

RESEARCH

Open Access



Unveiling the narrative: Italian media's role in shaping public health messaging during COVID-19 - A corpus-assisted discourse analysis

Carlotta Fiammenghi¹, Elisabetta Ceretti², Loredana Covolo², Nicola Pelizzari^{2*}, Laura Brunelli³, Anna Mirella Taranto⁴, Luana Penna⁴, Silvio Brusaferrero^{3,4} and Umberto Gelatti²

Abstract

Background Effective public health communication during emergencies is critical for influencing behaviour and maintaining trust. Newspapers remain a key channel for disseminating institutional guidance, yet the extent to which their coverage aligns with international recommendations (e.g., World Health Organisation, WHO and Pan American Health Organisation, PAHO guidance) is not well established. Italy, the first European country to experience major COVID-19 outbreaks, provides a useful case for assessing the extent to which mainstream newspaper coverage aligned with WHO and PAHO recommendations for COVID-19 reporting.

Methods A quali-quantitative, corpus-assisted discourse analysis was conducted on 5,621 articles (over 2.7 million words) published in seven major Italian newspapers and included in the national press review of the Italian National Institute of Health. Lexical and discursive patterns were examined using SketchEngine, combining quantitative keyword, collocation, and concordance analyses with qualitative close reading. Findings were compared against reporting recommendations issued by the WHO and the PAHO.

Results Italian newspapers largely reflected key elements of WHO/PAHO guidance, emphasizing preventive measures, vaccine-related content, and reliance on expert and institutional sources (e.g., scientists, researchers, and public health professionals). However, recurrent use of non-people-first labels for affected individuals and occasional instances of "false balance" in vaccine coverage were also identified.

Conclusions Mixed corpus-based and qualitative approaches can support systematic assessment of newspaper reporting against WHO and PAHO recommendations for COVID-19 coverage. While many aspects of guideline-consistent reporting were observed, gaps remain in the use of non-stigmatising, people-first language and in evidence-weighted framing of scientific disagreement. Strengthening collaboration between journalists and public health institutions may support clear, ethical, and trustworthy communication in future emergencies.

Keywords Infodemics, Corpus-assisted discourse analysis, COVID-19, Media reporting

*Correspondence:

Nicola Pelizzari
nicola.pelizzari@unibs.it

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

The role of the traditional newspaper press in shaping public opinion and fostering informed citizenship has been a subject of considerable academic interest. For example, it has been shown that fake or misleading information could significantly distort public understanding and influence decision-making processes [1, 2]; and recently it has been confirmed that misleading information may be more pervasive than previously thought, and therefore harder to counteract, especially in mainstream print news reporting [2]. Indeed, despite the rise of digital media, recent research has underscored the continued relevance of traditional journalism in the news landscape: although there was a surge in online news use during the 2020 COVID-19 pandemic in the USA, the UK, France, and Germany, traditional news outlets (including established newspapers that publish both in print and online) accounted for most of this increase, as people kept turning to reliable mainstream news sources even in times of crises [3]; and in Italy, traditional media were found to be the most frequently used communication channels to inform the public about COVID-19 vaccines [4].

The importance of newspaper reporting, particularly during public health crises, was promptly understood by leading public health organisations worldwide, such as the World Health Organisation (WHO) and the Pan American Health Organisation (PAHO), which published guidelines and advice for journalists at the very onset of the COVID-19 pandemic [5, 6]. These guidelines covered the content and style of newspaper articles dealing with the pandemic, with precise recommendations as to the lexical words to be preferred when describing pandemic trends and preventive measures. They also described the attitudes towards controversies and rumours that journalists should adopt, as well as the relationships they should establish with sources, encouraging them to rely mostly on national, state, local and governmental authorities. Finally, both the WHO and the PAHO acknowledged that health authorities were increasingly required to deal with journalists as part of their daily job tasks.

Despite the publication of these internationally recognised guidelines, the extent to which mainstream newspaper coverage in different countries actually complies with WHO and PAHO recommendations remains understudied. Recent research employing corpus-based discourse analysis has demonstrated the value of systematically examining media coverage of public health emergencies. Studies have analysed COVID-19 newspaper discourse across diverse national contexts [7–9], including China, the United Kingdom, Brazil, and the United States, using corpus-assisted approaches that combine quantitative analysis of linguistic patterns with qualitative close reading to reveal how lexical choices, framing strategies, and source selection shape public understanding

of scientific uncertainty, institutional authority, and preventive measures, and to identify both overt and subtle discursive features that may influence public health outcomes [10, 11].

In the Italian context, a few studies have examined newspaper coverage of COVID-19, focusing on narrative framing, institutional discourse, and the representation of scientific expertise [12–14]. However, systematic evaluation of Italian newspaper compliance with specific international public health communication guidelines such as those issued by WHO and PAHO remains an underexplored area. Understanding this gap is important because public health institutions also depend on media to disseminate authoritative information, and research on organisational health literacy emphasises that effective health communication requires institutional capacity to engage productively with media professionals through evidence-based assessment of their communication practices [15].

Unlike communication by individuals, which can vary widely in accuracy and impact, institutional communication is typically characterised by a structured and authoritative dissemination of information, aiming to ensure the consistent delivery of evidence-based guidance, foster public trust, and influence health behaviours on a broad scale. Government health agencies and international organisations have the capacity to mobilise resources, reach diverse populations, and provide reliable information; by establishing clear, coherent, and credible communication, these institutions help to mitigate misinformation and disinformation, reduce public uncertainty, and enhance the effectiveness of public health interventions [16, 17]. However, it should be noted that the communication of such messages does not always occur via official channels. Indeed, it is frequently through interactions with the press that public health institutions can reach the widest possible audience. Consequently, journalists become the intermediaries who disseminate the message entrusted to them by the institutions. It is therefore essential to comprehend the mechanisms that govern the processes and alterations that this message may undergo during such a transition. Similarly, it is essential that public health officials are (made) aware of the potential power of their own words in shaping and directing public sentiment, establishing relations and trust, conscious that each uttered word matters and brings with it a wealth of possible meanings.

Italy was the first European country hit by the COVID-19 pandemic [18] and as such, it could be considered a testbed for its handling by the media; the relevance of such early reporting was underscored by the PAHO's advice to "*find examples from media coverage in other countries where the epidemic is more advanced and other kinds of measures have been implemented*" [5].

The present study is situated within a larger project analysing the press review of the Italian National Institute of Health (Istituto Superiore di Sanità, ISS), the leading technical-scientific body of the Italian National Health Service, during the COVID-19 pandemic [19].

In this project, the same corpus of 5,621 articles has already been used to examine how COVID-19 vaccines were portrayed in the press and to explore instances of subtle misinformation and false balance in vaccine coverage. The present article builds on that work by moving from the specific case of vaccination to a broader assessment of how Italian newspaper coverage included in the ISS press review aligns with WHO and PAHO guidelines on COVID-19 reporting, with particular reference to language choices, the description of affected individuals, the emphasis placed on preventive measures, and the role of expert sources. Accordingly, this study addresses the following research question: to what extent did newspaper articles included in the ISS press review comply with WHO and PAHO recommendations regarding (RQ1) the naming of the disease and the description of affected individuals, (RQ2) the avoidance of stigmatising or hyperbolic language, (RQ3) the emphasis on preventive measures, and (RQ4) the reliance on trusted expert sources?

To this end, a corpus-assisted, quali-quantitative discourse analysis [20] is employed to compare lexical and discursive patterns in the corpus with the specific “dos and don’ts” formulated in WHO and PAHO guidelines. Using the ISS press review as a sampling frame, the study examines how authoritative Italian newspapers mediated institutional public health communication during the first two years of the pandemic and provides empirical insights that can support reflection on future collaboration between public health institutions and the media.

Materials and methods

To answer our research questions, we first compiled a corpus of newspaper articles, then applied quantitative corpus-linguistic analyses to identify patterns, and finally performed qualitative analyses to interpret the meaning of terms in context. Each of these steps is outlined in more detail in the ensuing subsections.

Corpus compilation

We compiled a corpus of newspaper articles included in the national press review of the Italian National Institute of Health and printed between 1st January 2020 and 31st December 2021 in a selection of seven Italian national newspapers. This selection was balanced between circulation and publisher, to include the top-circulating daily newspapers owned by six different publishers; the full list is available in the Appendix. The articles were retrieved in .txt format through the database Factiva (a

research tool owned by Dow Jones & Company providing access to newspaper archives). The data collection phase was 1st September 2022–31st December 2022. The texts were uploaded to the SketchEngine [21], where they were tagged for part-of-speech and then analysed following a corpus-assisted discourse analysis approach [20], retrieving wordlists, keyword lists, collocations, and concordances. The corpus has been designed to be representative of general trends, and analyses focus on these trends across the entire corpus. Consequently, outlet-specific comparisons have been intentionally excluded from the analysis.

Quantitative analysis: wordlists, keywords, and collocations

Wordlists calculate the most frequent terms in the corpus in their lemma form, word form, or in fixed combinations of two or more elements (n-grams). Wordlists can be compiled based on raw frequencies, showing how many times an element appears in the corpus, or normalised frequencies, necessary to compare corpora or sub-corpora of different sizes and obtained by dividing the number of occurrences of an element within the corpus by the total number of words in the corpus and then multiplying the result by a common factor (set at 1000 in the present study, given the corpus size). A list of exclusion of function words (determiners, pronouns, conjunctions, prepositions, and auxiliary verbs) was uploaded to the corpus so as to eliminate them from the final wordlist, as they are frequent elements in any language, not representative of any particular discourse. The exclusion list can be found in the Appendix.

Keyword lists consist of the most frequently used elements in a corpus compared to a reference corpus and are used to understand the distinguishing features of a text or discourse. The reference corpus used for the study was ItTenTen20, which includes about 12 billion words from Italian texts from the web, updated to December 2020, and available on SketchEngine. Keyword lists are generated by the software based on a keyness score, calculated through the formula: $(\text{fpm focus} + N) / (\text{fpm ref} + N)$, where fpmw = frequency per million words, or normalised frequency; N was set at 100 in the present study, so as to exclude extremely rare elements in the corpus.

Collocations are defined as the words or lemmas which occur more frequently than could be expected by chance alone in a words’ immediate surroundings and help to define its positive or negative connotations [22]. We identified the relevant collocations using SketchEngine’s default combination of T-score, Mutual Information (MI) and LogDice, instrumental in the selection of meaningful, but not too rare, elements. The interval was also kept by default, three positions to the left and three positions

to the right of the node, in order to obtain collocations used in the word’s immediate co-text, accounting for the mobility of the elements in the Italian sentence. The WordSketch function on SketchEngine can be used to group collocations according to their logical-grammatical function. Overlapping meanings between two or more words can also be explored using the Thesaurus function, which automatically generates a list of words that belong to the same semantic category and have similar collocations by calculating a similarity score based on the percentage of shared collocates. Further information on statistics in SketchEngine are retrievable here: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf> (last accessed 12th January 2026).

Qualitative analysis: concordances and close reading

Concordances show each occurrence of a given linguistic element within its immediate co-text and can be expanded and annotated. Close reading of concordances allows researchers to explore and clarify how words are used and what meanings they acquire in context. When combined with frequency and keyness measures, concordance analysis enables a systematic linkage between frequency and meaning of linguistic elements in a text, and therefore enabled us to address: (RQ1) the presence and contextual use of language used to name and describe the disease and affected individuals; (RQ2) the presence and contextual use of hyperbolic or potentially stigmatising language; (RQ3) the overt mention of preventive measures and their discursive framing; and (RQ4) the presence and framing of expert sources in newspaper coverage.

The latter aspect (RQ4) required particular attention to instances of false balance, defined as the journalistic practice of presenting opposing perspectives as epistemologically equivalent even when only one is supported by scientific consensus. False balance is difficult to detect

through lexical analysis alone, as it does not hinge on the mere presence of divergent viewpoints but rather on the relative epistemic weight attributed to them. For this reason, it becomes visible only when considering newspaper articles (and, in some cases, sequences of articles) as coherent discursive products. Accordingly, instances of false balance were identified through qualitative close reading, following the framework proposed by Boykoff and Boykoff [23] and subsequently applied to medical journalism by Clarke [24].

Figure 1 graphically illustrates the workflow from corpus compilation to quantitative and qualitative analysis.

Role of the funding source

Istituto Superiore di Sanità (ISS) provided the full corpus of newspaper articles used for the linguistic analysis.

Results

The 2020 corpus is made of 2859 articles and 1,646,927 words; the 2021 corpus is made of 2762 articles and 1,072,391 words. Tables showing composition of the sub-corpora are available in the Appendix.

The complete lists of the first 200 content words and the first 200 keywords in the 2020 and 2021 corpora can be found in the Appendix, with raw, normalised frequencies, and keyness score. Based on their core meaning, the most prominent words in the 2020 corpus can be divided into four main groups: infection trends (22% of words, 36.5% of keywords); the measures implemented to reduce contagion (11.5% of words, 13% of keywords); the people and institutions responsible for such implementation (7% of words, 22.5% of keywords); and the words needed to describe the new virus (6% of words, 15.5% of keywords). Similarly, the four most prominent semantic groups in the 2021 corpus are: infection trends (16% of words, 24.5% of keywords); the people and institutions responsible for public health management (7.5% of words, 11.5% of keywords); vaccination (5.5% of words, 13% of keywords); and the measures implemented to reduce contagion (4.5% of words, 9.5% of keywords).

As stated in the Introduction and in the Methods, corpus-assisted, quali-quantitative discourse analysis was used to gauge whether the newspaper articles included in the corpus complied with the guidelines and recommendations provided by WHO [6] and PAHO [5], with a specific focus on four main aspects: the name of the disease and the description of affected individuals (RQ1); the use or the avoidance of stigmatising and hyperbolic language (RQ2); the attention paid to preventive measures (RQ3); and the reliance on trusted expert sources (RQ4).

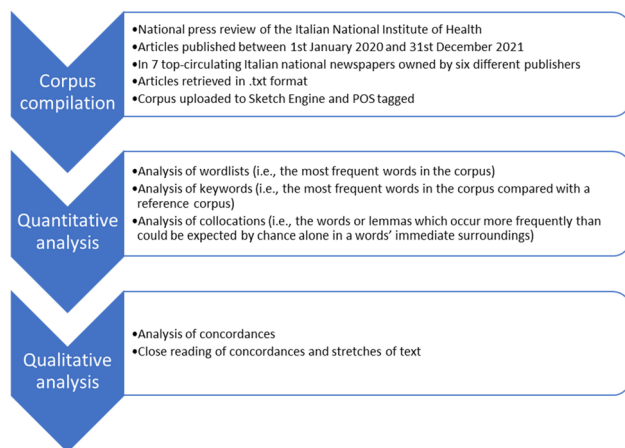


Fig. 1 - Flowchart showing the workflow from corpus compilation to quantitative and qualitative analysis

The name of the disease and the description of affected individuals (RQ1)

WHO recommendations on disease naming suggest talking about the new coronavirus disease (COVID-19) instead of attaching locations or ethnicity to the disease, by explicitly stating that “this is not a ‘Wuhan Virus,’ ‘Chinese Virus’ or ‘Asian Virus’” [6]. In line with these recommendations, the newspaper articles in the corpus predominantly referred to the new coronavirus disease using the terms “COVID” and “COVID-19”, rather than attaching geographic or ethnic labels. Quantitative corpus analysis shows that “COVID” and “COVID-19” are more frequent lemmas and stronger keywords than “Wuhan” or “cinese” (“Chinese”) in both years under study. In the 2020 corpus, “COVID” occurred 1768 times (normalised frequency 1.07; keyness score 9.3) and “COVID-19” 1393 times (normalised frequency 0.84; keyness score 7.2), whereas “Wuhan” and “cinese” displayed substantially lower keyness scores (3.4 and 2.9, respectively). In the 2021 corpus, “COVID” further increased in frequency (2792 occurrences; normalised frequency 2.60; keyness score 14.2), while “COVID-19” remained a keyword (keyness score 4.2), and neither “Wuhan” nor “Chinese” appeared as keywords. These quantitative patterns seem to indicate consistent adherence to recommended disease naming practices across the two years.

With regard to references to individuals affected by the disease, official guidelines recommend using people-first language to talk about “people who have COVID-19” and “people who are being treated for COVID-19” instead of referring to people with the disease as “COVID-19 cases” or “victims” [6 p. 3]. However, the corpus seems to show limited compliance with such guidelines: quantitative analysis reveals that the lemmas “caso” (“case”) and “vittima” (“victim”) are frequent and salient in both sub-corpora. In 2020, “caso” occurred 6450 times (normalised frequency 3.91), while “vittima” displayed a keyness score of 2.7. In 2021, “caso” remained highly frequent (6,418 occurrences; normalised frequency 5.98; keyness score 4.0), and “vittima” occurred 516 times (normalised frequency 0.31). These quantitative findings seem to indicate that non-people-first labels were widely used in the corpus when referring to individuals with COVID-19.

Similarly, guidelines recommending the use of expressions such as “people who may have COVID-19” or “presumptive cases” instead of talking about “COVID-19 suspects” or “suspected cases” [6 p.3] were also not fully reflected in the corpus under study. Indeed, qualitative close reading of concordances further highlighted that “caso” was predominantly used to refer to individuals with COVID-19, rather than to those with suspected infection: in 95% of occurrences in the 2020 corpus (6127 instances) and 78.5% of occurrences in the 2021 corpus (5038 instances), “caso” was used to denote people who

had COVID-19 rather than people who were suspected of having the disease.

Stigmatising and hyperbolic language (RQ2)

In relation to the use of hyperbolic or alarmist language, guidelines recommend not to repeat or share unconfirmed rumours, and to “avoid using hyperbolic language designed to generate fear like ‘plague,’ ‘apocalypse’ etc.” [6 p. 3]. The corpus shows substantial compliance with these recommendations: quantitative analysis indicates that overtly alarmist terms such as “piaga” (“scourge”), “peste” (“plague”), and “apocalisse” (“apocalypse”) are rare and do not appear among the most frequent lemmas or keywords. In the 2020 corpus, “piaga” occurred four times and “peste” ten times, while “apocalisse” did not occur; in 2021, “piaga” occurred three times, “peste” twice, and “apocalisse” did not occur. These low frequencies seem to indicate limited reliance on overtly fear-inducing terminology as identified by WHO guidelines. Nevertheless, quantitative and qualitative analysis of wordlists and concordances showed that metaphors of war and uncontrolled spread persisted, both in 2020 and in 2021, e.g. with heavily connotated lemmas such as “guerra” (“war”), “battaglia” (“battle”), and “lotta” (“fight”) and “allarme” (“alarm”) recurring in both corpora, and being strong collocates of lemmas referring to preventive measures such as vaccines in the 2021 corpus.

The attention paid to preventive measures (RQ3)

Official guidelines encouraged reporters to talk positively and emphasise the effectiveness of prevention and treatment measures, instead of emphasising or dwelling on the negative, or messages of threat [6 p. 3], and indeed, preventive measures constitute a salient topic in both sub-corpora. In 2021, vaccination emerged as a major focus of coverage: 5.5% of content words and 13% of keywords are semantically related to vaccination. Qualitative discourse analyses also show that COVID-19 vaccines are generally framed positively, with emphasis on effectiveness and safety, although problematic instances of false balance can be detected (see Fiammenghi et al. 2025 for a more detailed analysis of the phenomenon of false balance in the corpus) [19]. Preventive measures more broadly are also prominent across both years: in the 2020 corpus, 11.5% of content words and 13% of keywords are semantically related to preventive measures (above all, distancing, isolation, and protective masks), while in the 2021 corpus the corresponding figures are 4.5% of content words and 9.5% of keywords. The qualitative analysis of the occurrence of these terms in context shows that they are consistently framed as necessary and effective measures to prevent contagion, and their use is not only strongly encouraged but also explained and described. Together, these quantitative indicators can be taken to

show sustained attention to prevention measures in the articles in the corpus.

The reliance on trusted and expert sources (RQ4)

PAHO recommendations insisted on the need to “quote scientists, researchers, public health professionals, academics, and specialists [...] who can clarify public concerns” [5 p. 3] and to “interview those on the front lines” [5 p. 13]. The analysis of sources quoted or interviewed in newspaper articles shows alignment with these recommendations: quantitatively, interviews account for a substantial proportion of articles in both corpora, with 449 interviews in 2020 (16%) and 474 interviews in 2021 (17%). In 2020, most interviewees were medical researchers (36%) or public health officials (36%), followed by politicians (19%), researchers in other fields (2.38%), and a small number of patients (three articles); the remaining 5% fell into an “other” category. In 2021, public health officials constituted the majority of interviewees (63.5%), followed by medical researchers (15%) and politicians (14.5%); one interview involved a patient, and the remainder were classified as “other”.

Qualitative analysis further shows that the Italian *Istituto Superiore di Sanità*, the country’s main public health institution, is consistently portrayed as an authoritative and reliable source of health information. Linguistic strategies such as impersonalisation, collectivisation, functionalisation, and anonymisation are used to construct the Institute as a social actor in the discourse. Close reading of concordances indicates that references to the Institute most frequently occur in connection with its regular reports monitoring contagion trends (18% of occurrences in the 2020 corpus) and the vaccination campaign (26% of occurrences in the 2021 corpus). Collocational analysis also shows that reporting verbs such as “say”, “confirm”, and “explain” commonly co-occur with “Istituto Superiore di Sanità”, with the Institute itself – rather than individual spokespersons – used metonymically as the grammatical subject of these verbs. Table 1 presents a summary of the results.

Discussion

Studies in linguistics have long argued that the different uses of language can have direct consequences, as words can shape public understanding, forge relationships, and modify people’s behaviours, starting with Austin’s groundbreaking 1975 volume on pragmatics, tellingly titled *How to Do Things With Words* [26] and continuing with corpus-based work on health communication and discourse. Public health communication studies have found evidence that effective communication campaigns can improve patients’ compliance with preventive and screening health measures [27, 28]; conversely, fake news, mis- and disinformation conveying anti-vaccination

messages have been found to decrease public willingness to adhere to such campaigns [29, 30]. Furthermore, ineffective communication potentially erodes trust in public health institutions and authorities, which can have long-term negative consequences for population health [31, 32].

Public health communication, however, is not only performed by public health institutions themselves (for example, when devising awareness-raising campaigns), but may also often happen through the mainstream press, which covers public health topics and reports the words uttered by public health institutions and spokespersons. Authoritative international public health institutions have been sensitive and alert as to the importance of effective communication since the beginning of the COVID-19 pandemic, developing guidelines and recommendations to journalists on how to ethically and effectively report on the pandemic [5, 6]. Despite this effort, authors have argued that the press coverage of the pandemic was at times problematic, with negative consequences on public understanding and compliance with preventive measures [31].

Recent studies examining COVID-19 news coverage in relation to specific reporting guidelines have similarly found a mixed picture, with partial adherence to recommendations alongside persistent problematic framings such as stigmatising labels, alarmist metaphors and unbalanced presentation of expert views [7–9, 33].

In line with this work, the present analysis, which for the first time applies a quali-quantitative corpus-assisted discourse approach to the 2020 and 2021 press review of ISS, the leading Italian public health institution, shows that Italian newspapers largely followed WHO and PAHO guidance on emphasising preventive measures and using institutional sources, while diverging from recommendations on non-stigmatising terminology and the careful handling of controversy.

The results uncovered a wealth of linguistic practices that aligned with official recommendations and could be considered best journalistic practices when covering public health issues, for example, the ample space devoted to preventive measures and vaccination and the heavy reliance on official sources of information, with spokespeople from public health institutions and scientific experts frequently quoted or interviewed to provide an informed perspective on the latest developments of the pandemic. Regarding the naming of the disease and the description of affected individuals, however, the analysis showed that journalists generally adopted the recommended technical terminology for the virus and disease but frequently referred to “COVID-19 cases” and “victims” who “spread” the virus and “infected” others, rather than to “people who have COVID-19”, thereby maintaining potentially stigmatising patterns.

Table 1 Selection and summary of recommendations on use of language included in the WHO’s and PAHO’s guide on reporting about COVID-19, and compliance in the corpus under study

Dos and don’ts from the official guidelines on COVID-19 reporting	Compliance in the corpus under study	Further details from the analysis
<p>The new coronavirus disease</p> <p>DO - talk about the new coronavirus disease (COVID-19)</p> <p>DON’T - attach locations or ethnicity to the disease, this is not a “Wuhan Virus”, “Chinese Virus” or “Asian Virus”. [6 p. 2]</p>	<p>YES</p> <p>COVID-19 and COVID are more frequent lemmas and stronger keywords in the corpus compared to “Wuhan” and “cinese” (“Chinese”), both in the 2020 and 2021 corpora.</p>	<p>2020</p> <p>“COVID”: freq. 1768, raw freq. 1.07, keyness score 9.3</p> <p>“COVID-19”: freq. 1393, raw freq. 0.84, keyness score 7.2</p> <p>Wuhan: keyness score 3.4</p> <p>Cinese: keyness score 2.9</p> <p>(see also Appendix)</p> <p>2021</p> <p>“COVID”: freq. 2792, raw freq. 2.60, keyness score 14.2</p> <p>“COVID-19”: keyness score 4.2</p> <p>Wuhan: -</p> <p>Chinese: -</p> <p>(see also Appendix)</p>
<p>People who have COVID-19</p> <p>DO - talk about “people who have COVID-19”, “people who are being treated for COVID-19”;</p> <p>DON’T - refer to people with the disease as “COVID-19 cases” or “victims”. [6 p. 3]</p>	<p>NO</p> <p>Both “caso” (“case”) and “vittima” (“victim”) are frequent lemmas and strong keywords in the corpus. [23]</p>	<p>2020</p> <p>“Caso”: freq. 6450, raw freq. 3.91,</p> <p>“Vittima”: keyness score 2.7</p> <p>(see also Appendix)</p> <p>2021</p> <p>“Caso”: freq. 6418, raw freq. 5.98, keyness score 4.0</p> <p>“Vittima”: freq. 516, raw freq. 0.31</p> <p>(see also Appendix)</p>
<p>People who may have COVID-19</p> <p>DO - talk about “people who may have COVID-19” or “people who are presumptive for COVID-19”</p> <p>DON’T - talk about “COVID-19 suspects” or “suspected cases”. [6 p. 3]</p>	<p>NO</p> <p>The qualitative close reading of the concordances of “caso” revealed the noun is used to mean “people with COVID-19” in 95% (6127) of occurrences in 2020 and 78.5% (5038) of occurrences in 2021.</p>	
<p>Hyperbolic language</p> <p>DO - speak accurately about the risk from COVID-19, based on scientific data and latest official health advice.</p> <p>DON’T - repeat or share unconfirmed rumours, and avoid using hyperbolic language designed to generate fear like “plague”, “apocalypse” etc. [6 p. 3]</p>	<p>YES/NO</p> <p>Neither “piaga”, “peste” (“plague”) nor “apocalisse” (“apocalypse”) are among the most frequently used lemmas and keywords in the corpus under study.</p> <p>However, metaphors of war and examples of heavily connoted language persist in both corpora, with the lemmas “guerra”, “battaglia”, “lotta” and “allarme” being prominent collocates of “vaccination” in the 2021 corpus.</p>	<p>2020</p> <p>Piaga: freq. 4</p> <p>Peste: freq. 10</p> <p>Apocalisse: -</p> <p>2021</p> <p>Piaga: freq. 3</p> <p>Peste: freq. 2</p> <p>Apocalisse: -</p>
<p>Preventive measures</p> <p>DO - talk positively and emphasise the effectiveness of prevention and treatment measures. [...]</p> <p>DON’T - emphasise or dwell on the negative, or messages of threat. [6 p. 3]</p>	<p>YES</p> <p>The qualitative analysis of the occurrence of terms relating to preventive measures (e.g. isolation, quarantine, and masks) in context shows that they are consistently framed as necessary and effective to prevent contagion.</p> <p>Vaccines are one main topic in the 2021 corpus.</p>	<p>5.5% of words and 13% of keywords are semantically related to the topic of vaccination. Qualitative discourse analyses show that COVID-19 vaccines are generally framed positively with emphases on effectiveness and safety, although problematic instances of false balance can be detected [19]</p>
<p>DO - emphasise the effectiveness of adopting protective measures to prevent acquiring the new coronavirus, as well as early screening, testing and treatment. [6 p. 3]</p>	<p>Protective measures are one main topic in both the 2020 and the 2021 corpus.</p>	<p>11.5% of words and 13% of keywords are semantically related to the topic of preventive measures in the 2020 corpus. 4.5% of words and 9.5% of keywords are semantically related to the topic of preventive measures in the 2021 corpus.</p>

Table 1 (continued)

Dos and don'ts from the official guidelines on COVID-19 reporting	Compliance in the corpus under study	Further details from the analysis
Trusted and expert sources DO - quote scientists, researchers, public health professionals, academics, and specialists [...] who can clarify public concerns. (5 p. 3) DO - interview those on the front lines. [5 p. 13]	YES Interviews occupy a large space both within the 2020 corpus and within the 2021 corpus.	449 interviews are present in the 2020 corpus (16%), 474 interviews in the 2021 corpus (17%). Most of the people interviewed in the 2020 corpus are medical researchers (36%) or public health officials (36%); 19% are politicians, whether national or local; 2.38% are researchers in other fields (statistics, economics and law); 3 articles in the 2020 corpus are interviews with patients. We assigned the remaining 5% to the category "other", including pub and restaurant managers. Most of the people interviewed in the 2021 corpus are public health officials (63.5%); 15% are medical researchers, and 14.5% are politicians. One text is an interview with a patient, and the remainder fall into the category "other".
DO – use trusted and reliable sources. [...] Treat scientific spokespersons in a dignified way. [5 pp. 19–21]	The Italian <i>Istituto Superiore di Sanità</i> is often portrayed as an authoritative and reliable source of health information in the corpus. Linguistic strategies are consistently used to impersonalise, collectivise, functionalise, and anonymise the Institute as a social actor in the discourse [25].	The close reading of the occurrences of the phrase "Istituto Superiore di Sanità" in both corpora shows that the Institute is mentioned when citing its regular reports monitoring contagions (18% of occurrences in the 2020 corpus) and the vaccination campaign (26% of occurrences in the 2021 corpus). Other strong collocates of the phrase "Istituto Superiore di Sanità" include reporting verbs such as "say", "confirm", and "explain" (see also Appendix); the Institute itself, and not one of its spokespersons, is used metonymically as subject of such verbs.

In terms of avoiding stigmatising or hyperbolic language more broadly, overtly sensational expressions were relatively rare, yet metaphors of war and uncontrolled spread persisted in some headlines and leads, which may have amplified anxiety and reinforced binaries between "infected" and "non-infected" groups. With respect to the emphasis placed on preventive measures, Italian newspapers devoted substantial space to vaccination, distancing, masks and hygiene, often foregrounding these measures in headlines and early paragraph positions, thus aligning with WHO and PAHO recommendations to highlight actionable behaviours. Finally, the analysis of reliance on trusted expert sources revealed a strong orientation towards institutional expertise: spokespersons from national and regional health authorities, as well as scientific experts, were frequently quoted and interviewed. However, the prominence of the written interview format sometimes created instances of false balance, in which divergent expert opinions were presented without sufficient contextualisation of the underlying strength of evidence. Three main areas should be considered when devising institutional communication interventions: (1) the use of specific terms and labels; (2) the reporting and interplay of quotations by researchers, scientists and public health officers; and (3) the broader messages emerging about institutions and their spokespersons.

These different areas would require different levels of attention and investment: while the accurate use of specific terms could be supported, also thanks to ICT-based tools, more sustained efforts are needed for the last two areas, ideally fostering medium- to long-term alliances

between public health officials and journalists. In this regard, adopting an organisational health literacy perspective could prove especially valuable, not only in crisis situations but also in times of stability, by promoting communication practices that are clear, inclusive, and responsive to the generally low levels of health literacy in the population [15].

The study has several limitations. Although the corpus built for the study is of good size and stems from a reasoned selection of widely circulated and editorially heterogeneous newspapers, we did not include newspapers characterised by extreme positions. This is in accordance with previous studies on media representations of public health crises that have focused on mainstream outlets rather than highly partisan titles [7–9]. However, the corpus did not include all the articles published by Italian newspapers during the pandemic and was dependent on the source and search string chosen for compilation; therefore, the results should be interpreted with caution, even within the Italian context. In addition, the corpus was compiled from the ISS national press review, which represents a curated selection of newspaper articles; since the full criteria guiding inclusion in this press review are not publicly detailed, the dataset should be interpreted as a structured sample rather than an exhaustive representation of Italian newspaper coverage. Furthermore, some of the results of the quantitative analysis could vary slightly by modifying the parameters chosen (in particular, different collocations could be obtained by selecting a different range of positions to the right and/or left of the node, or by modifying the statistical tests used), although we

minimised the risk of obtaining biased results by adopting standard tests and intervals. Finally, keyword analysis relied on itTenTen20 as a reference corpus, which reflects broad Italian web language rather than a curated newspaper register; this choice may affect keyness values and should be considered when interpreting keywords as indicators of journalistic salience. Moreover, the analysis focused on how institutional messages were represented in newspaper texts, not on the original drafting and internal circulation of institutional communication, so the findings should not be interpreted as a direct evaluation of institutional communication practices themselves.

Despite these limitations, we believe that such an analysis can contribute to a better understanding of how institutional public health messages are mediated in newspaper coverage during public health crises, and of the importance of instructing and educating public health officials on how to deal with the newspaper press. The COVID-19 pandemic was an unprecedented public health emergency, and Italy was the first Western country to be dramatically hit by it. Most public health institutions and journalists alike were not adequately prepared to face the communication challenges posed by the emergency, which were compounded by the complex information environment we are currently living in [32, 33]. However, important lessons were learned during the pandemic years, which could be further implemented in case a new public health emergency of a similar scale arises. These lessons also apply to non-crisis contexts: effective communication and collaboration between public health officers and the media are essential even in times of stability. Media can shape public expectations and perceptions of healthcare. Social listening during crises may also help institutions monitor public sentiment and emerging concerns, enabling more timely and targeted communication [32, 34, 35].

At the same time, we believe that public health institutions should systematically adopt multidisciplinary tools to analyse their communication campaigns in the mainstream newspapers ex-post. We believe that further research should be promoted to set up specific indicators and their intervals as well as to understand the meaning of the emerging quantitative values. This would be extremely beneficial to foster public health messages even in non-crisis times.

Conclusions

The corpus-assisted discourse analysis of the press review of the leading Italian public health institution, presented in the paper, has illustrated the best and the worst practices implemented by authoritative Italian newspapers covering the first two years of the COVID-19 pandemic, against the backdrop of WHO's and PAHO's recommendations to journalists. Quantitatively, the corpus showed

high salience of preventive measures and vaccination, and frequent reliance on institutional and expert sources, indicating substantial compliance with recommendations to emphasise actionable behaviours and trusted authorities. Qualitative analysis, however, revealed recurrent use of disease labels and metaphors that risk reinforcing stigma, and interview formats that occasionally produced false balance between evidence-supported and marginal positions, highlighting areas of partial or non-compliance that may affect public understanding.

Such an analysis has practical implications, as it highlights the need for joint training of public health professionals and newspaper writers, where knowledge can be exchanged on how to effectively communicate public health messages. This training should provide journalists with the tools to assess medico-scientific information, while public health officials could gain insights into the inner workings of media coverage, become more alert as to the power of their own words, and more able to craft their original message so that it can be more effectively translated into an effective mainstream message.

Finally, the study calls for continued interdisciplinary exploration of institutional communication and journalistic practices, to devise optimal communication strategies which could help public health authorities interact with the press in the event of new public health emergencies. In this perspective, it would be important to examine whether, in Italy and internationally, any systematic actions are being taken to incorporate lessons from COVID-19, such as clearer terminology, calibrated sourcing practices and joint training initiatives, into preparedness plans for possible future pandemics.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-026-27037-5>.

Supplementary Material 1.

Acknowledgements

The authors wish to thank Istituto Superiore di Sanità for access to the national press review and supporting the study.

Authors' contributions

CF contributed to methodology, formal analysis, investigation, and writing of the original draft, as well as review and editing of the manuscript. NP contributed to methodology, analysis, investigation, and writing of the original draft, and participated in writing – review and editing. EC contributed to the conceptualization of the study, writing – review and editing, and provided supervision. LC contributed to the conceptualization of the study, writing – review and editing, and provided supervision. MAT contributed to writing – review and editing, and funding acquisition. LP contributed to writing – review and editing. LB contributed to writing – review and editing. SB provided supervision and contributed to funding acquisition. UG provided supervision and contributed to funding acquisition. All authors read and approved the final manuscript.

Funding

This study was supported by Istituto Superiore di Sanità (ISS).

Data availability

The datasets generated and/or analysed are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

This study analysed publicly available newspaper articles included in the press review of the Italian National Institute of Health (Istituto Superiore di Sanità). The dataset consists exclusively of published media texts and does not contain personal data or involve human participants. According to the guidelines of the University of Brescia Ethics Committee, research based solely on publicly available textual materials does not require formal ethical approval or informed consent. The Declaration of Helsinki is therefore not applicable to this study.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Foreign Languages and Literatures, University of Verona, Lungadige Porta Vittoria, 41, Verona 37129, Italy

²Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, University of Brescia, Viale Europa, 11, Brescia 25121, Italy

³Department of Medicine, University of Udine, Piazzale Massimiliano Kolbe, 4, Udine 33100, Italy

⁴Italian National Institute of Health, Viale Regina Elena, 299, Roma 00161, Italy

Received: 14 October 2025 / Accepted: 12 March 2026

Published online: 17 March 2026

References

- Trehan T, Papanikitas A. Defining and detecting fake news in health and medicine reporting. *J R Soc Med*. 2020;113(9):302–5. <https://doi.org/10.1177/0141076820907062>.
- Allen J, Howland B, Mobius M, Rothschild D, Watts DJ. Evaluating the fake news problem at the scale of the information ecosystem. *Sci Adv*. 2020;6(14):eaay3539. <https://doi.org/10.1126/sciadv.aay3539>.
- Altay S, Nielsen RK, Fletcher R. Quantifying the infodemic: people turned to trustworthy news outlets during the 2020 coronavirus pandemic. *J Quant Descr Digit Media*. 2022;2:1–29. <https://doi.org/10.51685/jqd.2022.020>.
- Falcone R, Ansani A, Coli E, Marini M, Sapienza A, Castellfranchi C, Paglieri F. Trusting COVID–19 vaccines as individual and social goal. *Sci Rep*. 2022;12:9470. <https://doi.org/10.1038/s41598-022-13675-3>.
- Pan American Health Organisation, World Health Organisation. COVID–19: an informative guide—advice for journalists. Washington (DC): PAHO. 2020 Apr. Available from: <https://iris.paho.org/handle/10665.2/52392>. Accessed 1 Aug 2025.
- International Federation of Red Cross and Red Crescent Societies, UNICEF, World Health Organisation. Social stigma associated with COVID–19. 2020 Feb 24. Available from: [https://www.unicef.org/media/65931/file/Social%20stigma%20associated%20with%20the%20coronavirus%20disease%202019%20\(COVID-19\).pdf](https://www.unicef.org/media/65931/file/Social%20stigma%20associated%20with%20the%20coronavirus%20disease%202019%20(COVID-19).pdf) Accessed 1 Aug 2025.
- Li Z, Zhao R, Lou B. Corpus-based critical discourse analysis of reporting practices in English news reports on public health event in China and United States. *Front Psychol*. 2023;14:1137382. <https://doi.org/10.3389/fpsyg.2023.1137382>.
- Li L, Gao F. Contrasting semantic and sentimental features between Chinese and American economic news discourse in the epidemic era: a corpus-based critical discourse analysis. *Soc Sci Humanit Open*. 2023;7(1):100435. <https://doi.org/10.1016/j.ssoho.2023.100435>.
- Brookes G, Taylor C, Goodman S. COVID–19 news in China and the UK: a corpus-based critical discourse analysis of China Daily and The Times. *Corpora*. 2024;19(3):285–315. <https://doi.org/10.3366/cor.2024.0315>.
- Brookes G, Atkins S, Harvey K. Corpus linguistics and health communication: using corpora to examine health (–care) language. In: Harrington K, Mahlberg M, editors. *The Routledge handbook of corpus approaches to discourse analysis*. London: Routledge; 2022. pp. 653–75. <https://doi.org/10.4324/9780367076399-43>.
- Brookes G, Collins LC. *Corpus linguistics for health communication: a guide for research*. 1st ed. London: Routledge; 2023. <https://doi.org/10.4324/978103099659>.
- Bonifacio F, Darelli V, Turrini V, Zaffaroni LG. Crisis as revelation and opportunity: the Italian press storytelling of the COVID–19 pandemic first phase. *Comunicazioni Sociali*. 2024;1:76–89. https://doi.org/10.26350/001200_000201.
- Ophir Y, Sangalang A, Cappella JN, Yang Cynthia Chung J, Brennan E. The framing of COVID–19 in Italian media and its relationship with community mobility: a mixed–method approach. *J Health Commun*. 2021;26(3):161–73. <https://doi.org/10.1080/10810730.2021.1899344>.
- Denti O. COVID–19 crisis management in newspaper articles. *Linguistica e Comunicazione di Massa*. 2023;10(2):49–70. <https://doi.org/10.7358/lcm-2023-002-edit>.
- Pelizzari N, Covo L, Ceretti E, Fiammenghi C, Gelatti U. Defining, assessing, and implementing organizational health literacy: barriers, facilitators, and tools—a systematic review. *BMC Health Serv Res*. 2025;25:599. <https://doi.org/10.1186/s12913-025-12775-w>.
- Reynolds B, Quinn Crouse S. Effective communication during an influenza pandemic: the value of using a crisis and emergency risk communication framework. *Health Promot Pract*. 2008;9(4 Suppl):S13–7. <https://doi.org/10.1177/1524839908325267>.
- Rubinelli S, Fiordelli M, Zanini C, Diviani N. It is time for health institutions to invest in persuasive communication to combat low–quality information: a lesson learned from the COVID–19 infodemic. *Front Commun (Lausanne)*. 2022;7:1006302. <https://doi.org/10.3389/fcomm.2022.1006302>.
- Remuzzi A, Remuzzi G. COVID–19 and Italy: what next? *Lancet*. 2020;395(10231):1225–8. [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9).
- Fiammenghi C, Ceretti E, Covo L, Pelizzari N, Taranto MA, Penna L, Brunelli L, Brusafiero S, Gelatti U. A multi–disciplinary exploration of the press review of the Italian National Institute of Health during the COVID–19 pandemic. *Hum Vaccin Immunother*. 2025;00(00):1–17. <https://doi.org/10.1080/21645515.2025.2567680>.
- Baker P. Corpora and discourse analysis. In: Gee JP, Handford M, editors. *The Routledge handbook of discourse analysis*. London: Routledge; 2013. pp. 11–34.
- Kilgarriff A, Baisa V, Bušta J, Jakubíček M, Kovář V, Michelfeit J, Rychlý P, Suchomel V. *The Sketch Engine: ten years on*. *Lexicography*. 2014;1(1):7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- Stubbs M. *Corpus semantics*. In: Riemer N, editor. *The Routledge handbook of semantics*. London: Routledge; 2015. pp. 106–22.
- Boykoff MT, Boykoff JM. Balance as bias: global warming and the US prestige press. *Global Environ Change*. 2004;14(2):125–36. <https://doi.org/10.1016/j.gloenvcha.2003.10.001>.
- Clarke CE. A question of balance: The autism–vaccine controversy in the British and American elite press. *Sci Commun*. 2008;30(1):77–107. <https://doi.org/10.1177/1075547008320262>.
- Van Leeuwen TA. *Text and practice*. London, England: Routledge; 1996.
- Austin JL. *How to do things with words*. Oxford: Oxford University Press; 1975. <https://doi.org/10.1093/acprof:oso/9780198245537.001.0001>.
- Unlu A, Truong S, Tammi T, Lohiniva AL. Exploring political mistrust in pandemic risk communication: mixed–method study using social media data analysis. *J Med Internet Res*. 2023;25:e50199. <https://doi.org/10.2196/50199>.
- Zafri NM, Afroj S, Nafi IM, Hasan MMU. A content analysis of newspaper coverage of COVID–19 pandemic for developing a pandemic management framework. *Heliyon*. 2021;7:e06544. <https://doi.org/10.1016/j.heliyon.2021.e06544>.
- Baccolini V, Renzi E, Isonne C, et al. COVID–19 vaccine hesitancy among Italian university students: a cross–sectional survey during the first months of the vaccination campaign. *Vaccines (Basel)*. 2021;9:1292. <https://doi.org/10.3390/vaccines9111292>.
- Bilal A, Hoogensen Gjørnv G, Lanteigne M, et al. Comprehensive security, disinformation, and COVID–19: an analysis of the impacts of mis– and disinformation and populist narratives during the pandemic. *Open Res Eur*. 2023;3:209. <https://doi.org/10.12688/openreseurope.16733.1>.
- Sorath F, Shiwani S, Sindhu F, et al. A systematic review of the attitudes, beliefs, and acceptance of the COVID–19 vaccine in the western and eastern hemispheres. *Cureus*. 2024;16:e73161. <https://doi.org/10.7759/cureus.73161>.
- Muğaloğlu EZ, Kaymaz Z, Mısırlı ME, Laçın–Şimşek C. Exploring the role of trust in scientists to explain health–related behaviours in response to the

- COVID–19 pandemic. *Sci Educ (Dordr)*. 2022;31:1281–309. <https://doi.org/10.1007/s11191-022-00323-5>.
33. Kania U. Snake flu, killer bug, and Chinese virus: a corpus–assisted critical discourse analysis of lexical choices in early UK press coverage of the COVID–19 pandemic. *Front Artif Intell*. 2022;5:970972. <https://doi.org/10.3389/frai.2022.970972>.
 34. Allen J, Watts DJ, Rand DG. Quantifying the impact of misinformation and vaccine–skeptical content on Facebook. *Science*. 2024;384:eadk3451. <https://doi.org/10.1126/science.adk3451>.
 35. Zago D, Cautero P, Scarpis E, et al. TV medical dramas: assessing the portrayal of public health in primetime. *Front Public Health*. 2024;12:1432528. <https://doi.org/10.3389/fpubh.2024.1432528>.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.