious side effects.
- Repeat and reinforce: It is important to get accurate information about the COVID-19 vaccines from reliable sources, such as WHO or your health-care provider. Vaccines are a safe and effective way to protect yourself and others from COVID-19.

Prebunking as a proactive response technique

What is prebunking?

Prebunking is a proactive response approach to address rumours, false information and disinformation by preemptively providing accurate information to the public before the false information spreads, and equipping individuals with the skills needed to identify inaccurate information.

The goal of prebunking is to inoculate individuals against false information by providing them with accurate information before they are exposed to falsehoods, as well as by making them aware of the tactics used to spread false information.

This approach is similar to a vaccine which works by providing the body with a small, harmless amount of an antigen to build immunity against disease. In both cases, the goal is to prevent harm by preparing the individual's body or mind to recognize and resist harmful agents before they encounter them in real-life scenarios. Prebunking helps to "immunize" individuals against false information by providing them with the cognitive tools and knowledge needed to identify and reject false information.

Behavioural psychology explains why prebunking works

Prebunking is built on inoculation theory, which was developed in the 1960s by social psychologists, and it aims to train people to recognize tactics used to manipulate information, much like vaccines train the immune response against a virus. Research has shown that prebunking can be more effective than debunking in reducing the belief in and spread of misinformation *(34)*.

The key aspects of prebunking are described below.

1. **Inoculation:** Prebunking works by inoculating people against false information. By exposing people to small doses of false information and then immediately providing them with accurate information, prebunking can help people build up resistance to future false information on the same topic *(34–36)*.
2. **Social proof:** Social proof is a psychological phenomenon where people assume the actions of others to reflect correct behaviour for a given situation *(37)*. Just as vaccines work by building herd immunity in populations, prebunking helps to build a community of individuals who are better equipped to recognize and resist false information. This can ultimately help to reduce the overall spread and impact of harmful false information. Prebunking uses social proof to nudge people towards more accurate beliefs and behaviours. This means that by providing accurate information that is supported by social norms and trusted sources, prebunking can encourage people to adopt more accurate beliefs and behaviours *(38)*.
3. **Self-affirmation:** Prebunking works by using self-affirmation techniques to build resistance against persuasion. By providing people with opportunities to affirm their values and identity before presenting them with information that challenges their beliefs, prebunking can help people feel less defensive and more open to considering new information.
    - For example, a RCCE-IM campaign might ask individuals to reflect on a personal value that is important to them, such as their family structure or their health decisions, and then ask them to make a short video or a social media post about why that value is important to them. This activity could be completed before individuals are exposed to false information about a health topic, such as the safety of a particular medication.
    - Research suggests that this type of self-affirmation activity can increase individuals' confidence in their own values and beliefs, which in turn can make them more resistant to the influence of false information. When people are confident in their own values and beliefs, they are less likely to be swayed by false information that conflicts with those values *(39,40)*.

4. **Memory bias (misinformation effect):** Prebunking helps counteract memory bias, which is the tendency to remember false information even after it has been corrected *(41)*. By providing accurate information before false information is encountered, prebunking can help people remember accurate information instead of false information *(42).*
5. **Trust enhancement:** Prebunking enhances trust between the public and the authoritative sources of information. By providing accurate information in a proactive and contextualized way, prebunking can demonstrate the credibility of the authoritative sources and enhance trust *(36)*.

Prebunking techniques used in the public health context

Prebunking can be used in the public health context to inoculate against harmful health information as in the examples listed below.

- Researchers have used brief inoculation videos to train people in the detection of flawed arguments, as an example of passive inoculation *(43).* The videos expose participants to a single misleading technique, providing both a forewarning and explanation of the manipulation technique. This exposure helps to enhance the participants' ability to detect and resist false information.
- Games can be an effective technique to inoculate against false claims.
    - The Social Decision-Making Lab at the University of Cambridge supported by the World Health Organization built the game GoViral! *(44).* When a player enters the game, they are encouraged to "walk a mile in the shoes of a manipulator to get to know their tactics from the inside" and "see it as ruining the magician's trick so that we don't fall for it next time around." In this simulation exercise, players learn how filter bubbles create echo chambers of false information and to manipulate negative emotions to stoke outrage and build influence. This method of exposing players to false information is also referred to as active inoculation. Unlike passive inoculation, where individuals are directly informed why the information is false, active inoculation requires them to learn by actively constructing the misinformation themselves in a controlled environment.
    - A similar game called Bad News *(45)* educates players on six prevalent tactics used in spreading fake news:
        1. **Impersonation:** Pretending to be someone else or representing a group to make the information seem more credible.
        2. **Polarization:** Exploiting political divisions to create a wider gap between groups.
        3. **Emotional Language:** Using excessive emotional words to twist the original news, provoking intense feelings.
        4. **Conspiracy Creation:** Crafting or encouraging conspiracy theories to interpret recent happenings.
        5. **Trolling:** Targeting users, celebrities or organizations to give the illusion of widespread agreement or disagreement with a statement.
        6. **Discrediting:** Attacking the credibility of individuals, institutions or well-accepted truths to sow doubt among the audience.

A comparison between debunking and prebunking

Debunking is a reactive approach for addressing rumours and disinformation by correcting false information after it has already been disseminated. In contrast, prebunking is preventive and proactive and aims to prevent the impact of false information (Table 6).

Table 6. Comparing debunking and prebunking

Debunking	Prebunking
Reactive response	Proactive response
Identify false information signals	Identify **potential** for false information to spread
Refute false information through response/ message development	**Anticipate false information** and prepare a response
Communicate the refutation to the audience	Communicate **accurate information** before false information spreads
Occurs after the false information has spread	Occurs **before** the false information has a chance to spread

Leverage community engagement for response design

Community engagement helps shape contextually relevant responses by capturing insights into local knowledge, attitudes and perceptions that influence behaviour. These are two ways through which community engagement supports response in its design phase:

- **Providing community insights:**
  Community engagement creates an essential feedback loop between the designers of the response and the community. Obtaining direct insights from the community can support the design of long-term interventions. CSOs and community actors know the intended audiences, their attitudes, practices and beliefs and can be a valuable source of insights to design the most effective response based on these.
- **Co-designing message and interventions:** CSOs and community actors should be involved in the development and testing of messages to ensure appropriateness and understanding. Involving them in the design of interventions can ensure that they resonate more with target audiences and are more effective to achieve desired outcomes.

Outreach empowers individuals to make informed decisions to protect their health. This is achieved by refuting false information and promoting accurate information and advice.

Phase 5: Outreach

Description

Outreach is the final phase in the process, where the response plan is implemented and key messages are targeted to the intended audiences. The goal of this phase is to engage the audience and promote behaviour change that supports the intended public health outcomes.

The outreach phase should also link back to the first phase of signal detection through a feedback loop. Once a response intervention is implemented, it is important to continue monitoring the situation to ensure that it is reaching the intended audiences and has the desired impact. The feedback loop can also provide insight as to how messages are being perceived by the target audiences, allowing for continuous improvement of the response messages and the overall IM process.

Furthermore, the feedback loop can also provide valuable information to the signal detection phase as any new signals or emerging issues can be detected early on and incorporated into the IM process, enabling the process to be more proactive and responsive to emerging issues.

Key steps in the outreach phase

The key activities in this phase are summarized in Table 7.

Table 7. Key outreach phase activities

Activity	Description
Disseminating messages	Disseminate the messages through the most effective channels to reach the intended audience(s), such as social media platforms, email newsletters, websites, leaflets and other online and offline communication channels.
Amplifying messages	Use a variety of strategies to amplify the messages and increase their reach, such as partnering with influencers, engaging with online and offline communities, or paid promotion of social media posts.
Monitoring feedback	Monitor feedback from the target audience(s) (how they receive and perceive the messages) and adjust messaging as needed based on their response and engagement.
Evaluating impact	Evaluate the impact of the messaging on the target audience(s) (whether they accept and uptake advice) and make adjustments to the outreach strategy as needed to improve effectiveness.
Maintaining engagement	Continue to engage with the target audience(s) over time, building trust and establishing a relationship that supports ongoing communication and engagement.

Outreach case study

This case study describes how Ireland's Health Service Executive (HSE) used social media for two-way communication, social listening and countering false information *(46).*

Context: When the COVID-19 pandemic hit, Ireland's HSE started mapping the information needs of people contacting their call centre and used this to develop a script, which answered the most frequently asked questions about COVID-19.

The social media team was also receiving a large number of information requests, most of them as direct messages via HSE's Twitter and Instagram accounts. The HSE social media team collaborated with technical experts on answering these questions.

Muiriosa Ryan, Social Media Manager at HSE remembered:

> When the government was going to announce a new initiative on COVID-19 testing or a change in the travel rules, we knew the public were going to have a lot of questions. HSE's call centre, content and social media teams worked together with HSE's public health experts to keep the common talking points and FAQ [further answer questions] document on COVID-19 and the information on HSE's website up to date and relevant. Answering questions on COVID-19 from 07:00 until 22:00, seven days a week, became a routine task for the social media team.

Surge resources

In 2019, HSE's social media team consisted of four staff: one manager, two executives and an assistant. In 2020, the team was assigned three extra staff bringing the total to seven.

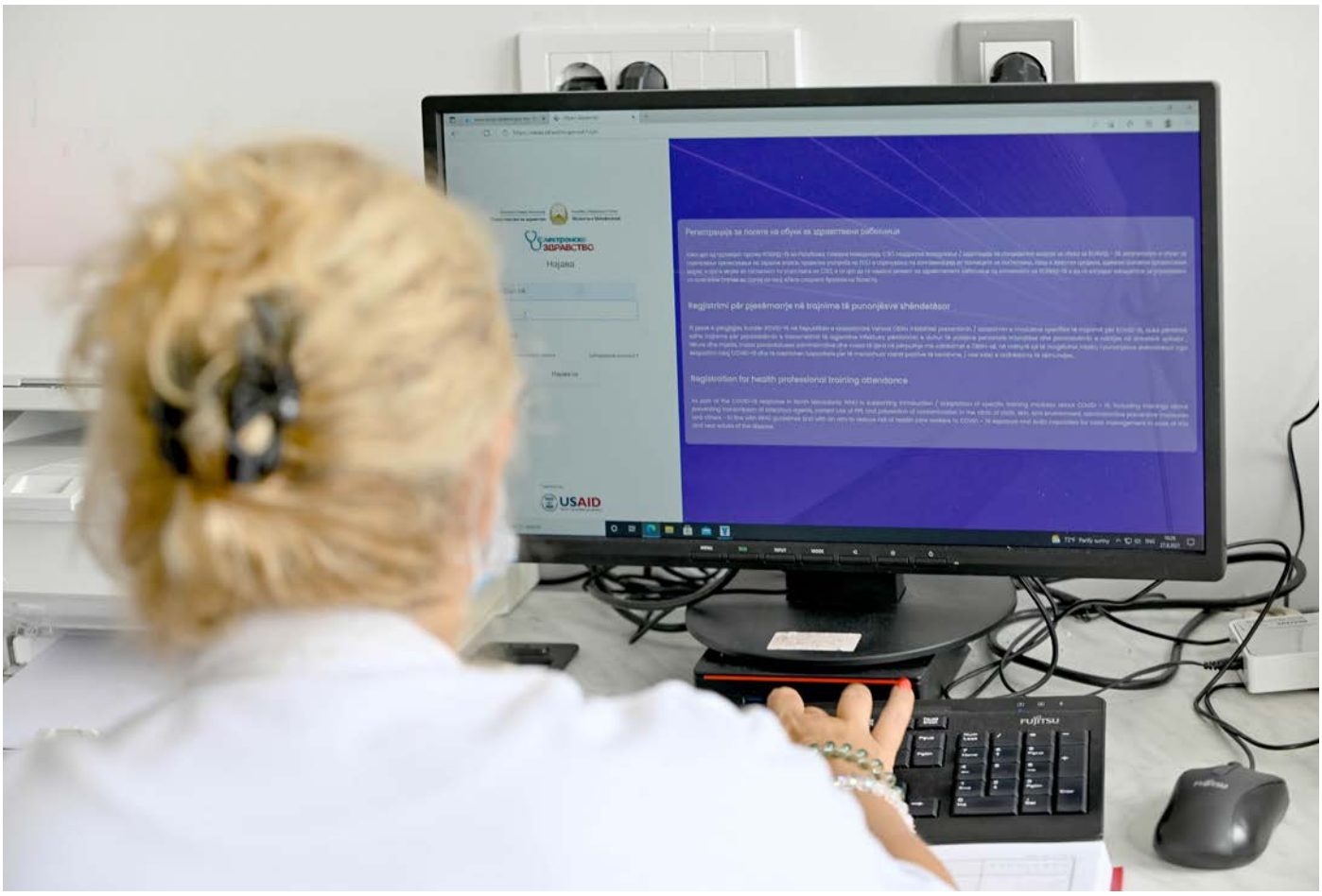
HSE's social media team counters false information

Countering online misinformation from anti-vaccine campaigners and their allies immediately became a high priority for the HSE social media team. The Social Media Manager Muiriosa Ryan also stated that, "Twitter put a button on its site for all users in Ireland linking to HSE's vaccine website to make reliable information more accessible," and "They [the social media companies] have generally been pretty good at taking down misinformation when we report it. Content that HSE reports gets fast tracked for action seven days a week. Our biggest challenge is finding the time to keep up with all the misinformation being posted".

Leverage community engagement for outreach

Engaging communities in the outreach phase bridges the gaps between health institutions and the public. These are some ways in which community engagement can support the outreach process:

- **Supporting message deployment:** Nongovernmental entities often hold a high level of trust and respect within the communities they serve. Trusted influencers, CSOs and community actors can encourage target audiences to refute false information and accept and uptake accurate advice.
- **Selecting communication channels:** It is vital to involve CSOs and community actors to identify the most suitable channels to reach our target audiences. CSOs and other community groups often have their own newsletters, websites and other online and offline communication channels that may be used to reach target audiences. Furthermore, CSOs can support offline responses for example through community sessions focusing on topics related to a specific false narrative or by organizing sessions with public health experts and community members.
- **Providing feedback on outreach:** CSOs and community partners are best positioned to get feedback on how messages are perceived and their influence on behavioural change. Therefore, it is recommended to work with CSOs to track feedback from message recipients and to help shape iterations and follow-up.
- **Build back better together:** Involve partners and stakeholders in lessons learned and "building back better" efforts. Intra- and after-action reviews with communities are essential to identify resource gaps, the most effective measures, challenges and recommendations to strengthen future responses.



References[2]

1. Wardle C, Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe; 2017 (https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c).

2. Advancing infodemic management in risk communication and community engagement in the WHO European Region: implementation guidance. Copenhagen: WHO Regional Office for Europe; 2022 (https://www.who.int/europe/publications/i/item/WHO-EURO-2022-5842-45607-65433).

3. Lohiniva AL, Sibenberg K, Austero S, Skogberg N. Social Listening to Enhance Access to Appropriate Pandemic Information Among Culturally Diverse Populations: Case Study From Finland. JMIR Infodemiology. 2022;2(2):e38343. doi: 10.2196/38343.

4. Policy guide on children and digital connectivity. New York: United Nations Children's Fund; 2018 (https://www.unicef.org/esa/media/3141/file/PolicyLab-Guide-DigitalConnectivity-Nov.6.18-lowres.pdf).

5. Lotto M, Hanjahanja-Phiri T, Padalko H, Oetomo A, Butt ZA, Boger J et al. Ethical principles for infodemiology and infoveillance studies concerning infodemic management on social media. Front Public Health. 2023 Mar 23;11:1130079. doi: 10.3389/fpubh.2023.1130079.

6. Roozenbeek J, Zollo F. Democratize social-media research - with access and funding. Nature. 2022 Dec;612(7940):404. doi: 10.1038/d41586-022-04407-8.

7. Purnat TD, Nguyen T, and Briand S (editors). Managing Infodemics in the 21st Century: Addressing New Public Health Challenges in the Information Ecosystem. Cham: Springer Cham; 2023.

8. Google alerts. Mountain View: Google; 2023 (https://www.google.com/alerts).

9. Google trends. Mountain View: Google; 2023 (https://trends.google.com/trends/).

10. Followerwonk. Seattle: Followerwonk; 2023 (https://followerwonk.com/).

11. Hoaxy2 beta. Bloomington: Indiana university Observatory on social media; 2023 (https://hoaxy.osome.iu.edu/).

12. Bot Sentinel. Hasbrouck Heights: Bot Sentinel; 2023 (https://botsentinel.com/).

13. Talkwalker. Luxembourg: Talkwalker; 2023 (https://www.talkwalker.com/social-media-analytics-search).

14. Tools That Fight Disinformation Online. Santa Monica: RAND; 2023 (https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html).

15. WHO releases a public health taxonomy for social listening on monkeypox conversations. In: World Health Organization [website]. Geneva: World Health Organization; 2022 (https://www.who.int/news/item/26-09-2022-who-releases-a-public-health-taxonomy-for-social-listening-on-monkeypox-conversations).

16. Fact-checking. Florence: European digital media observatory (https://edmo.eu/fact-checking/).

2 All online weblinks were accessed 8 August 2023

17. Vaccine Misinformation Management Field Guide. New York: United Nations Children's Fund; 2020 (https://vaccinemisinformation.guide/).

18. Kahneman D, Sibony O, Sunstein CR. Noise: a flaw in human judgment. Boston: Hachette Book Group; 2021.

19. Coronavirus: Outcry after Trump suggests injecting disinfectant as treatment. London; British British Broadcasting Corporation; 2020 (https://www.bbc.com/news/world-us-canada-52407177).

20. WHO/UNICEF How to build an infodemic insights report in 6 steps. Geneva: World Health Organization; 2023 (https://www.who.int/publications/i/item/9789240075658).

21. Rumour Tracker Programme: A community based approach to address information gaps and misinformation on COVID-19. Geneva: World Health Organization; 2022 (https://cdn.who.int/media/docs/default-source/science-translation/case-studies-1/cs13_rumourtracking.pdf?sfvrsn=829a4b42_4)

22. Technical note from the WHO Technical Advisory Group on behavioural insights and science for health. Geneva: World Health Organization; 2021 (https://www.who.int/publications/m/item/technical-note-from-the-who-technical-advisory-group-on-behavioural-insights-and-science-for-health).

23. Coronavirus disease (COVID-19) advice for the public: Mythbusters. In: World Health Organization [website]. Geneva: World Health Organization; 2022 (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters).

24. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psychol Sci. 2017;28(11):1531-1546. doi: 10.1177/0956797617714579.

25. VERA FILES FACT CHECK: Viral advisory against drinking cold water amid heat NOT TRUE. Quezon City: Vera Files; 2021 (https://verafiles.org/articles/vera-files-fact-check-viral-advisory-against-drinking- cold-w).

26. Nyhan B. Why the backfire effect does not explain the durability of political misperceptions. Proc Natl Acad Sci U S A. 2021;118(15):e1912440117. doi: 10.1073/pnas.1912440117.

27. Nourbakhsh A, Liu X, Li Q, Shah S. Mapping the echo-chamber: detecting and characterizing partisan networks on Twitter [conference paper]. Proceedings of the 2017 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation; 2017 (http://sbp-brims.org/2017/proceedings/papers/challenge_papers/MappingTheEcho-Chamber.pdf)

28. Swire-Thompson B, Miklaucic N, Wihbey JP, Lazer D, DeGutis J. The backfire effect after correcting misinformation is strongly associated with reliability. J Exp Psychol Gen. 2022 Jul;151(7):1655-1665. doi: 10.1037/xge0001131.

29. Albarracín D, Albarracín J, Chan MS, Hall Jamieson K. Creating Conspiracy Beliefs: How Our Thoughts Are Shaped. Cambridge: Cambridge University Press; 2022.

30. Soldá A. Overconfidence as an interpersonal strategy [PhD thesis]. Brisbane: Queensland University of Technology; 2020. doi: 10.5204/thesis.eprints.135191.

31. Amazeen M. The Debunking Handbook 2020. Fairfax: George Mason University Center for Climate Change Communication; 2020 (https://www.bu.edu/com/research/the-debunking-handbook-2020/).

32. Conger K. How misinformation, medical mistrust fuel vaccine hesitancy. In: Stanford Medicine [website]. Stanford: Standford Medicine; 2021 (https://med.stanford.edu/news/all-news/2021/09/infodemic-covid-19.html).

33. MacFarlane D, Tay LQ, Hurlstone MJ, Ecker UKH. Refuting Spurious COVID-19 Treatment Claims Reduces Demand and Misinformation Sharing. J Appl Res Mem Cogn. 2021;10(2):248-258. doi: 10.1016/j.jarmac.2020.12.005.

34. Roozenbeek J, Van der Linden S, Nygren T. Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. Cambridge (MA): Harvard Kennedy School; 2023 (https://misinforeview.hks.harvard.edu/article/global-vaccination-badnews/).

35. Garcia L, Shane T. A guide to prebunking: a promising way to inoculate against misinformation. New York: First Draft; 2021 (https://firstdraftnews.org/articles/a-guide-to-prebunking-a-promising-way-to-inoculate-against-misinformation/).

36. Harjani T, Roozenbeek J, Biddlestone M, van der Linden S, Stuart A, Iwahara M et al. A Practical Guide to Prebunking Misinformation. United Kongdom: University of Cambridge, BBC Media Action, Jigsaw; 2022 (https://interventions.withgoogle.com/static/pdf/A_Practical_Guide_to_Prebunking_Misinformation.pdf).

37. Cialdini RB. Influence: Science and practice. Vol. 4. Boston: Pearson education; 2009.

38. The gentle science of persuasion, part three: Social proof. Tempe: Arizona State University; 2007 (https://news.wpcarey.asu.edu/20070103-gentle-science-persuasion-part-three-social-proof).

39. Iles IA, Gillman AS, Platter HN, Ferrer RA, Klein WMP. Investigating the Potential of Inoculation Messages and Self-Affirmation in Reducing the Effects of Health Misinformation. Sci Commun. 2021;43:6. doi: 10.1177/10755470211048.

40. Carnahan D, Hao Q, Jiang X, Lee H. Feeling fine about being wrong: The influence of self-affirmation on the effectiveness of corrective information. Hum Commun Res. 2018;44(3): 274–298. doi: 10.1093/hcr/hqy001.

41. Loftus EF. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. Learn Mem. 2005 Jul-Aug;12(4):361-6. doi: 10.1101/lm.94705.

42. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio L, Brashier N et al. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol. 2022;1:13–29. doi: 10.1038/s44159-021-00006-y.

43. Lewandowsky S, van der Linden S. Countering Misinformation and Fake News Through Inoculation and Prebunking. Eur Rev Soc Psychool. 2021;1–38. doi:10.1080/10463283.2021.1876983.

44. Social decision-making lab at the University of Cambridge, Drog, Tilt, Gusmanson, United Kingdom cabinet office. GoViral! [online game]. Cambridge: University of Cambridge; 2023 (https://www.goviralgame.com).

45. Social decision-making lab at the University of Cambridge, Gusmanson. Bad News [online game]. Cambridge: University of Cambridge; 2023 (https://www.getbadnews.com).

46. Risk communication and community engagement: a compendium of case studies in times of COVID-19. Copenhagen: WHO Regional Office for Europe; 2022 (https://apps.who.int/iris/handle/10665/363343).

Further reading

1. Wang S, Pang MS, Pavlou P. Cure or Poison? Identity Verification and the Posting of Fake News on Social Media. J Manag Inf Syst. 2021;38:1011–1038. Doi: 10.1080/07421222.2021.1990615.
2. Kolluri NL, Murthy D. CoVerifi: A COVID-19 news verification system. Online Soc Netw Media. 2021;22:100123. doi: 10.1016/j.osnem.2021.100123.
3. Tschiatschek S, Singla A, Rodriguez M, Merchant A, Krause A. Fake News Detection in Social Networks via Crowd Signals. WWW '18: Companion Proceedings of the The Web Conference 2018. 2018;517-524. doi: 10.1145/3184558.3188722.
4. Torres R, Gerhart N, Negahban A. Combating fake news: An investigation of information verification behaviors on social networking sites [conference paper]. Hawaii International Conference on System Sciences. 2018. doi: 10.24251/HICSS.2018.499.
5. Ullrich EKH, Lewandowsky S, Cook J, Schmid P, Fazio L, Brashier N et al. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psych. 2022;1(1): 13–29.
6. Van der Linden S. Foolproof: why we fall for false information and how to build immunity. New York: Harper Colins; 2023.
7. van der Linden S. Misinformation: susceptibility, spread, and interventions to immunize the public. Nat Med. 2022;28(3):460-467. doi: 10.1038/s41591-022-01713-6.
8. Young K, Hyunji L. Debunking misinformation in times of crisis: Exploring misinformation correction strategies for effective internal crisis communication. J Contingencies Crisis Manag. 2022;31. doi: 10.1111/1468-5973.12447
9. Mourali M, Drake C. The Challenge of Debunking Health Misinformation in Dynamic Social Media Conversations: Online Randomized Study of Public Masking During COVID-19. J Med Internet Res. 2022 Mar 2;24(3):e34831. doi: 10.2196/34831.
10. Whitehead HS, French CE, Caldwell DM, Letley L, Mounier-Jack S. A systematic review of communication interventions for countering vaccine misinformation. Vaccine. 2023;41(5):1018-1034. doi: 10.1016/j.vaccine.2022.12.059.
11. How 'prebunking' can fight fast-moving vaccine lies. In: PBS News Hour [website]. Washington DC: PBS; 2021 (https://www.pbs.org/newshour/health/how-prebunking-can-fight-fast-moving-vaccine-lies)
12. University of Cambridge. Social media experiment reveals potential to 'inoculate' millions of users against misinformation. Rockville: ScienceDaily; 2022 (www.sciencedaily.com/releases/2022/08/220824152220.htm).
13. Bond S. False information is everywhere. 'Pre-bunking' tries to head it off early. In: npr [website]. Washington DC: npr; 2022 (https://www.npr.org/2022/10/28/1132021770/false-information-is-everywhere-pre-bunking-tries-to-head-it-off-early).
14. Google to Expand False information 'Prebunking' in Europe. In: VOA [website]. Washington DC: VOA; 2023 (https://www.voanews.com/a/google-to-expand-misinformation-prebunking-in-europe/6960557.html).

The World Health Organization (WHO) is a specialized agency of the United Nations created in 1948 with the primary responsibility for international health matters and public health. The WHO Regional Office for Europe is one of six regional offices throughout the world, each with its own programme geared to the particular health conditions of the countries it serves.

Member States

Albania
Andorra
Armenia
Austria
Azerbaijan
Belarus
Belgium
Bosnia and Herzegovina
Bulgaria
Croatia
Cyprus
Czechia
Denmark
Estonia
Finland
France
Georgia
Germany
Greece
Hungary
Iceland
Ireland
Israel
Italy
Kazakhstan
Kyrgyzstan
Latvia
Lithuania
Luxembourg
Malta
Monaco
Montenegro
Netherlands (Kingdom of the)
North Macedonia
Norway
Poland
Portugal
Republic of Moldova
Romania
Russian Federation
San Marino
Serbia
Slovakia
Slovenia
Spain
Sweden
Switzerland
Tajikistan
Türkiye
Turkmenistan
Ukraine
United Kingdom
Uzbekistan

World Health Organization
Regional Office for Europe
UN City, Marmorvej 51,
DK-2100 Copenhagen Ø,
Denmark

TEL +45 45 33 70 00
FAX +45 45 33 70 01
EMAIL eurocontact@who.int
WEB www.who.int/europe

WHO/EURO:2024-8271-48043-71198