UNIVERSITY OF SPLIT
SCHOOL OF MEDICINE

Luka Ursić

A CRITICAL ASSESSMENT OF HEALTH SYSTEMS GUIDANCE AND THE UNDERLYING DISCOURSE IN THE H1N1 AND COVID-19 PANDEMICS

Dissertation thesis

Split, 2025

UNIVERSITY OF SPLIT
SCHOOL OF MEDICINE

Luka Ursić

A CRITICAL ASSESSMENT OF HEALTH SYSTEMS GUIDANCE AND THE UNDERLYING DISCOURSE IN THE H1N1 AND COVID-19 PANDEMICS

Dissertation thesis

Mentor: Prof. Damir Sapunar, MD, PhD

Split, 2025

The research presented here was funded by the Croatian Science Foundation (grant agreement IP-2019-04-4882: “Professionalism in Health – Decision making in practice and research, ProDeM”. One part of the study entitled “Scientific discourse during health crises: A linguistic and content analysis of editorials on the response to the H1N1 and COVID-19 pandemics” was conducted within the Stehr-Boldt fellowship programme at the University of Zürich Institute of Biomedical Ethics and History of Medicine. The funders had no role in the design and conduct of the studies; the collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript produced within these studies; or the decision to submit them to a specific scholarly journal.

This dissertation is based on the following published manuscript:

Ursić L, Žuljević MF, Vuković M, Bralić N, Roje R, Matas J, et al. Assessing the quality and completeness of reporting in health systems guidance for pandemics using the AGREE-HS tool. J Glob Health. 2023 Oct 27;13:06050.

ACKNOWLEDGEMENTS

CONTENT

1. INTRODUCTION ---------- 1

1.1. Development of guidelines during pandemics ---------- 1

1.2. Definition of a pandemic ---------- 2

1.2.1. H1N1 pandemic ---------- 3

1.2.2. COVID-19 pandemic ---------- 3

1.3. Developing and assessing health systems guidance – status and gaps ---------- 4

1.3.1. Practical use by GDG members ---------- 5

1.3.2. Uptake and use by policymakers and other stakeholders ---------- 6

1.3.3. Guideline evaluations and assessments ---------- 6

2. OBJECTIVES ---------- 8

2.1. Aims and hypotheses ---------- 9

3. MATERIALS AND METHODS ---------- 10

3.1. Study 1: AGREE-HS assessment ---------- 11

3.1.1. Data sources and search strategy ---------- 11

3.1.2. Screening process and eligibility criteria ---------- 11

3.1.3. AGREE-HS tool ---------- 12

3.1.4. Statistical analysis ---------- 14

3.1.5. Software ---------- 14

3.2. Study 2: levels of evidence underlying the HSG ---------- 15

3.2.1. Data sources and search strategy ---------- 15

3.2.2. Screening process and eligibility criteria ---------- 15

3.2.3. Extraction of recommendations and classification of evidence ---------- 16

3.2.4. Statistical analysis ---------- 17

3.2.5. Software ---------- 17

3.3. Study 3: linguistic analysis of editorials ---------- 18

3.3.1. Data sources and search strategy ---------- 18

3.3.2. Screening process and eligibility criteria ---- *18*

3.3.3. Data extraction ---- *19*

3.3.4. Sentiment analysis ---- *19*

3.3.5. Statistical analysis ---- *19*

3.3.6. Software ---- *19*

4. RESULTS ---- 21

4.1. Study 1: AGREE-HS assessment ---- 22

4.1.1. Overall and organisation-specific scores ---- *23*

4.1.2. Recommendations for use ---- *28*

4.1.3. Assessor's comments ---- *29*

4.2. Study 2: levels of evidence underlying the HSG ---- 32

4.2.1. Characteristics of HSG, their recommendations, and their supporting statements ---- *33*

4.2.1. Classification of the evidence underlying the recommendations ---- *34*

4.3. Study 3: linguistic analysis of editorials ---- 36

4.3.1. LIWC analysis ---- *37*

5. DISCUSSION ---- 39

5.1. Summary of findings ---- 40

5.2. Transparency and completeness of reporting in HSG: implications and a need to establish standards ---- 41

5.1.1. Strengths and limitations ---- *43*

5.3. Evidence base for pandemics: lessons learned for future outbreaks ---- 44

5.2.1. Strengths and limitations ---- *45*

5.4. Scientific communication: (im)partiality in a time of disinformation and misinformation ---- 46

5.3.1. Strengths and limitations ---- *48*

6. CONCLUSION ---- 50

6.1. Practical implications of findings and way forward ---- 51

7. SUMMARY ---------- 53

8. LAY SUMMARY ---------- 55

9. CROATIAN SUMMARY ---------- 57

10. CROATIAN LAY SUMMARY ---------- 59

11. REFERENCES ---------- 61

12. CURRICULUM VITAE ---------- 72

13. SUPPLEMENT ---------- 78

LIST OF ABBREVIATIONS

AGREE II – Appraisal of Guidelines for Research & Evaluation II

AGREE-HS – Appraisal of Guidelines for Research and Evaluation – Health Systems

CAMARADES – Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies

CDC – United States Centers for Disease Control and Prevention

CI – confidence interval

COI – conflict of interest

CPG – clinical practice guideline

ECDC – European Centre for Disease Control and Prevention

EQUATOR – Enhancing the QUAlity and Transparency Of health Research

GDG – guideline development groups

GRADE – Grading of Recommendations Assessment, Development and Evaluation

HSG – health systems guidance

IHR – International Health Regulations

IQR – interquartile ranges

IRIS – Institutional Repository for Information Sharing

LIWC – Language Inquiry and Word Count

MD – median

PHEIC – public health emergency of international concern

R^2 – coefficient of determination

RCT – randomised controlled trial

SARS-CoV-2 – severe acute respiratory syndrome coronavirus 2

SD – standard deviation

SyRF – Systematic Review Facility

WHO – World Health Organization

WoS – Web of Science

$\bar{x}$ – sample mean

1. INTRODUCTION

1.1. Development of guidelines during pandemics

The International Health Regulations (IHR), broadly revised in 2005 as a legal framework proscribing the World Health Organization (WHO) Member States' rights and obligations in responding to disease outbreaks, define a public health emergency of international concern (PHEIC) as an "extraordinary event" that constitutes "a public health risk through the international spread of disease" which could "require a coordinated international response" (1). Seven such PHEICs have been declared in the 21st century (2). However, only two – the 2009 H1N1 and the COVID-19 outbreaks – have affected most countries and regions worldwide (3) and have been considered pandemics (2) (**Figure 1**).

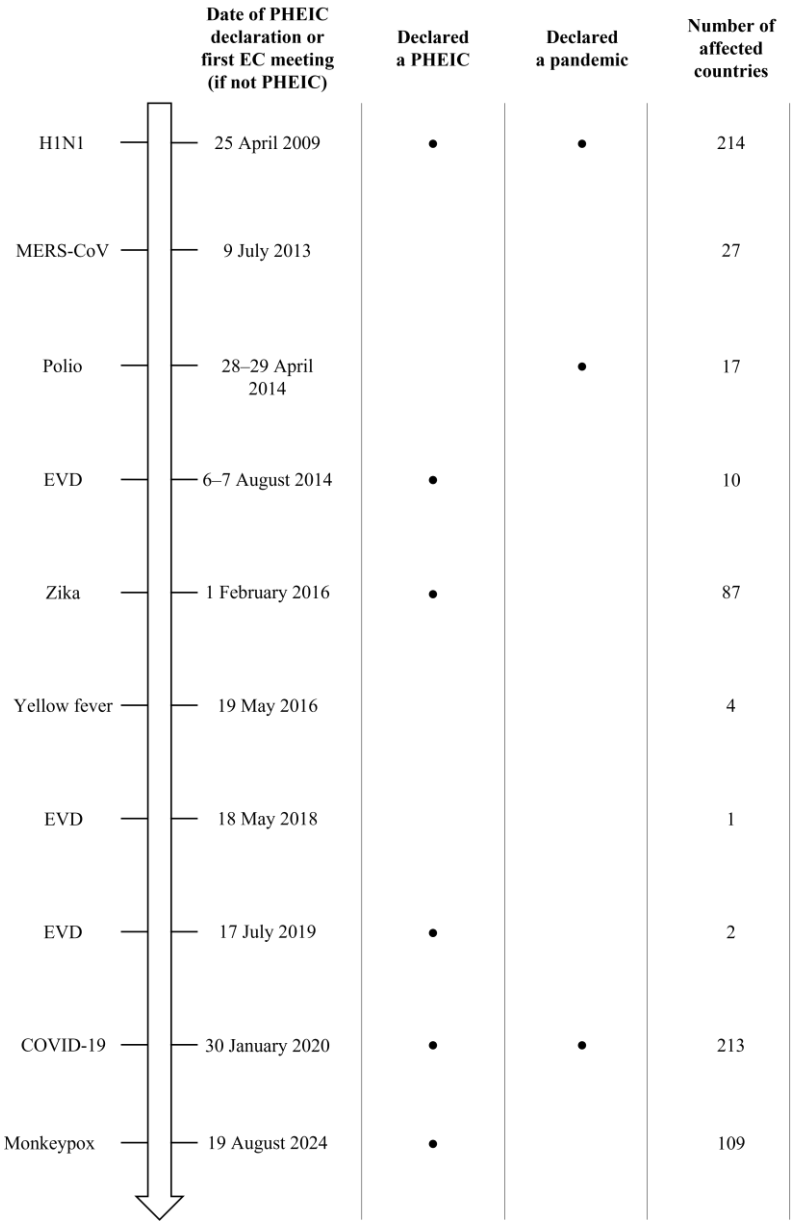

Figure 1. PHEICs of the 21st century. Adapted from published data (2, 3). Date of Monkeypox declaration and the number of affected countries added from WHO reports (4, 5). EVD – Ebola virus disease, PHEIC – public health emergency of international concern.

Through the IHR, the WHO is obliged to "publish, in consultation with Member States, guidelines to support States Parties in the development of public health response capacities" during PHEICs, while cooperating with and coordinating between other national bodies and governments in these efforts (1). In the WHO's terminology, such guidelines are defined as documents "developed by the World Health Organization containing recommendations for clinical practice or public health policy" and can be broadly categorised as "standard", "consolidated", "interim", "emergency (rapid response)", and "rapid advice" (6). The latter three types become relevant during disease outbreaks; while they are issued during such periods of high pressure and are therefore usually based on limited evidence, they are still expected to rely on robust development processes (6, 7). However, they are ultimately legally non-binding to the Member States, which are allowed to diverge from them within their own contexts (8).

In the context of Europe and the USA, the WHO's role in providing guidance to policymakers during infectious disease outbreaks is mirrored by two other organisations – the European Centre for Disease Prevention and Control (ECDC) (9) and the US Centers for Disease Control and Prevention (CDC) (10). Similar to the WHO, their recommendations and guidelines are non-binding, but are still regarded as key sources for evidence-informed decision-making in healthcare at local, state, and national levels (9–11).

To a varying extent, all three organisations saw their decisions and actions challenged during the H1N1 and COVID-19 pandemics. The WHO, for example, was criticised for overestimating the severity of the H1N1 pandemic and not reporting conflicts of interest (CoIs) among guideline development group (GDG) members (12, 13). The CDC was similarly disparaged for lacking transparency and rigour in producing its COVID-19 guidance (14). While the ECDC mainly received praise, some experts raised concerns regarding its limited mandate and highlighted shortcomings in its communication towards the public (15).

1.2. Definition of a pandemic

The IHR criteria for defining a PHEIC have been contested previously (16), as has their practical application in designating disease outbreaks (3). The exact definition of a pandemic and its meaning in comparison to a PHEIC has been even more dubious (2). The term currently carries no legal connotation within the WHO terminology and is defined only within the context of pandemic influenza (17). Specifically, a PHEIC would connote a disease outbreak affecting or having the potential to affect several member states; a pandemic, meanwhile, would indicate its global spread beyond that of an epidemic (18), making "all pandemics (…) PHEICs", but not *vice versa* (17).

The uptake and implications of these two terms in national contexts are even more limited: a recent analysis of 28 WHO Member States' legislation found that only 16.7% and 37.5% referenced PHEICs and pandemics, respectively, with almost half only doing so post-2020 (2). The authors concluded that this ambiguity and misalignment between national legislation and WHO frameworks bear significant implications for global health, especially given the organisation's reliance on normative power when declaring a PHEIC or a pandemic and providing guidance to Member States (2). Historically, the simultaneous use of the two designations for H1N1 and COVID-19 – *i.e.* "pandemic" and "PHEIC" – weakened the WHO's role in leading countries towards a collective, equitable response to these crises (2, 3).

1.2.1. H1N1 pandemic

The WHO designated the H1N1 outbreak a pandemic on 11 June 2009 (19, 20), following the Emergency Committee's initial declaration of a PHEIC on 25 April 2009 (21). The research community and the general population reacted with worry about whether the organisation overestimated the threat of the influenza virus (22–25). These mounting concerns had been exacerbated by the WHO's removal of the criterion of severity from its definition of a pandemic just prior to the H1N1 outbreak, as well as the erasure of pandemic guidelines from its website (26).

Subsequent inquiries into the WHO's response to the pandemic found that these concerns were not unwarranted, as they uncovered potential CoIs among members of its GDGs due to their undisclosed ties to the pharmaceutical industry (12, 27, 28). Considering the Member States' decisions to stockpile vaccines and antivirals, which went on to be unused at massive expense to their public health funds (12), these issues were debated extensively among researchers, with open calls to overhaul the WHO's guideline development processes (29–34). Ultimately, all the concerns were reiterated by the Council of Europe's Parliamentary Assembly, which called on the organisation to adopt mechanisms that would address these gaps in the future (35).

1.2.2. COVID-19 pandemic

The WHO Emergency Committee declared the outbreak of a novel coronavirus in China a PHEIC on 30 January 2020 (36). Less than two months later, on 11 March 2020, the organisation's director rang the alarm "loud and clear" by characterising the PHEIC as a pandemic, citing that more than 118,000 cases of COVID-19 had been detected in 114 countries (37). In contrast to the H1N1 pandemic, initial reactions focussed on whether the WHO had

reacted quickly enough to the outbreak in China, and if its limited funding and overdependence on Member States restricted its ability to coordinate the global response (38, 39).

More importantly, early analyses of the WHO's interim clinical practice guidelines (CPGs) for COVID-19 noted their methodological weaknesses due to the "(lack of) discussion on the applicability of the guidelines, inadequate recording of conflicts of interest, a narrow range of included stakeholders, and insufficient planning for updating the document" (40), echoing the issues observed regarding the organisation's guideline development processes during the H1N1 pandemic. Other researchers similarly observed inconsistencies in the WHO's early recommendations on mask wearing and raised worries about how this would be perceived by the general public (41).

Similar gaps were observed in the USA as well, where the CDC had been criticised for failing to transparently communicate its guidance to the public (42, 43) and accurately report key data (44), as well as its decision to opt for an individualistic, rather than collective and equitable approach to formulating its masking recommendations (45). In late 2022, the agency's internal review highlighted a need for more robust reporting and clearer presentation of its public health guidance, as well as the use of "evidence-based rating systems" in the production thereof (46). Regarding the ECDC's response in the European context during the early pandemic, stakeholders highlighted that the agency's ability to provide guidelines to EU countries was stymied by its limited mandate, relatively low number of staff, inadequate technical capacity, and general lack of visibility, noting that these factors limited the uptake of its guidance by policymakers (47).

1.3. Developing and assessing health systems guidance – status and gaps

Existing research has identified challenges in the development of public health guidelines and health systems guidance (HSG) beyond the scope of either of the three organisations or the two pandemics. In contrast to CPGs, which assist practitioners in diagnosing and managing patients (48), HSG are targeted at decision-makers within a healthcare system and constitute "systematically developed recommendations to manage challenges related to health system governance, financial and delivery arrangements, and the implementation strategies needed to get appropriate programs and services to those who need them" (49). Aside from the experience of GDGs, studies have also explored issues with the uptake of such guidance among decision-makers, and have assessed their quality and completeness in terms of methodological rigour, transparent reporting of CoIs, and other factors such as equity and cost considerations.

1.3.1. Practical use by GDG members

From a development perspective, GDG members reported challenges in using existing, standardised evidence assessment frameworks to evaluate complex interventions. For example, individuals who had previously used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework for this purpose found it to be inflexible, as they had to grade evidence to be of lower quality when it came from nonrandomised or observational studies or due to reasons of performance bias (*e.g.* studies which lacked blinding simply because it was impossible to establish) (50).

Similar observations were made in public health specifically, where a scoping review that sought to understand challenges related to the use of the GRADE approach in this context recorded five main barriers (51):

1. "incorporating the perspectives of diverse stakeholders";
2. "selecting and prioritising health and "nonhealthy" outcomes";
3. "interpreting outcomes and identifying a threshold for decision-making";
4. "assessing certainty of evidence from diverse sources, including nonrandomised studies";
5. "addressing implications for decision makers, including concerns about conditional recommendation".

In general, guidelines issued in the context of public health emergencies should still adhere to robust and transparent development processes (7, 52). However, this is seemingly difficult to realise in practice. For example, GDG members with experience in developing rapid guidelines saw limited or weak evidence, time pressure, and lack of funding and personnel as the main obstacles to the process (53). They also held different opinions on whether the production of such guidelines can adhere to rigorous evidence synthesis methodologies due to the limited timeframe, whether all stakeholders (*e.g.* health economists) should be included and all considerations (*e.g.* costs or values and preferences) given in the development process, and whether external peer reviewers should be engaged to evaluate its outcomes (53). Researchers who had conducted rapid evidence reviews during the COVID-19 pandemic experienced issues with aligning their work with the needs of local, national, and international decision-makers and clinical stakeholders and including them in the research process, rendering their findings less useful in practice (54).

1.3.2. Uptake and use by policymakers and other stakeholders

The uptake of guidelines by target audiences following their development is likewise affected by diverse factors, with a prominent one being the adaptation of guidelines developed at global levels to local, national contexts. For example, a scoping review that explored how WHO-developed guidelines are adopted in Member States found the non-existence or restrictiveness of existing health system legislation, a lack of experience and capacity at both the provider and infrastructure levels, and inadequate funding for implementation and uptake to be the main barriers to the uptake process (55). Another review that mapped the evidence on the adaptation and implementation of HSG in low- and middle-income countries found unclear reporting on the methodology for tailoring said guidelines to their specific context (56). Interviews with researchers who worked with policymakers to translate evidence from guidelines into practice highlighted that the latter want to be involved in the research itself to ensure it addresses their needs, but that this process is time-consuming and requires consideration of the context where the evidence is being implemented, as well as any potential divergences between the researchers' and policymakers' interests (57).

1.3.3. Guideline evaluations and assessments

Most studies on guidelines issued during PHEICs and similar crises focussed on CPGs. For example, an evaluation of WHO CPGs issued for four public health emergencies (H1N1, H7N9, Middle East respiratory syndrome coronavirus, and Ebola outbreaks) found they fell short of expected standards, scoring low for rigour of development, editorial independence, and applicability, as *per* the Appraisal of Guidelines for Research & Evaluation II (AGREE II) tool (58). A similar assessment of early COVID-19 guidelines had similar findings (40).

Few studies assessed public health guidelines or HSG issued during PHEICs or similar crises. An evaluation of mental health and psychosocial support guidelines issued during emergencies using the Appraisal of Guidelines for Research and Evaluation – Health Systems (AGREE-HS) tool found them to be of low quality, especially regarding the reporting of development methodologies, expert involvement, funding disclosures, and implementability considerations (59). Similar findings emerged regarding the HSG appraised in a review of post-COVID-19 syndrome rehabilitation guidelines (60). More recently, a study assessing the public health emergency response plan of Kerala (a state in India) for the Nipah virus found it to be of moderate quality, albeit lacking in terms of specific outcome measures and evaluation frameworks for the proposed interventions, as well as considerations of their ethical implications for society (61). One preprint reporting on an assessment of WHO CPGs, HSG,

and integrated guidelines (*i.e.* those integrating components of both) for epidemics found that, when assessed with the AGREE-HS tool, the latter two received similar scores, with shortcomings in reporting on the composition of the GDG and the implementability of their recommendations (62). One post-COVID-19 analysis of 130 public health guidelines issued by national and international organisations and societies found they rarely considered the values and preferences in their recommendations, while those that did infrequently used non-systematic processes (63).

2. OBJECTIVES

2.1. Aims and hypotheses

The goal of this thesis was to deepen our understanding of how the HSG for the H1N1 and COVID-19 pandemics were produced by three globally relevant organisations – the WHO, the CDC, and the ECDC. Specifically, it aimed to determine the gaps in the reporting of the processes used to develop the HSG and their recommendations, as well as the strength of their evidence base. It also sought to analyse the scientific community's discourse on the response of healthcare systems worldwide to the crisis, as it had a significant impact on public trust in science, policymakers, and research evidence.

Three studies were designed to explore these issues, the first of which aimed to evaluate the completeness and transparency of reporting in HSG issued by the WHO, the CDC, and the ECDC for the H1N1 and the COVID-19 pandemics using the AGREE-HS tool (64). The second study, leaning on the dataset and expanding the findings of the first one, sought to determine the evidence base underlying the HSG for the COVID-19 pandemic and explore how it changed over time. The third study, designed as a linguistic and content analysis of editorials published in scholarly journals, investigated the scientific discourse on the public health response to the H1N1 and the COVID-19 pandemics.

Due to its exploratory nature, the first study aimed to answer the following research question:

1. According to the AGREE-HS tool, how completely and transparently were the development processes of the HSG issued by the WHO, CDC, and the ECDC for the H1N1 and COVID-19 pandemics reported in the HSG themselves?

The second and third studies tested the following hypotheses:

1. There will be no change in the levels of evidence underlying the HSG issued by the ECDC and the WHO during the COVID-19 pandemic between the earlier and later stages of the pandemic.
2. The editorials published on the public health response during the COVID-19 pandemic will have higher scores for negative tone and emotion and personal/person-centred language, certitude, and all-or-none thinking; and lower scores for analytical thinking compared to those published during the H1N1 pandemic.

3. MATERIALS AND METHODS

3.1. Study 1: AGREE-HS assessment

The first study was done in two stages, beginning with a robust extraction and screening process. In the first stage, the repositories and websites of the WHO, the ECDC, and the CDC were systematically searched for relevant HSG, which were then screened for eligibility according to pre-determined criteria. The second stage comprised the piloting of the AGREE-HS tool, training for all evaluators on how to use it, and the full evaluation of the analytical sample. While no single reporting guideline fit this study design, the record extraction and screening processes were reported *per* the PRISMA-S guidelines (65).

3.1.1. Data sources and search strategy

As the study focussed on HSG issued by the WHO, the ECDC, and the CDC specifically, records were extracted from their institutional repositories rather than bibliographic databases such as PubMed or Web of Science (WoS). Specifically, the WHO Institutional Repository for Information Sharing (IRIS), the CDC Stacks, and the ECDC's main repository were queried using sensitive search strategies, as they had relatively low searchability and allowed for limited uses of Boolean operators and wildcard symbols. The organisations' H1N1-dedicated webpages, the WHO Committee Approved Guidelines, and the MMWR. Morbidity and Mortality Weekly Report journal were additionally searched for relevant records (Table S1 in the **Supplement**). The endpoint of the extraction was 17 May 2022.

3.1.2. Screening process and eligibility criteria

After extraction and deduplication, the titles and abstracts were screened for eligibility, followed by the full text of the records that were deemed to be of interest. This was done in duplicate to reduce bias and ensure the inclusion of all relevant HSG, whereby one researcher screened all the records in pair with a dedicated researcher for each organisation. Three researchers then proceeded to pilot the AGREE-HS tool on a randomly extracted subsample of included HSG and, following discussion, screened a randomly allocated set of the remaining records to determine their eligibility for the analysis.

Discrepancies at any stage were resolved through discussion. The interrater agreement for the screening process was calculated using Cohen's kappa for pairs, Fleiss' kappa for triplicate screening, and raw percentage agreement (66, 67).

The eligibility criteria were informed by the AGREE-HS tool, whereby any HSG issued by the three organisations and targeted towards decision-makers within healthcare systems were included for analysis. Topically, the HSG had to provide recommendations on health system-level strategies and approaches towards nonpharmaceutical interventions (masking, contact tracing, *etc*.), infection prevention and control measures, healthcare system preparedness, provision of healthcare services or supplies, or resource allocation (inclusive of human resources such as healthcare staff).

As the analysis focussed on HSG only, records dealing with the clinical management of H1N1 or COVID-19, providing technical guidance or criteria for testing devices or PPE, or containing only checklists and tools intended for the implementation of specific strategies were excluded during screening. This was also the case for policy briefs or similar guideline summaries (inclusive of those intended for public use), records providing infographics or answers to frequently asked questions, raw datasets, and risk reports or assessments.

3.1.3. AGREE-HS tool

The AGREE-HS tool was developed in 2018 by the AGREE-HS Research Team at McMaster University in Ottawa, Canada (64). It belongs to the wider family of AGREE tools, which have been used extensively in evaluating various aspects of CPGs (68). In contrast to these tools, however, AGREE-HS is mainly meant to be used on HSG, defined as "systematically developed statements to assist with decisions about appropriate options for addressing health system challenges, the implementation of these options, and the monitoring and evaluation of the implementation efforts" (64). Aside from suggesting a methodological and reporting framework for the GDGs working on developing HSG, it can also be used to assess the quality and completeness of reporting thereof. In this sense, the tool has been validated and tested in several studies following a robust development process that relied on surveys and expert consultation (59, 69–71).

The AGREE-HS tool comprises five main items (**Table 1**) rated on a Likert-type scale of 1–7 (lowest to highest quality). The evaluators grade each item based on an exhaustive set of criteria, downgrading their scores in cases where specific criteria are not met, such as when the composition of the GDG is not clearly reported. Additionally, the evaluators can leave comments elaborating on each grading decision and complete two additional items on whether they would recommend the use of the HSG in an "appropriate" context or in their own context, with response options being "Yes", "Yes, with modifications", and "No".

Table 1. AGREE-HS tool items and their description (64)

AGREE-HS item	Description
Topic	"This item addresses the description of the health system challenge, the causes of the challenge and the priority accorded to it, and relevance of the guidance."
Participants	"This item addresses the composition of the health systems guidance development team and the management of competing interests and funder influence."
Methods	"This item addresses the use of systematic methods and transparency in reporting; the use of the best available and up-to-date evidence; the consideration of effectiveness and cost-effectiveness of the potential options; and the weighting of benefits and harms in the guidance document."
Recommendations	"This item addresses the outcomes orientation and comprehensiveness of the guidance; the ethical and equity considerations drawn upon in its development; the details for its operationalization; the sociocultural and political alignment of the guidance; and the updating plan."
Implementability	"This item addresses the barriers and enablers to implementing the recommendations; the cost and resource considerations in implementing the recommendations; the affordability of implementation and anticipated sustainability of outcomes; the flexibility and transferability of the guidance; and the strategies for disseminating the guidance, monitoring its implementation and evaluating its impact."

AGREE-HS – Appraisal of Guidelines Research & Evaluation – Health Systems

For this analysis, an Excel sheet was designed containing the five AGREE-HS items and the two final recommendations, as well as additional comment fields for each decision. Prior to piloting the spreadsheet in a training session, five evaluators were provided with the AGREE-HS manual and checklist, as well as relevant literature. The training session was followed by a discussion aimed at clarifying any discrepancies and misunderstandings in how to use the tool then clarified, which helped ensure robustness and consistency throughout the full evaluation. Finally, the records were randomised to the five evaluators, with each record being evaluated in duplicate. The evaluators all had training and experience in biomedical research, but came from a diverse set of disciplines (social sciences, humanities, and biomedicine).

3.1.4. Statistical analysis

The characteristics of the included HSG were summarised using frequencies and percentages. Due to their ordinal nature, both the domain-specific and total AGREE-HS scores were presented using medians (MDs) and interquartile ranges (IQRs), with the latter calculated *per* the recommended formula (64), as follows:

$$\text{AGREE} - \text{HS SCORE} = \frac{\text{Obtained score} - \text{minimum score}}{\text{Maximum score} - \text{minimum score}}$$

The scores between the pandemics and between each organisation were compared using Mann-Whitney's U and the Kruskal-Wallis test, respectively, while any statistically significant differences observed in the latter analysis were further explored using the Dwass-Steel-Critchlow-Fligner *post-hoc* test. Lastly, ordinal regression was used to determine which of the domains served as a predictor of an HSG being recommended for use.

3.1.5. Software

The records from the WHO- and CDC-based sources (with the exception of its H1N1 guidance page) were exported and screened in EndNote, version X9 (Clarivate, London, UK). As the CDC's H1N1-dedicated webpage and the ECDC's sources did not allow for this approach, they were scraped using *R*, version 4.2.1 (R Core Team, Vienna, Austria) or Python, version 3.8.8 (Python Software Foundation, Delaware, USA) and managed in Excel, version 1808 (Microsoft Corp., Redmond, Washington, USA). The deduplication process was handled either manually in EndNote or automatically using a Python script. Record allocation for eligibility screening and AGREE-HS evaluation, and the interrater agreement analysis were done in *R*. Further details on the scraping and deduplication process are available in Text S1 in the **Supplement**.

All statistical analyses were performed and related figures generated in jamovi, version 2.3.16 (jamovi project, Sydney, Australia), Python, version 3.8.8 (Python Software Foundation, Delaware, USA), and MedCalc, version 20.218 (MedCalc Software Ltd, Ostend, Belgium).

3.2. Study 2: levels of evidence underlying the HSG

For the analysis of the levels of evidence underlying these HSG, the dataset obtained in the first study was updated until 18 July 2024, after which all statements and recommendations from the HSG were extracted along with their supporting "evidence" in the form of referenced studies. These were then categorised manually according to their study design and methodology. As with the first study, the screening process was reported *per* the PRISMA-S guidelines (65).

3.2.1. Data sources and search strategy

The data source for the analysis was the analytical sample from the first study (72), which was expanded through manual retrieval of prior versions of the included HSG from the initially extracted records and through manual searches of the institutional repositories. However, the HSG issued for the H1N1 pandemic by all organisations and those issued by the CDC for the COVID-19 pandemic were excluded *a priori*, as they did not clearly present their evidence base, preventing further analyses. This latter shortcoming was identified by the CDC itself in a review of its COVID-19 guidance (46).

This manual search was supplemented by an updated search from 17 May 2022 (endpoint date for extraction in the first study) up to 18 July 2024 (date of last search). In contrast to the first study, these searches were not done using specific keywords, but rather included all records published in the repositories within the given period, as this annulled the possibility of any HSG being overlooked or not extracted, and as the number of records published in this period was manageable for screening. While the search was precisely filtered for the period from 17 May 2022 to 18 July 2024 in the ECDC repository, this could not be done for the WHO IRIS repository due to limitations with its search capacity. Rather, all results were filtered to years 2022, 2023, and 2024, and any records published before 17 May 2022 were discarded, as they had already been included in the first study. Deduplication was not performed at this stage.

3.2.2. Screening process and eligibility criteria

Following the methodology of the first study, titles/abstracts and full-texts of records from the updated extraction were screened in duplicate, with one researcher screening all records in pair with another dedicated researcher for each organisation. The three researchers followed the eligibility criteria from the first study in which they had likewise participated, annulling the need for additional training or piloting. The manually retrieved, prior version of

the HSG were not screened for eligibility, as these were explicitly linkable to already included records (*e.g.* through direct hyperlinks or clear indications of a HSG being updated) and were therefore automatically included. All disputes were resolved through discussion.

3.2.3. Extraction of recommendations and classification of evidence

Two researchers entered the recommendations from the HSG into a pre-designed, piloted extraction sheet, identifying them based on the taxonomy proposed by Lotfi and colleagues as "formal" or "informal" (**Table 2**) (73).

Table 2. Criteria for formal and informal recommendations, as defined by Lotfi and colleagues (73)

Criteria	Formal recommendation	Informal recommendation
Actionable	Yes	Yes
Based on prioritised question	Yes	Yes
Population defined	Yes	Yes
Intervention defined	Yes	Yes
Comparator defined	If applicable	If applicable
Strength, direction, and certainty of evidence	Yes	No
Supported by systematic reviews and evidence synthesis	Yes	No, but can be based on select evidence
Resulted from a deliberative, structured recommendation development process	Yes	No

After extracting the recommendations, the researchers manually extracted any supporting statements and related evidence (*i.e.* cited studies, based on in-text citations and reference lists). They then categorised these studies based on their type (preprint, published article, existing guideline/HSG, or other literature (policy documents, news articles, testing standards, *etc.*)) and their study design (only for preprints and published articles). Due to the diversity of study types, the latter classification was first done manually by three researchers for a subset of the sample to ensure accuracy, after which categories were formed for use by two researchers in the remainder of the classification. This helped with both the accuracy of classification and consistency across the sample, and acted as a preparatory step that would facilitate subsequent analyses using the Joanna Briggs Levels of Evidence framework (74, 75). Discrepancies or issues at both stages were resolved through discussion.

3.2.4. Statistical analysis

Descriptive statistics were used to summarise the characteristics of the ECDC and WHO HSG, as well their recommendations and the supporting statements. Categorical variables were presented as frequencies and percentages, while continuous ones were summarised using MDs and IQRs.

3.2.5. Software

For the screening process, the ECDC records scraped from the organisation's website using *R*, version 4.2.1 (R Core Team, Vienna, Austria), and were exported and managed in Microsoft Excel, version 1808 (Microsoft Corp., Redmond, Washington, USA). The WHO records were imported into EndNote X9 (Clarivate, London, UK) and screened in the Rayyan screening management software (76). The data extraction process and levels of evidence evaluation were managed in Excel, and all statistical analyses were performed in jamovi, version 2.3.16 (jamovi project, Sydney, Australia).

3.3. Study 3: linguistic analysis of editorials

For the third study, editorial material related to the H1N1 and COVID-19 pandemics was extracted from bibliographic databases and screened for thematic eligibility according to pre-determined criteria. In terms of analyses, the Linguistic Inquiry and Word Count (LIWC) software was used to explore the sentiment of the retrieved editorials. The protocol for this study is available on the Open Science Framework (77). The PRISMA-S guidelines were followed in reporting the record extraction and screening process (65).

3.3.1. Data sources and search strategy

The WoS, MEDLINE (*via* WoS), and Scopus databases were searched up to 22 May 2024 using a search strategy developed in collaboration with a local medical librarian, comprising keywords related to H1N1, COVID-19, and the healthcare system response to the two crises. In detail, this included keywords related to nonpharmaceutical interventions; misinformation/disinformation; healthcare resource allocation and management; healthcare system preparedness; mandates/policies/guidance related to research, education, or new technologies; and lessons learned. While no specific language restrictions were set, the search was limited to the years 2009–14 for the H1N1 and 2019–24 for the COVID-19 pandemic in order to capture similar temporal periods, and filtered to "Editorial Material" in WoS and "Editorial" in Scopus to exclude other types of records. The full search strategy is available in Text S2 in the **Supplement**.

3.3.2. Screening process and eligibility criteria

Following deduplication, the titles/abstracts and full-texts of all records were screened in duplicate according to pre-determined, piloted eligibility criteria (Text S3 in the **Supplement**). As these editorials often did not have an abstract, an accelerated screening process was adopted, whereby any records for which discrepancies existed in the researchers' inclusion/exclusion decisions were included for full-text screening, at which point further disagreements were resolved through discussion.

In terms of eligibility criteria, editorials, viewpoints, or similar opinion pieces on the healthcare system response to either the H1N1 or the COVID-19 pandemic at the local, regional, national, or global level were included for analysis. Editorials not discussing the two pandemics or discussing clinical aspects thereof, original research articles and similar reports (inclusive of literature reviews), comments on the findings of a single study, reports on the

output of a single healthcare system-related project, audio-visual material, and editorials presenting articles in thematic issues were excluded.

3.3.3. Data extraction

The following metadata were extracted for each included editorial: title, authors, year of publication, and journal of publication. After article PDFs were retrieved automatically *via* EndNote, version X9 (Clarivate, London, UK), their text was extracted and cleaned using a Python script, removing any irrelevant content (titles, authorship bylines, boxes, images, tables, references, *etc*.) and retaining only the main text for analysis. In this process, data on the corresponding authors, funding, and conflicts of interest were extracted separately from the texts for further exploratory analyses.

3.3.4. Sentiment analysis

The LIWC-22 software was used for the sentiment analysis. The software, which has now undergone several iterations since its conception in the 1990s, utilises pre-determined, validated "dictionaries" (78, 79). These dictionaries are formed out of word categories, some of which are functional (*e.g.* pronouns, nouns, verbs), while others are subjective (*e.g.* words denoting positive or negative emotion, social processes). In the initial development process, the latter categories were built through a two-stage selection process, where two groups of three judges consecutively rated the words for inclusion; the final agreement rate after the second phase ranged from 93% to 100% (78). This process was repeated and built upon with the development of new versions of the software in 1997 and 2007 (78), and again in 2015 (80), with the latest, most comprehensive iteration being published in 2022 (79).

3.3.5. Statistical analysis

The LIWC scores were reported using MDs and IQRs due to the non-normal distribution of the data. For the same reason, the scores between the editorials published during the H1N1 pandemic and the COVID-19 pandemic were compared using Mann-Whitney's U test.

3.3.6. Software

The records retrieved from the databases were exported into EndNote, version X9 (Clarivate, London, UK) and afterwards managed in the Systematic Review Facility (SyRF), a web-based screening management platform developed by the Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies (CAMARADES) research group at the

University of Edinburgh, Scotland, UK (81). The text from the PDF versions of the included records was extracted using Python, version 3.13, *via* the Adobe PDF Extract application programming interface. The “regex”, “wordsegment”, “pandas”, and “json” packages were used to clean and manage the textual data at this stage. The LIWC-22, version 1.11.0 (Pennebaker Conglomerates, Austin, TX, USA) was used for the sentiment, while the NLP pipeline was set up in Python, version 3.13.

4. RESULTS

4.1. Study 1: AGREE-HS assessment

The initial query retrieved 59,641 records across all sources, with 23,844 remaining post-deduplication. Following the title/abstract, full-text, and AGREE-HS eligibility screening stages, 108 HSG remained for analysis (**Figure 2**; Table S2 in the **Supplement**).

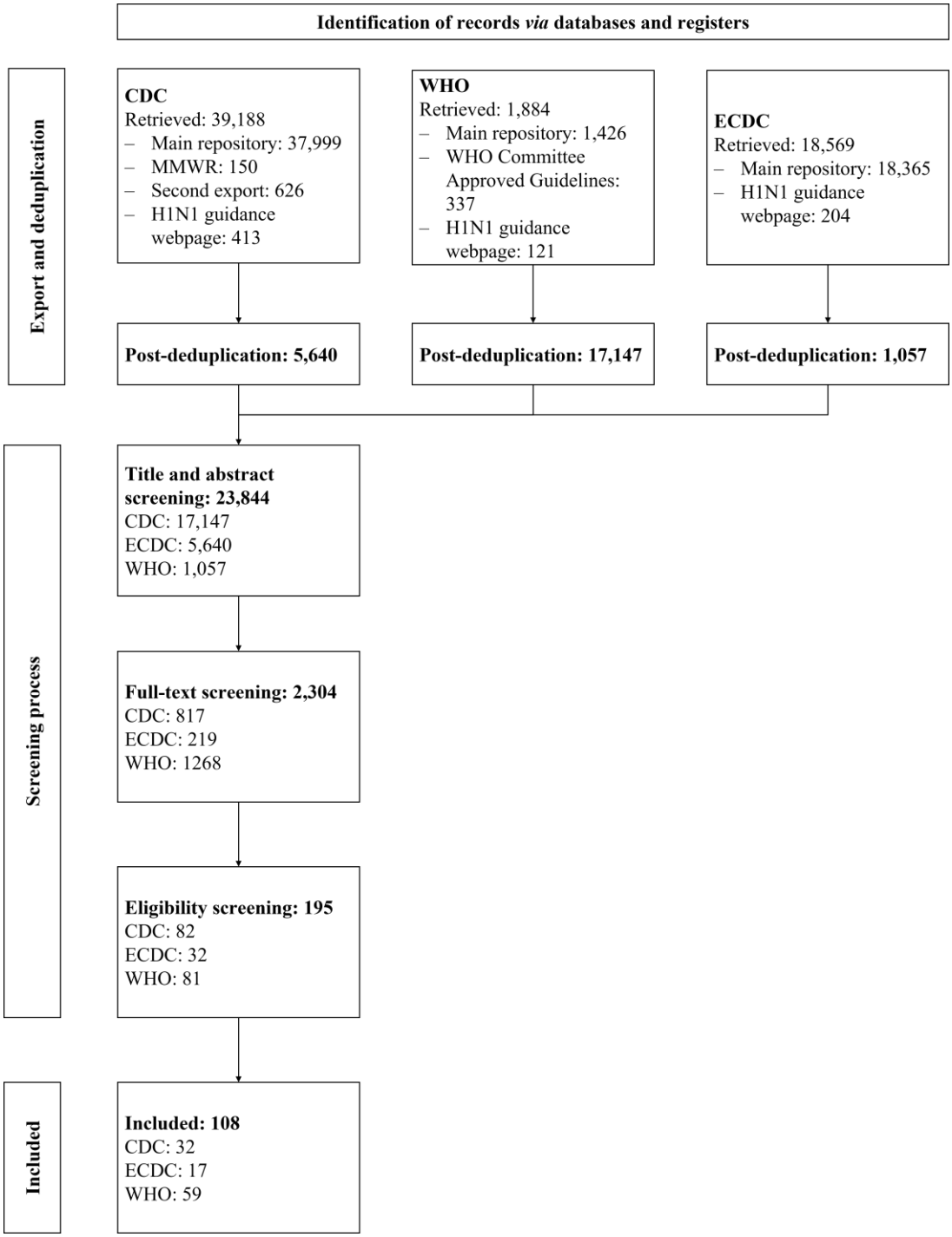

Figure 2. Record screening process. Adapted from related article (72) which was published in open access under CC BY 4.0 license. Reported according to PRISMA-S guidelines (65). CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, WHO – World Health Organization.

Most were produced for the COVID-19 pandemic (n = 92, 85.18%) and over half were developed by the WHO (n = 59, 54.63%). They were mainly global in scope, with the CDC and ECDC HSG expectedly relating to their national/regional contexts (**Table 3**).

Table 3. Characteristics of the included HSG

	Pandemic		**Geographical scope**	
Organisation	H1N1	COVID-19	Global	Regional or national
CDC	7	25	2	30
ECDC	2	15	0	17
WHO	7	52	59	0

CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease prevention and Control, WHO – World Health Organization

4.1.1. Overall and organisation-specific scores

The HSG were rated poorly in general, with the raw median score for the overall sample and for each organisation's HSG being below the midpoint of the total possible score (range = 10–70). The CDC HSG were rated more poorly than those of either the ECDC ($P < 0.001$) or the WHO ($P < 0.001$), which, in turn, did not differ significantly from one another (**Table 4**; Table S3 in the **Supplement**). However, the scores at the level of each HSG varied significantly across all three organisations, but especially the WHO, indicating significant differences in completeness of reporting (**Figure 3**).

Table 4. Overall AGREE-HS scores of HSG issued by the CDC, the ECDC, and the WHO

			P-value			
	Raw total score, MD (IQR)*	**Transformed total score, MD (IQR)†**	**Overall**	**CDC *vs.* ECDC**	**CDC *vs.* WHO**	**ECDC *vs.* WHO**
Overall	32.0 (26.0–42.0)	36.7 (26.7–53.3)				
Organisation			<0.001‡	<0.001‡	<0.001‡	0.455‡
CDC	24.0 (22.0–27.0)	23.3 (20.0–28.3)				
ECDC	35.0 (31.0–39.0)	41.7 (35.0–48.3)				
WHO	37.0 (32.0–45.5)	45.0 (36.7–59.2)				
Pandemic			0.015§			
H1N1	27.0 (25.8–30.0)	28.3 (26.3–33.3)				
COVID-19	33.5 (26.8–42.3)	39.2 (27.9–53.8)				

AGREE-HS – Appraisal of Guidelines for Research and Evaluation – Health Systems, CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, IQR – interquartile range, MD – median, WHO – World Health Organization

*Determined based on rating from both assessors (range = 10–70, midpoint = 40).

†Calculated *per* the formula outlined in the AGREE-HS manual (64): (obtained score − minimum score)/(maximum score − minimum score).

‡Kruskal-Wallis test for overall and Dwass-Steel-Critchlow-Fligner test for pairwise comparison.

§Mann-Whitney's U test.

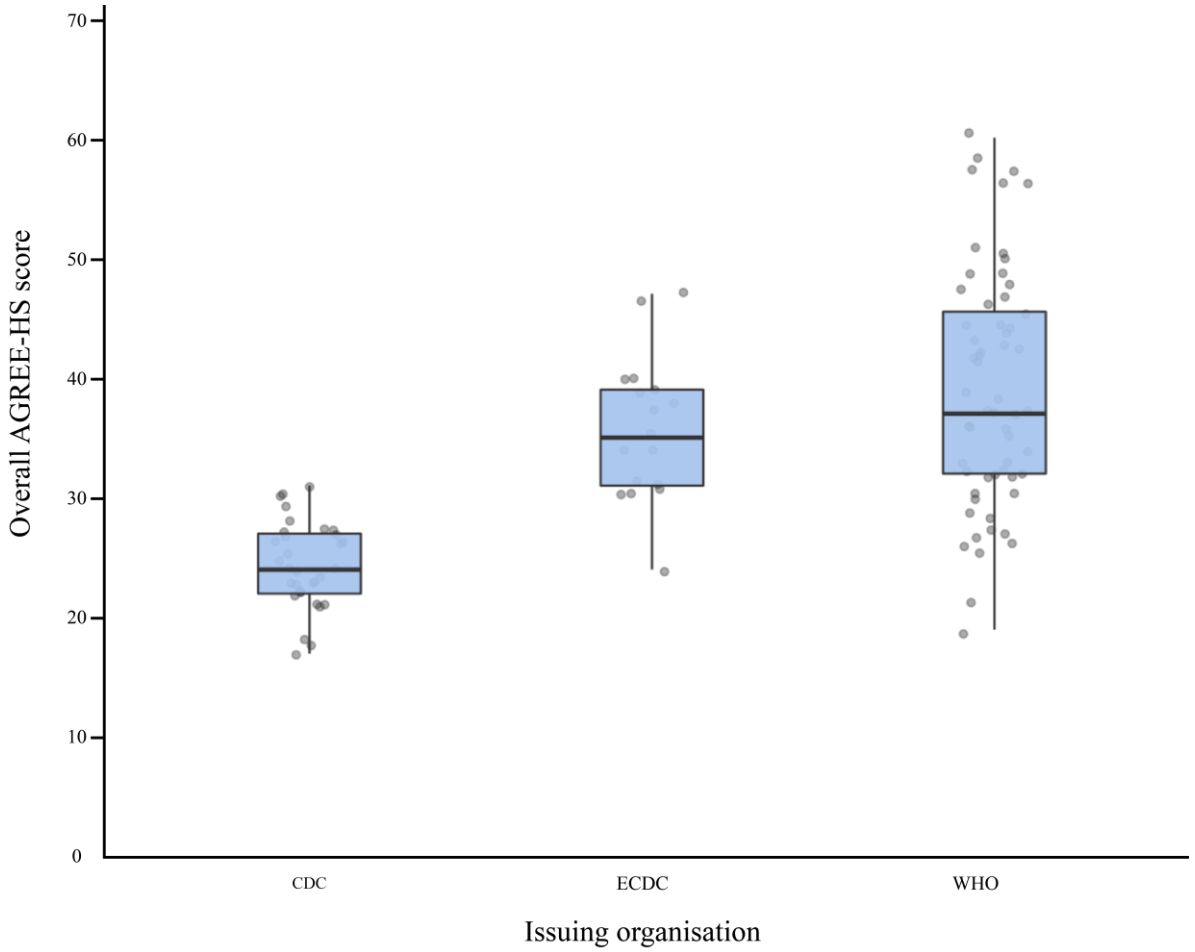

Figure 3. Overall AGREE-HS scores of HSG issued by the CDC, the ECDC, and the WHO. Reproduced from related article (72) which had been published in open access under CC BY 4.0 license. CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, WHO – World Health Organization.

In terms of individual domains, the HSG of all three organisations received low scores for reporting regarding their "Methods", "Participants", and "Implementability". In fact, only the ECDC and the WHO HSG received median scores at or above the midpoint of the scale (range: 1–7) for the "Topic" and "Recommendations" domains. Between the three organisations, the CDC HSG were rated more poorly across all five domains, while the ECDC and the WHO only differed on the "Participants" domain ($P = 0.010$), with the latter receiving marginally higher scores (**Figure 4**, **Table 5**).

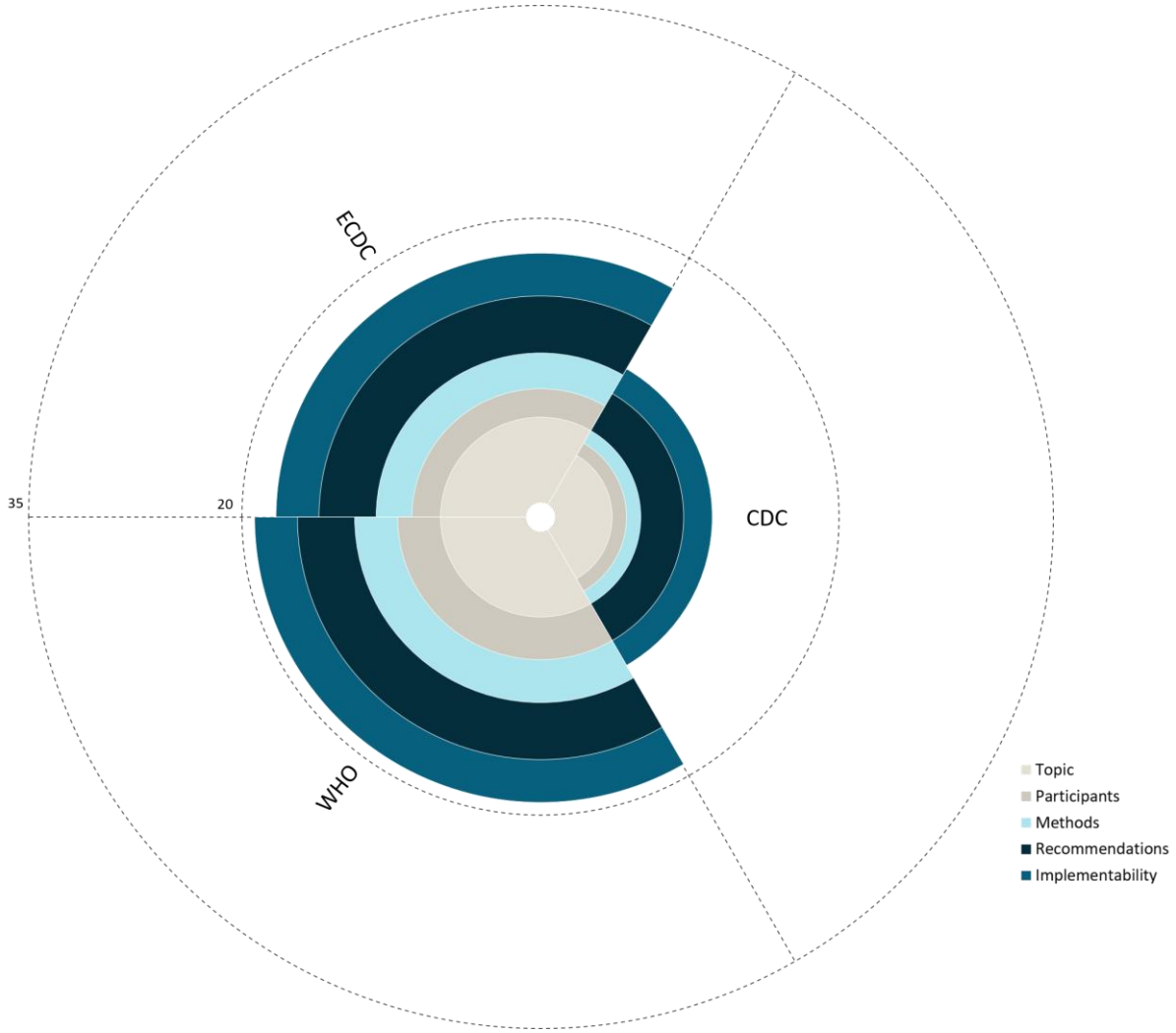

Figure 4. AGREE-HS domain scores for the HSG issued by the CDC, the ECDC, and the WHO. The inner and outer dashed circles represent the midpoint and the maximum total possible score, respectively. Reproduced from related article (72) which had been published in open access under CC BY 4.0 license. CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, WHO – World Health Organization.

Table 5. Assessment and comparison of AGREE-HS domains for the HSG issued by the CDC, the ECDC, and the WHO

	CDC, MD (IQR)	ECDC, MD (IQR)	WHO, MD (IQR)	*P*-value*			
				Overall	CDC *vs.* ECDC	CDC *vs.* WHO	ECDC *vs.* WHO
Topic	4.00 (3.00–6.00)	6.00 (4.25–6.00)	6.00 (5.00–7.00)	<0.001	<0.001	<0.001	0.801
Participants	1.00 (1.00–1.00)	2.00 (2.00–2.00)	3.00 (2.00–3.00)	<0.001	<0.001	<0.001	0.010
Methods	1.00 (1.00–2.00)	2.50 (2.00–3.00)	3.00 (2.00–4.00)	<0.001	<0.001	<0.001	0.534
Recommendations	3.00 (2.00–3.00)	4.00 (3.00–5.00)	4.00 (3.00–5.00)	<0.001	<0.001	<0.001	0.964
Implementability	2.00 (2.00–3.00)	3.00 (2.25–4.00)	3.00 (3.00–5.00)	<0.001	0.001	<0.001	0.587

AGREE-HS – Appraisal of Guidelines for Research and Evaluation – Health Systems, CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, IQR – interquartile range, MD – median, WHO – World Health Organization
*Kruskal-Wallis test for overall and Dwass-Steel-Critchlow-Fligner test for pairwise comparison.

4.1.2. Recommendations for use

The assessors mostly agreed on their decisions to recommend the HSG for use (**Table 6**). Looking at overlaps in their decisions, 56 (51.9%) of the HSG were not recommended for use by either assessor, while 24 (22.2%) were recommended either for immediate use (n = 4, 3.7%), unconditionally/given some modifications (n = 9, 8.3%), or exclusively following some modifications (n = 9, 8.3%). Most of the HSG recommended for use (either with or without modifications by one or both assessors) were developed for the COVID-19 pandemic (n = 18, 16.7%), with only two (0.9%) relating to the H1N1 pandemic. There were some discrepancies between the assessors, however, where a HSG was recommended for use by one, but not the other assessor (n = 3, 2.8%), or not recommended by one, but recommended by another following modification (n = 25, 23.1%).

When considering these recommendations at the level of each organisation, only six HSG produced by the CDC were recommended for use following modifications by one assessor, but not recommended for use by the other; the remaining 26 were not recommended for use by both assessors. Four ECDC HSG were recommended for use by at least one assessor (either unconditionally or following modifications) compared to 20 WHO HSG.

Table 6. Recommendations for use, n (%)

		Organisation			Pandemic	
Assessor recommendations	**Overall**	**CDC**	**ECDC**	**WHO**	**COVID-19**	**H1N1**
Yes/yes	4 (3.7)	0 (0.0)	0 (0.0)	4 (3.7)	4 (3.7)	0 (0.0)
Yes/yes, with modifications	11 (10.2)	0 (0.0)	4 (3.7)	7 (6.5)	10 (9.3)	1 (0.9)
Yes, with modifications/yes, with modifications	9 (8.3)	0 (0.0)	0 (0.0)	9 (8.3)	8 (7.4)	1 (0.9)
Yes/no	3 (2.8)	0 (0.0)	1 (0.9)	2 (1.9)	3 (2.8)	0 (0.0)
No/yes, with modifications	25 (23.1)	6 (5.6)	2 (1.9)	17 (15.7)	21 (19.4)	4 (3.7)
No/no	56 (51.9)	26 (24.1)	10 (9.3)	20 (18.5)	46 (42.6)	10 (9.3)

CDC – Centers for Disease Control and Prevention, ECDC – European Centre for Disease Prevention and Control, WHO – World Health Organization

Lastly, the "Participants" and "Methods" domains acted as predictors of a HSG being recommended for use by our assessors (P <0.001), with the latter having the most prominent influence in this relationship (**Table 7**; Table S4 in the **Supplement**). While the "Implementablity" domain also predicted these recommendations, we do note that this relationship was marginally significant (95% CI = 1.017, 2.00; P = 0.042).

Table 7. Predictors of HSG being recommended for use by study assessors

Predictor/domain	Estimate (95% CI)	SE	Z	OR (95% CI)	*P*-value
Topic	0.260 (−0.1090, 0.655)	0.194	1.34	1.30 (0.897, 1.93)	0.180
Participants	0.539 (0.2344, 0.858)	0.158	3.41	1.71 (1.264, 2.36)	<0.001
Methods	1.064 (0.6707, 1.496)	0.210	5.08	2.90 (1.956, 4.46)	<0.001
Recommendations	0.245 (−0.1519, 0.639)	0.201	1.22	1.28 (0.859, 1.89)	0.223
Implementability	0.349 (0.0167, 0.692)	0.172	2.04	1.42 (1.017, 2.00)	0.042

CI – confidence interval, OR – odds ratio, SE – standard error

4.1.3. Assessor's comments

The assessors left 1176 comments explaining their assessments. As their high numbers, varying length, and unstructured nature made a formal qualitative analysis unmanageable, only several comments were selected to contextualise the assessors' scores.

In rating the "Participants" item, the assessors mainly noted that the ECDC and the WHO failed to report on the GDG members' backgrounds, as well as the steps taken for the management of potential conflicts of interest or the influence of the funding organisation (Q1). They did, however, provide a list of the names and surnames of GDG members, in stark contrast to the practices of the CDC (Q2).

> ***Q1:*** *"The development team members are mentioned alongside their institution, but with insufficient data on their backgrounds or sectors, which prohibits us from determining their stake or contribution to the development process. There is also no data on their conflicts of interest or the influence of the funding organisation (or steps taken to limit it)."*
>
> ***Q2:*** *"No information is given for this [in the CDC HSG] whatsoever, so this AGREE item is completely not addressed. No [information regarding] funder influence is mentioned, nor are any potential conflicts of interest [disclosed]."*

In terms of the "Methods" domain, the assessors commented that the WHO and ECDC HSG cited literature reviews as the basis of their development process, without clarifying their screening process, search strategy, or other information that might be relevant for such an approach (Q3). This included a lack of reporting on approaches in evidence assessment approaches or in obtaining the GDG members' consensus on recommendations (Q4). The CDC HSG, meanwhile, lacked even the most basic data for this domain, which explains the assessors' low scores (Q5).

> ***Q3:*** *"The methodological basis is a literature review process and a discussion with experts from relevant fields, alongside the implementation of existing guidelines with robust methodological processes. However, more details could have been provided regarding the review – for example, the screening process, search strategy, etc."*
>
> ***Q4:*** *"Despite a mention of a robust/transparent methodological approach, it is actually not presented very well; the evidence [has been] composed [through] regular reviews and meetings of the development group, but we have no data on exact steps or consensus approaches. The evidence base, however, is robust and up-to-date, and linked to the recommendations, and shortcomings are discussed."*
>
> ***Q5:*** *"No methodological processes are presented in view of any review process for the evidence. There is also no mention of consensus in any form regarding the formulation of the recommendations (...) Evidence for the recommendations is not presented either, neither in the main guidance documents nor its annex."*

For the "Recommendations" domain, the assessors noted they were mostly clear and operationalisable. However, they observed that the HSG did not clearly present when, how, and by whom they would be updated (Q6). Regarding thresholds for implementing recommendations, they noted they were at times described in a non-specific, qualitative manner (Q7) and that they lacked considerations on their societal impacts or the infrastructure necessary for their execution (Q8).

> ***Q6:*** *"The recommendations are clear and succinct (...) annexes, accompanying guidance documents and prior versions contain enough information and clear-cut definitions (for example, concerning infection rates/levels of transmission in different settings) to make the guidance operationalizable. The only reason this guidance was not evaluated with the grade for highest quality is the plan for updating the recommendations, which is vague (i.e., only mentions that it will be updated when new knowledge emerges, without describing how or why whom)."*
>
> ***Q7:*** *"Only qualitative descriptors are given regarding the specific outcomes expected from implementing the recommendations (...) no "end-point" is described with a specific threshold."*
>
> ***Q8:*** *"However, there is a lack of discussion on the impact of these measures on society as a whole, including the resources needed for implementing whole-of-society testing strategies, contact tracing, etc."*

The HSG's shortcomings in the "Implementability" domain were related mainly to the absence of discussions on the affordability/cost-effectiveness of the measures, as well as their sustainability in practice (Q9, Q10).

> ***Q9:*** *"While barriers/enablers are discussed extensively, especially because of limited evidence, there is a lack of discussion (at least an extensive one) about the costs of interventions. This also affects transferability aspects, as low-income settings might not have substantial resources to dedicate to proper masking measures or public health masking."*
>
> ***Q10:*** *"The same is applicable to discussions regarding the sustainability of the travel-related measures; while it is mentioned and shortly discussed, such discussions warrant more detail. This could be done by projecting specific costs, giving thresholds/expected outcomes, and through similar means."*

4.2. Study 2: levels of evidence underlying the HSG

The updated search retrieved 11,375 records, all of which entered our screening process. Twenty-six were eligible for inclusion following the title/abstract and full-text screening, while another 13 were found through a review of their references. After this was added to the initial sample of 65 HSG from the first study, 66 prior versions of the included HSG were retrieved manually. Following deduplication, 170 HSG (with all versions included) or 78 unique HSG (excluding prior versions) remained for analysis (**Figure 5**).

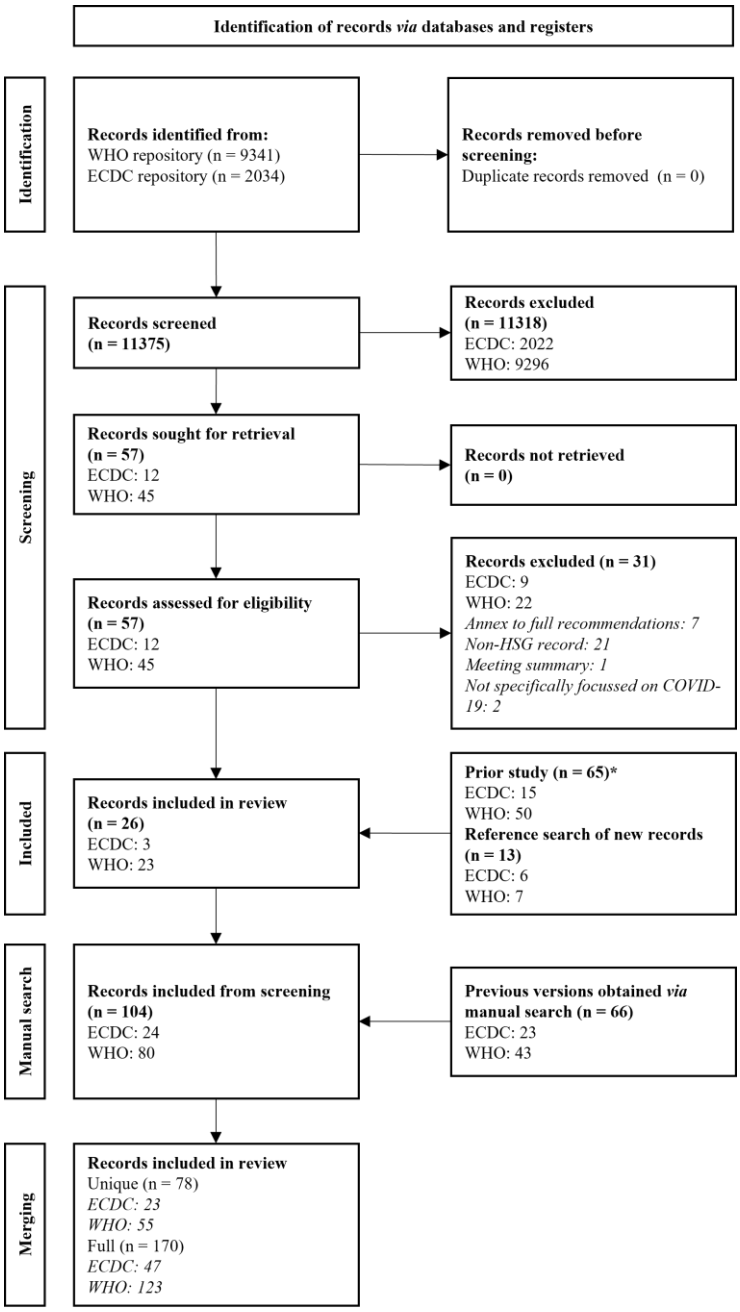

Figure 5. Flowchart for study of levels of evidence behind pandemic HSG. Reported according to PRISMA-S guidelines (65). ECDC – European Centre for Disease Prevention and Control, WHO – World Health Organization. *Two records excluded from initial study sample, as they did not present their evidence base.

4.2.1. Characteristics of HSG, their recommendations, and their supporting statements

This preliminary analysis included 17 HSG (10 unique records with 7 updates) published by the ECDC between 7 February 2020 and 28 January 2022. Six records were fully unique, meaning they were issued only once and never updated. The median update time was 185 days or approximately six months. Regarding the recommendations within the HSG, they were all exclusively informal, meaning that they took the form of actionable recommendations with defined populations, interventions, and (in applicable cases) comparators, but were not presented as if they were a product of systematic, deliberative evidence assessment processes. These recommendations were supported by 665 statements in the HSG citing 812 unique references (**Table 8**). As the analytical sample is small, inferential statistical analyses by pandemic period were not performed here; rather, only basic descriptive statistics are shown.

Table 8. Characteristics of included HSG and their recommendations*

Characteristic	
Without update	6
With update	4
Number of updates per HSG, range	2–4
Updated once	2
Updated twice	1
Updated thrice	1
Days between updates, MD (IQR)	185 (45–469)
Period of issuing†	
First	7
Second	6
Third	2
Fourth	2
Recommendations	
Formal	0
Informal	322
Number of statements	665
Number of unique references	812

HSG – health systems guidance, MD – median, IQR – interquartile range
*Values are presented as raw frequencies unless specified otherwise.
†Each period spanned 185 days from the publishing of the first HSG in the sample (7 May 2020) to the last one (28 January 2022), as this corresponded to the median update period for the HSG in our sample.

4.2.1. Classification of the evidence underlying the recommendations

Of the 812 unique references, over half (n = 461, 56.8%) were journal articles or preprints, with the latter accounting for just under one-fifth of this subsample (n/N = 81/461, 17.57%). Excluding bench research, just above half of the studies (n/N = 249/461, 54.01%) were observational in design. Systematic reviews, with or without meta-analyses, accounted for only one-tenth of the sample (n/N = 49/461, 10.62%), while only six (1.3%) randomised controlled trials (RCTs) were referenced across all 17 HSG (**Table 9**).

Table 9. Characteristics of the studies cited in support of the recommendations

	n (%)
Publication status	
Published in a journal	380 (46.8)
Preprint	81 (10.0)
Other guideline/HSG	103 (12.7)
Other literature*	248 (30.5)
Study design	
Viewpoints and non-structured literature reviews	63 (13.7)
Bench research	25 (5.4)
Case study	25 (5.4)
Case series	55 (11.9)
Modelling study	68 (14.8)
Qualitative study	1 (0.2)
Cross-sectional study	78 (16.9)
Cross-sectional study with modelling component	7 (1.5)
Cohort study	67 (14.5)
Cohort study with modelling component	8 (1.7)
Case-control study	9 (2.0)
Randomised controlled trial	6 (1.3)
Systematic, rapid, narrative, and literature reviews without synthesis or meta-analysis	25 (5.4)
Systematic review with meta-analysis or synthesis	24 (5.2)

HSG – health systems guidance
*Includes policy documents, risk assessments, risk reports, diagnostic test evaluations/standards, tools/checklists, news articles, etc.

Looking at preprints specifically, just over one-third (n/N = 29/81, 35.80%) reported on modelling analyses, while just under half (n/N = 37/81, 45.67%) described observational studies (**Table 10**). Stratified by periods of approximately six months (*i.e.* 185 days, the median update time of the HSG in our sample), an increase can be observed in the proportion of peer-reviewed literature and preprints referenced in the HSG. By study design, there is a drop over time in the proportion of viewpoints and non-structured literature reviews, case series, and cross-sectional studies, and an increase in that of cohort and case-control studies and RCTs. There is likewise an increase in the proportion of systematic reviews with meta-analyses and syntheses as to those without, except for a gap in the third period, where we find only one cited review (Table S5 in the **Supplement**).

Table 10. Characteristics of the studies cited in support of the recommendations, n (%)

Type of study	Published in a journal	Preprint
Viewpoints and non-structured literature reviews	63 (13.7)	0 (0.0)
Bench research	21 (4.6)	4 (0.9)
Case study	25 (5.4)	0 (0.0)
Case series	52 (11.3)	3 (0.7)
Modelling study	39 (8.5)	29 (6.3)
Qualitative study	1 (0.2)	0 (0.0)
Cross-sectional study	62 (13.4)	16 (3.5)
Cross-sectional study with modelling component	7 (1.5)	0 (0.0)
Cohort study	53 (11.5)	14 (3.0)
Cohort study with modelling component	8 (1.7)	0 (0.0)
Case-control study	5 (1.1)	4 (0.9)
Randomised controlled trial	4 (0.9)	2 (0.4)
Systematic, rapid, narrative, and literature reviews without synthesis or meta-analysis	25 (5.4)	0 (0.0)
Systematic review with meta-analysis or synthesis	15 (3.3)	9 (2.0)

4.3. Study 3: linguistic analysis of editorials

Our initial search retrieved 17,150 records, of which 1,781 were removed through deduplication. Following the title/abstract and full text screening, 2,954 records were left for analysis (**Figure 6**). Here we present a preliminary analysis of 200 editorials – 175 from the COVID-19 and 25 from the H1N1 pandemic.

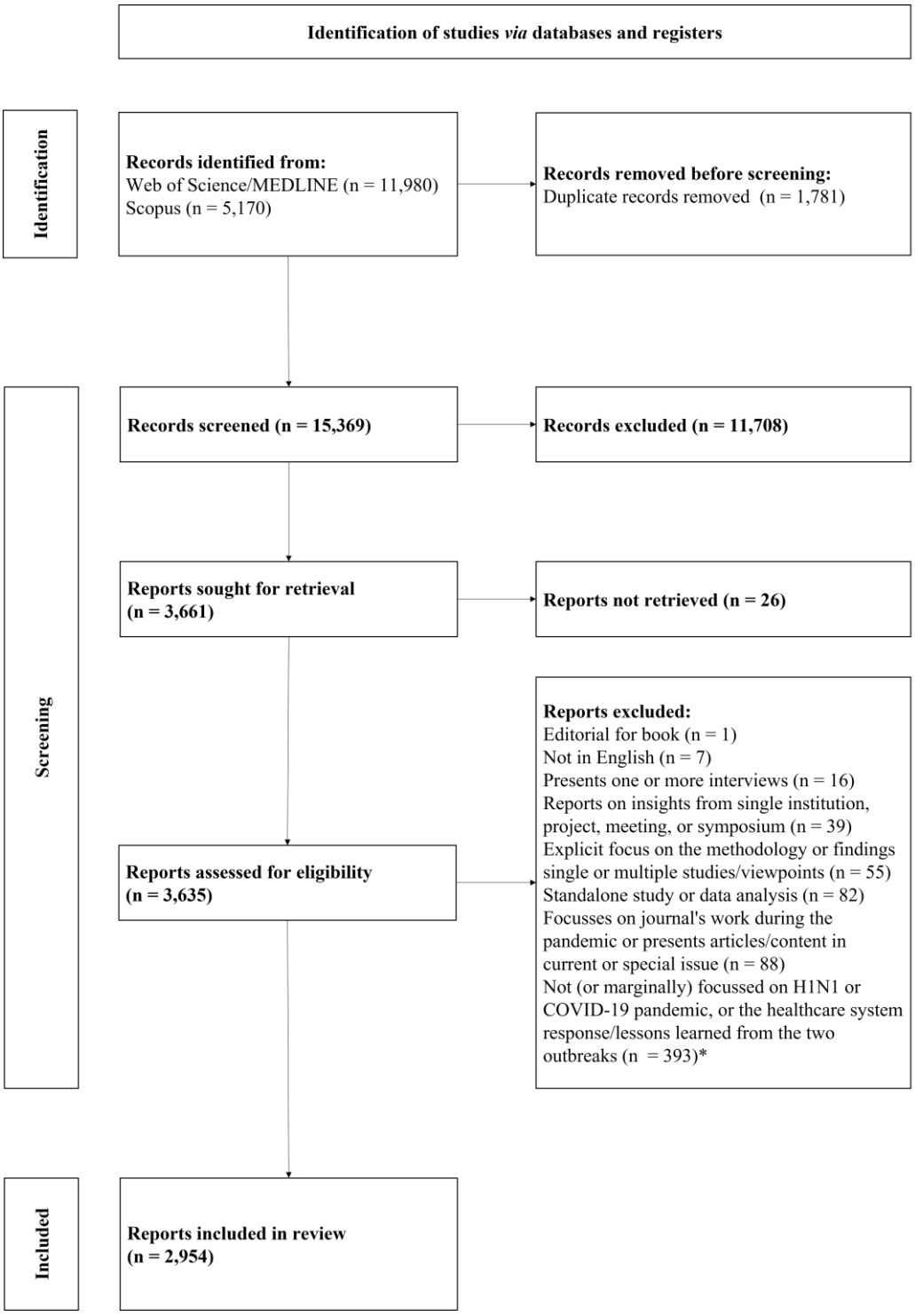

Figure 6. Flowchart for study on the linguistic analysis of editorials. Reported according to PRISMA-S guidelines (65).

4.3.1. LIWC analysis

The editorials for the COVID-19 pandemic had more words than those published during the H1N1 pandemic (MD = 1332, IQR = 947–1726 *vs*. MD = 838, IQR = 667–1207; $P < 0.001$). For the main summary variables, there were no differences in the proportion of words connoting analytical thinking, clout, or authenticity. There was a significant difference in tone, where the editorials related to COVID-19 had higher values than those on the H1N1 pandemic (MD = 28, IQR = 17.1–37.2 *vs*. MD = 15.6, IQR = 7.97–25.4; $P = 0.004$), indicating they had a more positive tone.

For the variables of interest, the editorials published during the H1N1 pandemic had higher values for negative tone (MD = 2.21, IQR = 1.74–3.07 *vs*. MD = 1.69, IQR = 1.22–2.31; $P = 0.007$) and certitude (MD = 0.390, IQR = 0.165–0.655 *vs*. MD = 0.230, IQR = 0.120–0.370; $P = 0.010$). For personal language, the only significant differences were found in the sentiment of need, which was more pronounced in COVID-19 editorials than those for the H1N1 pandemic (MD = 0.945, IQR = 0.617–1.23 *vs*. MD = 0.615, IQR = 0.402–0.818; $P = 0.010$). There were no differences in the proportion of words denoting social processes, negative emotion, analytical thinking, or all-or-none thinking (**Table 11**).

Table 11. Full results of preliminary LIWC analysis

LIWC variables	COVID-19	H1N1	Statistic	*P*-value
Word count	1332 (947–1726)	838 (667–1207)	1104	<0.001
Analytical thinking	92.8 (89.1–95.4)	90.3 (85.3–95.1)	1651	0.083
Clout	43.2 (35.4–51.2)	41.9 (35.2–50.2)	2080	0.904
Authentic	32.9 (21.8–45.7)	31.3 (25.0–43.3)	2060	0.846
Tone	28.0 (17.1–37.2)	15.6 (7.97–25.4)	1336	0.004
All-or-none thinking	0.340 (0.210–0.560)	0.400 (0.150–0.622)	2111	0.997
Certitude	0.230 (0.120–0.370)	0.390 (0.165–0.655)	1429	0.010
Negative tone	1.69 (1.22–2.31)	2.21 (1.74–3.07)	1391	0.007
Negative emotion	0.220 (0.110–0.403)	0.155 (0.0525–0.343)	1732	0.153
Social processes	7.20 (6.12–8.66)	6.82 (5.69–10.4)	2064	0.857
Social behaviour	3.10 (2.28–4.23)	3.55 (2.34–4.68)	1928	0.490
Prosocial behaviour	0.810 (0.458–1.34)	0.800 (0.425–1.01)	1888	0.400
Interpersonal conflict	0.140 (0.0600–0.263)	0.255 (0.00–0.455)	1846	0.315
Moralisation	0.120 (0.00–0.292)	0.155 (0.00–0.445)	2079	0.900
Need	0.945 (0.617–1.23)	0.615 (0.402–0.818)	1431	0.010
Want	0.00 (0.00–0.0625)	0.00 (0.00–0.128)	1884	0.311

LIWC – Language Inquiry and Word Count

There were noticeable differences in sentiment subcategories and categories associated with those tested for the hypothesis. For example, the editorials published for the H1N1 pandemic had higher values for cognition (MD = 21.2, IQR = 19.3–22.5 *vs*. MD = 18.0; $P < 0.001$), which is an overarching category for the sentiment of certitude. In terms of affect (an overarching category for positive/negative emotion and tone), they had more words connoting positive emotion (MD = 0.145, IQR = 0.108–0.235 *vs.* MD = 0.0850, IQR = 0.00–0.163; $P = 0.007$) and fewer connoting anger (MD = 0.00, IQR = 0.00–0.0900 *vs*. MD = 0.00, IQR = 0.00–0.00). The results for all LIWC categories are presented in Table S7 in the **Supplement**.

5. DISCUSSION

5.1. Summary of findings

Through three studies, this dissertation sought to provide a critical assessment of HSG for the H1N1 and COVID-19 pandemics issued by three globally relevant organisations, and to analyse the discourse among scholars that underlay the healthcare system response to these two crises.

The first study assessed the transparency and completeness of reporting in HSG issued by the WHO, the CDC, and the ECDC using the AGREE-HS tool. Overall, the HSG had significant shortcomings in reporting, but especially in how development processes and details on GDG members (including their CoIs) were presented. This was especially true for the HSG issued by the CDC, which were rated more poorly than those issued by the other two organisations. These two factors, captured by the tools' "Methods" and "Participants" domains, were significant predictors of an HSG being recommended for use by our assessors, which is likely why half of those included in our sample were not recommended for use.

The second study looked at the levels of evidence underlying the recommendations in these HSG. A preliminary analysis of 17 HSG issued by the ECDC showed that they exclusively contained informal recommendations, *i.e.* actionable statements with defined populations, interventions, and (if applicable) comparators, but without a clearly indicated strength, direction, and certainty of evidence, and without clear elaboration of whether they resulted from a deliberative, structured development process backed up by systematic reviews and evidence syntheses. The studies supporting these informal recommendations were mainly observational in design or based on statistical modelling, with one-fifth being published as preprints only. While an increase in the proportion of cohort and case-control studies and systematic reviews with meta-analyses can be seen in the later pandemic period, conclusions on an increase or drop in the levels of evidence can only be made once the full analysis is completed.

The preliminary results from the sentiment analysis of editorials published on the healthcare systems' response to pandemics partially contrasted the study hypothesis, where the editorials published for the H1N1 pandemic were more negative in tone and displayed more certitude than those published on the COVID-19 response. For personal language, the latter only had a higher proportion of words reflecting the sentiment of need; while this aligns with our hypothesis, there were no other differences for similar sentiment domains (*e.g.* social processes). The two groups of editorials did not otherwise differ in terms of negative emotion, analytical thinking, or all-or-none thinking.

5.2. Transparency and completeness of reporting in HSG: implications and a need to establish standards

The findings of the HSG assessment align with those of similar evaluations on CPGs issued during outbreaks. An AGREE II-based investigation of emergency CPGs issued by the WHO for four different outbreaks noted that they did not adequately present how they were developed or disclose the GDG members' CoIs (58), while analogous analyses of national and global CPGs issued early during the COVID-19 pandemic arrived at similar findings (40). These trends reflect those observed outside of outbreak-specific contexts, where a 2010 systematic review of 42 appraisal studies spanning two decades and encompassing 626 CPGs found moderate or low mean scores for their rigour of development, editorial independence, stakeholder involvement, and applicability (82). While the authors of this study observed improvement for all assessed AGREE-HS domains except for the one reflecting the independence of the GDG, they noted that the overall quality of CPG remained concerningly inadequate. These findings were echoed by a 2017 review of 415 CPGs published between 1992 and 2014 (83).

While CPGs have been studied extensively, as seen from the above-referenced research, we encounter few systematic analyses of HSG. Those published to date that have utilised the AGREE-HS tool applied to our sample arrived at similar findings, despite using a different set of HSG. In 2019, the team that developed the AGREE-HS tool assessed 85 HSG published mainly by the National Institute for Health and Care Excellence and the WHO between 2012 and 2017. They noted inadequate reporting on the methods used to produce the HSG; the composition of the GDG, their CoIs, and steps taken to mitigate the funder's influence on the development process; and the strategies needed for implementing the recommendations (49). A 2022 evaluation of 13 HSG for mental health and psychosocial support drew similar inferences (59), as did a 2023 scoping review and appraisal of 10 HSG on the responsibilities of pharmacists in dispensing opioids (84).

Our findings, however, partially contrast those of a recent preprint reporting on an AGREE-HS assessment of CPGs, HSG, and integrated guidelines issued by the WHO, wherein the assessors rated the included HSG or the HSG-related sections of integrated guidelines higher than we did in our study across most domains, and leaned more towards recommending them for use (62). This discrepancy could be related to Zhang and colleagues' inclusion of guidelines from other outbreaks and from later phases of the COVID-19 pandemic, which might have been better developed than the interim HSG from the early pandemic period we included

in our assessment. Lastly, the authors of an analysis focussed exclusively on the emergency response plan to the Nipah virus outbreak in Kerala, India, deemed it to be of moderate quality overall, but added that it "fared poorly on ethical considerations, updates, coverage of certain critical aspects, outcome measures, and evaluation" (61). It is difficult to compare this study to the analysis presented here, however, as its authors assessed only a single HSG from a specific national context.

The research presented above, which is mainly in concert with ours in terms of methods and findings and which likewise included (exclusively or not) WHO-developed HSG points to general gaps in reporting on the process of guideline development at leading healthcare organisations, but especially those related to development processes, GDG compositions, and GDG members' independence from external influence. This is concerning, given that such issues could affect guideline uptake: in the context of CPGs, for example, a 2015 realist literature review found "Stakeholder involvement" and "Evidence synthesis" to be two of six domains influencing guideline uptake (85). Similarly, a scoping review exploring the same topic in the context of LMICs concluded that the "attributes" of CPGs, including their perceived credibility (*i.e.* use of current evidence) and adaptability to local contexts, affected guideline uptake (86).

Our study likewise found these factors to be predictors of an HSG being recommended for use in our analysis, which is why they could be the first "targets" for interventions and improvement initiatives. The development and uptake of tools and checklists such as the AGREE-HS, whose target audience are HSG developers and whose goal is to provide a "methodological framework for developing and reporting HSG" (64), could be a good first step towards this goal. This had been achieved in biomedical research through the development of reporting guidelines – an initiative advanced by the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network (87) – which were shown to be effective even when their uptake is not ideal (88–91). Yet while many such tools exist for CPGs (92), to our knowledge, AGREE-HS seems to be the only one to focus on HSG specifically.

Taken as a whole, these findings would indicate that research on HSG is still in its early stages. Considering their high impact on policymakers and (consequently) health systems, and the identified need for a tool as the AGREE-HS during its development (69–71), we call for further designing, testing, and refinement of interventions that would improve the reporting of HSG development.

5.1.1. Strengths and limitations

The first strength of this study lies in our comprehensive querying of the organisations' repositories, which was supplemented with manual searches of their websites. The second strength is related to the screening being performed by two researchers, which ensured adherence to the pre-determined eligibility criteria and the inclusion of relevant records (93). This was extended to the AGREE-HS assessment, as each HSG (following random assignment) was assessed by two individuals who were trained in its use, ensuring uniformity and robustness of this process. Lastly, the inclusion of HSG published by three high-level organisations (the WHO, the ECDC, and the CDC) for two pandemics (COVID-19 and H1N1) lends relevance to our findings in the context of global health.

Some limitations, however, need to be noted. The limited searchability of the organisations' repositories and websites might have led to potentially relevant records not being captured, especially if they were not properly indexed. This could explain the low number of records retrieved for the H1N1 pandemic, which is another shortcoming of our analysis. For the AGREE-HS assessment itself, it should be noted that, as is true for all such approaches (68), it is inherently subjective and relies on the assessors' backgrounds, knowledge, and other factors that cannot be accounted for. This is also important for our use of two assessors only – while on the lower end of the recommended range of two to four assessors (64), this approach should have been methodologically sufficient, as noted by the AGREE team (68). Lastly, this analysis only included HSG published by 17 May 2022, *i.e.* in the early stages of the COVID-19 pandemic. They all contained interim recommendations and were therefore based on limited evidence and produced under significant time pressure, which might explain the low scores we found for the "Methods" domain. Yet it should be noted that guidelines published during crises should and can still adhere to high standards and be produced through rigorous methodologies, irrespective of the urgent nature of outbreaks and healthcare crises (58, 94).

5.3. Evidence base for pandemics: lessons learned for future outbreaks

As the relatively small sample size of 17 HSG issued by one organisation only prohibited the conduct of any inferential analyses, firm conclusions on the levels of evidence cannot be drawn from the findings presented here. Initial observations indicate positive changes in the levels of evidence towards the later pandemic periods, which would contrast our hypothesis. This, however, can only be confirmed through inferential statistics on the whole sample.

More discernible, however, is the large proportion of modelling and observational studies supporting the informal recommendations in these 17 HSG, as opposed to RCTs or systematic reviews. While these types of studies rank lower on evidence hierarchies in general, it is important to note that the distinction of lower- or higher-quality evidence is not absolute when it comes to health system or population-level interventions. For example, experts who used the GRADE approach in systematic reviews of complex interventions found that they often had to downgrade the best or only available evidence from non-randomised or observational studies to lower levels, which consequently affected their final recommendation and thus negatively impacted decision-making (50). This was also true in two studies reporting on interviews and discussions with experts developing systematic reviews and guidelines for public health interventions, who noted that the GRADE framework did not allow them to distinguish between types of observational studies when assessing the strength of evidence, or to consider non-epidemiological evidence in cases where it might be relevant (51, 95). In another qualitative study that explored how the WHO utilises evidence in developing recommendations, GDG members saw the "issue of prioritisation of RCT evidence" to be problematic in cases where they might have deemed other types of non-randomised studies to be important, and encountered challenges when assessing evidence on cost-effectiveness and affordability of interventions (96).

Some have thus argued that the use of observational and non-randomised studies, therefore, might not be unexpected in the context of healthcare system interventions, as they might offer better or more adequate evidence than RCTs (97). Speaking of COVID-19 specifically, others have noted that RCTs on public health interventions such as nonpharmaceutical measures, for example, were unfeasible or even unethical to conduct (98). These positions have been countered by researchers citing successful pragmatic RCTs conducted during the pandemic (99). Despite this, it is worth noting that evidence mapping research in specific countries such as the UK found a lack of RCTs and longitudinal studies,

which “has limited researchers’ and policymakers’ ability to assess the interventions and policies implemented” (100, 101). As suggested by Duval and colleagues (100) and by our preliminary findings, this would suggest that a better, more robust evaluation methodology should be built into public health interventions during future outbreaks during their design and implementation phases.

Of similar concern is the proportion of preprints cited in support of the recommendations in the HSG. As non-peer-reviewed scientific research, they are not supposed to inform guidelines, which is even disclosed explicitly on preprint servers such as medRxiv (102). Despite this, preprints have drawn massive media and public interest during the early stages of the COVID-19 pandemic (103, 104), indicating that they are also likely to affect public opinion. The rate of citing preprints in guidelines observed here is similar to that of just under 20% found by Fraser and colleagues in COVID-19 policy documents (104). Given the lack of clearly established, universal standards for the use of preprints for such purposes, we align with the proposals by Ravinetto and colleagues (102), who have called for consultations with stakeholders such as the WHO or the Committee on Publication Ethics with the aim of outlining “clear principles and policies for the publication and dissemination of non-peer-reviewed research results” and their further dissemination to the research community.

5.2.1. Strengths and limitations

One major strength of this study comes from the screening, data extraction, and recommendation/study classification being done in duplicate based on pre-established, piloted criteria developed in collaboration with a third researcher with expertise and experience in evidence-based medicine. This reduced the chance of records being overlooked during screening (93) or recommendations/studies being misclassified, lending robustness to our sample and the results. A second major strength is the updated search, which extended to 18 July 2024, well beyond the end of the COVID-19 PHEIC on 5 May 2023 (105), allowing us to encompass the whole pandemic period in our sample. The third strength of this study lies in the use of the validated, consensus-based framework proposed by Lotfi and colleagues to identify and classify recommendations from the HSG (73).

The main limitation, meanwhile, lies in our exclusion of HSG issued for the H1N1 pandemic and those issued by the CDC. This decision was informed by our first study (72), where we noted that these two groups of HSG did not clearly present their evidence base, preventing us from conducting this analysis. Furthermore, it is possible that, due to their phrasing, recommendations could have been classified as “informal”, while others might

understand them as “good practice statements”. The difference between the two, as outlined in the framework by Lotfi and colleagues (73), lies in the latter having clear-cut, positive net benefits, because of which a formal evidence synthesis would be unproductive for the GDG. However, Guyatt and colleagues previously noted that perceptions of this net benefit can be subjective, especially if a clear chain of evidence for a good practice statement is not given (106). This could have, therefore, affected our sample to some extent, as good practice statements might have been classified as informal recommendations and included in our analysis, despite our classification process being performed in duplicate. However, such cases likely only occurred because this “chain of evidence” was not clearly presented in any case. Lastly, the classification of the study designs was done manually, meaning that errors could have occurred in some cases. However, this manual approach is only a preparatory step for further analyses based on the Joanna Briggs Institute’s levels of evidence framework, which will allow for more accuracy and nuance in interpreting the findings (74, 75).

5.4. Scientific communication: (im)partiality in a time of disinformation and misinformation

Owing to their epistemic authority, experts – medical practitioners, scientists, and others – advise policymakers during healthcare crises, and can consequently have a significant impact on society (107). For this reason, and due to publishers’ and individual researchers’ unprecedented endorsement of specific candidates during the pandemic (108–110), experts’ direct criticisms of governmental responses (111), and even observed bias towards “an aggressive approach to COVID-19 mitigation” by outlets such as the BMJ (112), a hypothesis was set that the editorials and opinion pieces published in scholarly journals during the COVID-19 pandemic would be more negative in tone, less analytically driven, and more reflective of extreme, all-or-none thinking, and certitude.

While this analysis is too preliminary to draw any inferences, the finding that editorials published for the H1N1 pandemic were more negative and displayed more certitude is, at first, surprising, given these discussions. This could be explained by controversies related to the WHO’s declaration of the H1N1 pandemic and subsequent inquiries, which were extensively debated and documented in the scholarly community (12, 13, 28, 30–32, 35). These assumptions will be explored in further natural language processing analyses and related to the findings regarding the texts’ sentiment.

There are no direct thresholds to which these results can be compared, as published research mainly focussed on social media and news outlets, and exclusively on the COVID-19

pandemic. This focus is not unusual, given the increased social media presence of public health agencies and policymakers during this period, as observed in the USA (113). Looking at sentiment specifically, one mainly finds studies on datasets of posts made by the general public. One global analysis of 20 million tweets found high rates of sadness and anger early in the COVID-19 pandemic (114). Topic modelling analyses of similar corpora from Twitter found pronounced negative sentiment for tweets on vaccination symptoms/side effects, conspiracies related to vaccination and the source of the outbreak, or the pandemic's sociopolitical impact, and positive sentiment for its effect on the economy and healthcare, governmental responses, and vaccine effectiveness (115, 116). A mixed-methods analysis of Telegram chat groups debating the use of the green pass in Italy found negative sentiment in the group opposing its implementation; these discussions revolved around the relationship between the pass and vaccination in the context of legal and personal freedoms (117).

For experts, researchers, and medical practitioners, research has adopted different natural language processing methods such as topic modelling. Examples from the early pandemic period include content analyses of tweets shared by Texas public agencies (118), the Ministry of Health of Saudi Arabia (119), and decision-makers or public health agencies in Canada (120). Of relevance to this research and the role of experts in discussions on healthcare response, Drescher and colleagues observed that COVID-19-related tweets by German experts, as opposed to authorities, were more likely to be retweeted and liked by social media users (121). Such data is concerning, given that one study suggested that 52 physicians in the USA with a large number of followers across different platforms shared COVID-19 misinformation on social media (122), while another study of 262 tweets made by credentialed users on monkeypox found that four-fifths contained inaccurate information and overexaggerated the risk of the disease (123). The latter research also noted that, when accounting for the number of followers, the public was 974 times more likely to encounter inaccurate than accurate information on the platform (123).

The integrity of research centres on five key values: objectivity, honesty, openness, accountability, fairness, and stewardship (124). Their infringement, especially that of objectivity (where research should not be influenced by bias and motivations) and stewardship (active care for the research ecosystem), can introduce doubt about the perceived reliability of science in society (124). In this sense, the concept of stewardship presumes that resources dedicated to science often depend on "public demand, political considerations, concerns about national security, and even the prospects for our species' survival" (124).

Yet this perception – that science is affected by political or social circumstances – can significantly affect societal behaviour and trust in research. The results of four survey waves from 12 countries with 54,000 respondents worldwide showed that compliance with nonpharmaceutical measures during COVID-19 and willingness to get vaccinated were heavily related to trust in science, even more so than to trust in governments (125). Similarly, a survey of 12,037 individuals conducted in the USA during the early pandemic found that one-third believed science to be politically motivated, *i.e.* politicised – a belief which was associated with lower risk assessments regarding the disease's seriousness (126). Other research found that perceptions of science being polarised during the pandemic were associated with individuals' political orientation, which likewise affected their adherence to prevention measures (127, 128). Another global survey of 71,922 individuals in 2025 found trust in science to be moderately high overall, but noted it might be dependent on a country's political leadership, rather than individuals' political orientation (129).

Seeing the significant attention that publications by editors, journalists, and columnists attracted during the COVID-19 pandemic (130), there seems to be a need for further research into this type of unstructured, opinion-based communication, given its potential societal impact. While criticism from experts is always necessary, it should remain maximally fair, open, and objective, as well as neutral and unbiased. In this sense, editors and authors should be aware of the language they use when communicating in scholarly journals, as it should not exacerbate the negative sentiment and politicisation that is likely to occur in society during crises such as disease outbreaks. We hope that our study sets a benchmark for such investigations and serves to inform the development of editorial standards and guidelines for such types of publications.

5.3.1. Strengths and limitations

This study has several strengths. To the best of our knowledge, it is the first to analyse editorial materials, viewpoints, and other opinion pieces published in scholarly journals, and thus sets a benchmark to which future research on similar corpora could be compared. In terms of its methodology, the record retrieval and selection were based on a precise search strategy developed in collaboration with a medical librarian, and a screening process performed by two reviewers based on pre-determined, piloted eligibility criteria. Furthermore, the LIWC analysis was based on validated methods used in previous research (80), and their use allowed us to analyse a large textual corpus, which we could not have done through manual methods.

One limitation of this study, aside from the fact that only a preliminary analysis is presented here, lies in the lack of research, benchmarks, or standards to which the results of the

sentiment analysis could be compared. This prevents inference of whether their tone, for example, is more or less negative than it should be, as there are no thresholds or expectations for "ideal" tone. The search strategy is also partially limited, as it utilised filters for editorials and related materials in our search strategy. Such filters depend on the accuracy of the metadata in the databases, which in turn depends on the publishers or organisations that send it and the databases' own categorisation of these records. We mitigated this to some extent by searching three databases (WoS, MEDLINE *via* WoS, and Scopus), but the issue could have affected our sample to an unknown extent. Relatedly, the sample itself disproportionately includes editorial material published during the COVID-19 pandemic, which is expected, given both its global impact and the general increase in the output of academic publishing since the H1N1 pandemic. The results should, therefore, be interpreted with this in mind. Another limitation lies in the nature of the LIWC variables, which are dictionary-based and therefore do not reflect context (80). This will be mitigated through further analyses, which will contextualise this sentiment by words and topics to some extent. However, only a manual qualitative analysis (which would be extremely time-consuming with a dataset of this size) would address this issue fully and give the deepest insight into the themes emerging from these editorials and opinion pieces.

6. CONCLUSION

6.1. Practical implications of findings and way forward

The implications of the research presented in this thesis are threefold. First, there is a clear need to redesign how three key aspects of HSG development are disclosed in the final public record:

1. how conflicts of interest are managed among GDG members;
2. how the influence of external funders and parties is controlled in the development process;
3. how the HSG are methodologically developed, including how relevant research is selected and synthesised, and how the final recommendations are formulated.

A good step towards this would be the uptake of existing tools, such as the AGREE-HS used in the study presented here, or the development of new, more specific reporting guidelines akin to the ones used to report on biomedical research and CPGs. This would offer a simple way for GDGs to adhere to the highest standards of reporting and transparency, which would annul any controversies that might subsequently emerge in the public discourse. This is also an opportunity for research among experts who have developed HSGs for COVID-19, where their perspectives on gaps in HSG development processes could be collected through interviews and focus groups. This could inform both the targeted improvement of said processes and the production of new or the improvement and uptake of existing reporting guidelines for transparently reporting on the development of HSG.

The second implication relates to the use of evidence in developing recommendations within HSG. Given the complexities outlined above and the need for the rapid issuing of public health guidelines during pandemics, evidence outside of the gold standard (*i.e.* RCTs and systematic reviews with meta-analyses) could remain relevant in cases where other, higher-level research is scarce or ongoing. This was the case with the COVID-19 pandemic, where observational and non-randomised studies were frequently cited in support of recommendations in HSG. Current evidence assessment frameworks, however, do not fully reflect this, and GDGs face challenges in determining the strength and certainty of evidence for such research and subsequently formulating appropriate recommendations. Similarly, non-peer-reviewed preprints have been cited alongside peer-reviewed literature in these HSG, despite explicit disclaimers that they should not be used to inform decision-making and guideline development. While such rapid methods of communicating scientific knowledge can be extremely useful during emergencies, they should also be used with care and transparency. Yet global standards for this practice do not exist, and GDGs are left to evaluate the validity of research reported in

preprints themselves in contexts where other published, peer-reviewed studies do not exist. These gaps leave two avenues for action, both in terms of research and policy-making:

1. The development of better evidence assessment frameworks and processes that would incorporate evidence from observational and non-randomised studies, while still clearly presenting the resulting recommendations as being based on such research. Here, it would be important to disclose that the nature of this research limits our ability to draw inferences regarding causality, alongside any other caveats related to confounding and other factors.
2. The formulation of standards regarding the use of preprints for decision-making that would account specifically for the use of non-peer-reviewed literature in contexts of public health emergencies.

Lastly, more disease outbreaks will certainly occur in the future and, given the growing prominence of social media, will again be characterised by intense public discussions on the response of healthcare systems and governments worldwide. Researchers and experts communicating through scholarly outlets should strive to cut through this "noise" and offer objective criticism, while steering away from exacerbating any negative or politicised public discourse. Achieving this in the complex research ecosystem affected by external sociopolitical factors is challenging and requires coordinated efforts from both authors, editors, and publishers. The study presented here offers an initial insight into scientists' communication during pandemics based on editorials and opinion pieces published in scholarly journals, setting a baseline to which the findings of other future research could be compared. We urge that, prior to the testing of any interventions or standards in this sense, other large language analyses be performed on samples of scholarly publications sourced from both inside and outside of public health emergency contexts, and to explore how they are "picked up" and discussed by news outlets and on social media. This would help stakeholders (*i.e.* authors, editors, and publishers) identify targets for said interventions and shape them accordingly.

7. SUMMARY

A critical assessment of health systems guidance and the underlying discourse in the H1N1 and COVID-19 pandemics

Objectives: Through three studies, this thesis assessed the transparency and completeness of reporting of health systems guidance (HSG) issued by the World Health Organization (WHO), the US Centers for Disease Control and Prevention (CDC), and the European Center for Disease Prevention and Control (ECDC) for the H1N1 and COVID-19 pandemics, and explored their evidence base. Furthermore, given the intensity and politicisation of discussions on the healthcare systems' response to these pandemics, it also analysed editorials and opinion pieces published by experts and researchers in scholarly journals.

Materials and methods: The first study assessed HSG issued by the WHO, the CDC, and the ECDC through the Appraisal of Guidelines Research & Evaluation – Health Systems (AGREE-HS) tool and its five domains: "Topic", "Participants", "Methods", "Recommendations", and "Implementability". The second study classified the evidence underlying the COVID-19 HSG based on study design. The third study analysed the sentiment of editorials and opinion pieces through which researchers and editors commented on the healthcare response to the H1N1 and COVID-19 pandemics.

Results: An AGREE-HS assessment of 108 HSG published by the three organisations for the two pandemics highlighted gaps in the reporting of how they were developed, how and of whom the development groups were composed, and how CoIs and funder influence were managed. The CDC HSG scored significantly lower than those of either the ECDC or the WHO. Regarding the underlying levels of evidence, a preliminary analysis of 461 studies cited in 17 ECDC-produced COVID-19 HSG showed a predominance of observational and modelling studies supporting their recommendations, although this changed later in the pandemic as new evidence emerged. Just under a fifth of the referenced studies were preprints. The preliminary analysis of sentiment of 200 editorials showed that, compared to those related to the COVID-19 pandemic, the editorials published during the H1N1 pandemic had a more negative tone and higher certitude, and a lower proportion of words reflecting the sentiment of need.

Conclusions: These findings have three key implications. First, they highlight a need for greater transparency in reporting on how HSG are developed, as well as how conflicts of interest and funder influence are managed. Second, they call for updated evidence assessment frameworks and standards for incorporating observational studies and preprints in HSG development during public health emergencies. Finally, they underscore the importance of responsible expert communication during crises and the need for further research on how scholarly outputs influence public discourse.

8. LAY SUMMARY

A critical assessment of health systems guidance and the underlying discourse in the H1N1 and COVID-19 pandemics

Objectives: This thesis looked at the health systems guidance (HSG) developed by the World Health Organization, the US Centers for Disease Control and Prevention (CDC), and the European Center for Disease Prevention and Control for the H1N1 and COVID-19 pandemics. First, it evaluated how completely the processes of developing the HSG were presented. Then, it explored what kind of evidence was used to support the recommendations in the HSG issued during the COVID-19 pandemic. Lastly, it looked at how researchers communicated their opinions on the response of healthcare systems to these two pandemics in scientific journals.

Materials and methods: Through three studies, the thesis evaluated how fully the development of HSG at the organisations was presented through a tool made for this purpose, categorised the types of studies used as evidence for their recommendations based on their methods and design, and analysed opinion pieces written by experts in scientific journals during both pandemics.

Results: Many HSG did not completely present how they were developed and how conflicts of interest or funding influences were managed. This was especially true for the CDC's HSG when compared to the other two organisations. Most of the evidence used early in the COVID-19 pandemic came from studies based on observational designs or statistical modelling, rather than experimental designs and literature reviews. Expert commentary during the H1N1 pandemic was generally more negative and more confident in tone than during COVID-19.

Conclusions: The findings show a need for more transparent presentation of how HSG are developed during pandemics and for better methods for evaluating and using different types of evidence during such crises. There is also a need to explore and raise awareness of how experts communicate during public health crises, so that they can help keep the public dialogue more open and unbiased.

.

9. CROATIAN SUMMARY

Kritička procjena smjernica za zdravstvene sustave i pripadajućeg diskursa za vrijeme H1N1 i COVID-19 pandemija

Ciljevi: Ova disertacija procjenjuje transparentnost i cjelovitost izvještavanja o procesima razvoja smjernica za zdravstvene sustave koje su tijekom pandemija H1N1 i COVID-19 izdali Svjetska zdravstvena organizacija, Centar za kontrolu i prevenciju bolesti Sjedinjenih Američkih Država, te Europski centar za prevenciju i kontrolu bolesti, te istražuje razine dokaza na kojima su temeljene preporuke u smjernicama za COVID-19 pandemiju. Zbog intenziteta i politizacije javnih rasprava, dodatno su analizirani uvodnici i osvrti stručnjaka i znanstvenika objavljeni u znanstvenim časopisima na temu odgovora zdravstvenih sustava na pandemije.

Materijali i metode: Prvo istraživanje procjenjuje spomenute smjernice korištenjem *Appraisal of Guidelines Research & Evaluation – Health Systems* (*AGREE-HS*) alata za vrednovanje potpunosti izvještavanja i njegovih pet domena usredotočenih na tematiku smjernica, sudionike i metode njihovog razvoja, te preporuke i njihovu provedivost u praksi. Drugo istraživanje klasificira dokaze iza preporuka za COVID-19 smjernice na temelju ustroja vezanih istraživanja na osnovu procjene istraživača i prema razinama dokaza Instituta Joanna Briggs. Treće istraživanje analizira sentiment uvodnika koje su eksperti objavili u znanstvenim časopisima na temu odgovora zdravstvenih sustava na pandemije H1N1 i COVID-19.

Rezultati: Procjena 108 smjernica za dvije pandemije pomoću alata *AGREE-HS* pokazala je značajne nedostatke u izvještavanju o metodama izrade smjernica, ustroju skupina koje su ih razvijale, te načinima za sprečavanje utjecaja sukoba interesa i financijera na proces njihova razvoja. Pritom su smjernice Centara za kontrolu i prevenciju bolesti Sjedinjenih Američkih Država ocjenjene značajno niže od smjernica drugih organizacija. Preliminarna analiza 461 istraživanja citiranih u 17 COVID-19 smjernica Europskog centra za prevenciju i kontrolu bolesti pronašla je visok udio opservacijskih studija i studija temeljenih na statističkom modeliranju. Skoro petina spomenutih istraživanja su podijeljena kao preprinti – nerecenzirani znanstveni članci. Analiza sentimenta 200 uvodnika pokazala je da su oni objavljeni tijekom pandemije H1N1 imali izraženije negativan ton, viši stupanj sigurnosti u tvrdnjama i manji udio riječi koje odražavaju osjećaj potrebe, u usporedbi s onima iz razdoblja pandemije COVID-19.

Zaključci: Disertacija ukazuje na potrebu za transparentnim izvještavanjem o procesima razvoja smjernica za zdravstvene sustave, osmišljavanjem metoda procjena dokaza koje bi na prikladan način uzele u obzir opservacijske i druge nerandomizirane studije u vrijeme zdravstvenih kriza, te postavljanjem standarda za korištenje nerecenziranih preprint članaka u takvim kontekstima. Konačno, disertacija naglašava potrebu daljnjeg istraživanja znanstvene komunikacije u časopisima i njenog utjecaja na javno mnijenje.

10. CROATIAN LAY SUMMARY

Kritička procjena smjernica za zdravstvene sustave i pripadajućeg diskursa za vrijeme H1N1 i COVID-19 pandemija

Ciljevi: Ovo disertacija procjenjivala je smjernice za zdravstvene sustave koje su za vrijeme H1N1 i COVID-19 pandemija izdale Svjetska Zdravstvena Organizacija, Centri za kontrolu i prevenciju bolesti Sjedinjenih Američkih Država, te Europski centar za prevenciju i kontrolu bolesti. Cilj je bio utvrditi koliko su procesi razvoja smjernica potpuno i transparentno predstavljeni, na kakvim dokazima su preporuke u COVID-19 smjernicama temeljile, te kako su stručnjaci i znanstvenici raspravljali o odgovorima zdravstvenih sustava na dvije krize u znanstvenim časopisima.

Materijali i metode: Disertacija se temelji na tri istraživanja. Istraživači su prvo procjenjivali potpunost izvještavanja u smjernicama za dvije pandemije koristeći alat posebno razvijen za tu svrhu. U drugom istraživanju su kategorizirali istraživanja citirana u smjernicama za COVID-19 pandemiju na temelju njihova ustroja i korištenih znanstvenih metoda. Konačno, istraživači su koristili posebne računalne metode kako bi analizirali sentiment jezika članaka u kojima su znanstvenici raspravljali o odgovoru zdravstvenih sustava na obje zdravstvene krize.

Rezultati: Otkriveno je da u mnogim smjernicama nije jasno prikazano kako su nastale ni kako su organizacije sprečavale utjecaj sukoba interesa i financijera na razvoj preporuka. Preporuke Centara za kontrolu i prevenciju zaraza Sjedinjenih Američkih Država u tom pogledu su bile najslabije ocijenjene. Većina preporuka u smjernicama za COVID-19 pandemiju utemeljena je na istraživanjima koje su koristila opservacijski ustroj ili statističko modeliranje, osobito na početku pandemije. Komentari stručnjaka objavljeni u znanstvenim časopisima tijekom H1N1 pandemije bili su negativniji i samopouzdaniji nego oni objavljeni tijekom COVID-19 pandemije.

Zaključci: Disertacija pokazuje da je potrebno poboljšati transparentnost u izradi smjernica za zdravstvene sustave i razviti metode koje bi uzele u obzir različite vrste dokaza za vrijeme pandemija. Također je potrebno istraživati kako znanstvenici komuniciraju u javnosti tijekom zdravstvenih kriza kako bi mogli pomoći održavanju kvalitetnih i nepristranih javnih rasprava.

11. REFERENCES

1. World Health Organization. International Health Regulations. Geneva, Switzerland: World Health Organization; 2016 Jan 1 [cited 2024 Dec 6]. Available from: https://www.who.int/publications/i/item/9789241580496.

2. Wenham C, Stout L. A legal mapping of 48 WHO member states' inclusion of public health emergency of international concern, pandemic, and health emergency terminology within national emergency legislation in responding to health emergencies. Lancet (London, England). 2024;403:1504–12.

3. Mullen L, Potter C, Gostin LO, Cicero A, Nuzzo JB. An analysis of International Health Regulations Emergency Committees and Public Health Emergency of International Concern Designations. BMJ Global Health. 2020;5:e002502.

4. World Health Organization. Multi-country outbreak of monkeypox, External situation report #8 – 19 October 2022. Geneva (Switzerland): World Health Organization; 2022 Oct 19. Available from: https://www.who.int/publications/m/item/multi-country-outbreak-of-monkeypox--external-situation-report--8---19-october-2022.

5. World Health Organization. First meeting of the International Health Regulations (2005) Emergency Committee regarding the upsurge of mpox 2024. Geneva (Switzerland): World Health Organization; 2024 Aug 19 [cited 2024 Dec 6]. Available from: https://www.who.int/news/item/19-08-2024-first-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-upsurge-of-mpox-2024.

6. World Health Organization. WHO handbook for guideline development, 2nd Edition. Geneva (Switzerland): World Health Organization; 18 December 2014 [cited 2024 Dec 6]. Available: https://www.who.int/publications/i/item/9789241548960.

7. Garritty CM, Norris SL, Moher D. Developing WHO rapid advice guidelines in the setting of a public health emergency. J Clin Epidemiol. 2017;82:47–60.

8. Burci GL. The Legal Response to Pandemics: The Strengths and Weaknesses of the International Health Regulations. Journal of International Humanitarian Legal Studies. 2020;11:204–17.

9. Kokki M, Ammon A. Preparing Europe for future health threats and crises – key elements of the European Centre for Disease Prevention and Control's reinforced mandate. Euro Surveill. 2023;28:2300033.

10. Carande-Kulis V, Elder RW, Koffman DM. Standards Required for the Development of CDC Evidence-Based Guidelines. MMWR Suppl. 2022;71:1–6.

11. Congressional Research Service. The Centers for Disease Control and Prevention (CDC). Washington D.C. (USA): United States Congress; 2025 Mar 20 [cited 2025 May 6]. Available from: https://crsreports.congress.gov/product/pdf/IF/IF12241/5.

12. Cohen D, Carter P. WHO and the pandemic flu "conspiracies". BMJ. 2010;340:c2912.

13. Evans MR. The swine flu scam? J Public Health (Oxf). 2010;32:296–7.

14. Tanne JH. US CDC announces major changes after criticism of its responses to covid-19 and monkeypox. BMJ. 2022;378:o2074.

15. Gontariuk M, Krafft T, Rehbock C, Townend D, Van der Auwermeulen L, Pilot E. The European Union and Public Health Emergencies: Expert Opinions on the Management of the First Wave of the COVID-19 Pandemic and Suggestions for Future Emergencies. Front Public Health. 2021;9:698995.

16. Wilder-Smith A, Osman S. Public health emergencies of international concern: a historic overview. J Travel Med. 2020;27:taa227.

17. Villareal P. Pandemic: building a legal concept for the future. Washington University Global Studies Law Review. 2021;20:611.

18. Morens DM, Folkers GK, Fauci AS. What Is a Pandemic? J Infect Dis. 2009;200:1018–21.

19. Zarocostas J. World Health Organization declares A (H1N1) influenza pandemic. BMJ. 2009;338:b2425.

20. World Health Organization. DG Statement following the meeting of the Emergency Committee. Geneva (Switzerland): World Health Organization; 2009 Jun 11 [cited 2024 Dec 9]. Available from: https://www.who.int/news/item/11-06-2009-dg-statement-following-the-meeting-of-the-emergency-committee.

21. World Health Organization. First meeting of the IHR Emergency Committee. Geneva (Switzerland): World Health Organization; 2009 Apr 25 [cited 2024 Dec 9]. Available from: https://www.who.int/news/item/25-04-2009-first-meeting-of-the-ihr-emergency-committee.

22. Hayward A. Influenza A (H1N1) pandemic: true or false alarm. J Epidemiol Community Health. 2009;63:775–6.

23. Allam MF. Influenza A (H1N1) pandemic: true or false alarm. J Epidemiol Community Health. 2009;63:862.

24. Irving L, Hampson A. Influenza A (H1N1 09): public health lessons and questions. Aust Fam Physician. 2009;38:567.

25. Doshi P. Calibrated response to emerging infections. BMJ. 2009;339:b3471.

26. Kamradt-Scott A. What Went Wrong? The World Health Organization from Swine Flu to Ebola. Political Mistakes and Policy Failures in International Relations. 2017 Oct 9:193–215.

27. Godlee F. Conflicts of interest and pandemic flu. BMJ. 2010;340:c2947.

28. Zarocostas J. Swine flu pandemic review panel seeks access to confidential documents between WHO and drug companies. BMJ. 2010;340:c2792.

29. Butler D. Flu experts rebut conflict claims. Nature. 2010;465:672–3.

30. Kon P. WHO and pandemic flu. GlaxoSmithKline UK responds. BMJ. 2010;340:c3464.

31. Doshi P, Jefferson T. WHO and pandemic flu. Another question for GSK. BMJ. 2010;340:c3455.

32. Jefferson T, Doshi P. Time for change, WHO. BMJ. 2010;340:c3461.

33. Zarocostas J. WHO processes on dealing with a pandemic need to be overhauled and made more transparent. BMJ. 2011;342:d3378.

34. Penn CR. Future perfect? Improving preparedness through the experiences of the influenza A (H1N1) 2009 pandemic. Bull World Health Organ. 2011;89:470.

35. Council of Europe Parliamentary Assembly. The handling of the H1N1 pandemic: more transparency needed. Brussels (Luxembourg): Council of Europe; 2010 Mar 23 [cited 2024 Dec 12]. Available from: https://assembly.coe.int/CommitteeDocs/2010/20100329_MemorandumPandemie_E.pdf.

36. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Geneva, Switzerland: World Health Organization; 2020 Jan 30 [cited 2024 Dec 13]. Available from: https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

37. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19. Geneva, Switzerland: World Health Organization; 2020 Mar 11 [cited 2024 Dec 12]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

38. Gostin LO. COVID-19 Reveals Urgent Need to Strengthen the World Health Organization. JAMA. 2020;323:2361–2.

39. Brown TM, Ladwig S. COVID-19, China, the World Health Organization, and the Limits of International Health Diplomacy. Am J Public Health. 2020;110:1149–51.

40. Dagens A, Sigfrid L, Cai E, Lipworth S, Cheng V, Harris E, et al. Scope, quality, and inclusivity of clinical guidelines produced early in the covid-19 pandemic: rapid review. BMJ. 2020;369:m1936.

41. Chan AL, Leung CC, Cheng K. To wear or not to wear: WHO's confusing guidance on masks in the covid-19 pandemic. BMJ [Internet]. 2020 Mar 11 [cited 2024 Dec 12]. Available from: https://blogs.bmj.com/bmj/2020/03/11/whos-confusing-guidance-masks-covid-19-epidemic/.

42. Guharoy R, Krenzelok EP. Lessons from the mismanagement of the COVID-19 pandemic: A blueprint to reform CDC. Am J Health Syst Pharm. 2021;78:1739–41.

43. Benzian H, Johnston M, Stauf N, Niederman R. Presenting or Spinning Facts? Deconstructing the U.S. Centers for Disease Control Statement on the Importance of Reopening Schools Under COVID-19. Front Public Health. 2021;9:645229.

44. Krohnert K, Haslam A, Høeg TB, Prasad V. Statistical and Numerical Errors Made by the US Centers for Disease Control and Prevention During the COVID-19 Pandemic. SSRN [preprint]. 2023 Mar 23 [cited 2024 Dec 12]. Available from: https://dx.doi.org/10.2139/ssrn.4381627.

45. Tomori C, Ahmed A, Evans DP, Meier BM, Nair A. Your health is in your hands? US CDC COVID-19 mask guidance reveals the moral foundations of public health. EClinicalMedicine. 2021;38:101071.

46. Centres for Disease Control and Prevention. CDC Moving Forward Summary Report [Internet]. Atlanta (USA): Centres for Disease Control and Prevention; 2024 Feb 7 [cited 2024 Oct 25]. Available from: https://www.cdc.gov/about/cdc-moving-forward/index.html.

47. Gontariuk M, Krafft T, Rehbock C, Townend D, Van der Auwermeulen L, Pilot E. The European Union and Public Health Emergencies: Expert Opinions on the Management of the First Wave of the COVID-19 Pandemic and Suggestions for Future Emergencies. Front Public Health. 2021;9:698995.

48. National Center for Complementary and Integrative Health. Clinical Practice Guidelines 2021. Bethesda, Maryland (USA): National Center for Complementary and Integrative Health; 2021 [cited 2024 Dec 20]. Available from: https://www.nccih.nih.gov/health/providers/clinicalpractice.

49. Brouwers MC, Lavis JN, Spithoff K, Vukmirovic M, Florez ID, Velez M, et al. Assessment of health systems guidance using the Appraisal of Guidelines for Research and Evaluation – Health Systems (AGREE-HS) instrument. Health Policy. 2019;123:646–51.

50. Movsisyan A, Melendez-Torres GJ, Montgomery P. Users identified challenges in applying GRADE to complex interventions and suggested an extension to GRADE. J Clin Epidemiol. 2016;70:191–9.

51. Hilton Boon M, Thomson H, Shaw B, Akl EA, Lhachimi SK, López-Alcalde J, et al. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. J Clin Epidemiol. 2021;135:42–53.

52. Schünemann HJ, Santesso N, Vist GE, Cuello C, Lotfi T, Flottorp S, et al. Using GRADE in situations of emergencies and urgencies: certainty in evidence and recommendations matters during the COVID-19 pandemic, now more than ever and no matter what. J Clin Epidemiol. 2020;127:202–7.

53. Florez ID, Morgan RL, Falavigna M, Kowalski SC, Zhang Y, Etxeandia-Ikobaltzeta I, et al. Development of rapid guidelines: 2. A qualitative study with WHO guideline developers. Health Res Policy Syst. 2018;16:62.

54. Tricco AC, Garritty CM, Boulos L, Lockwood C, Wilson M, McGowan J, et al. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. J Clin Epidemiol. 2020;126:177–83.

55. Saluja K, Reddy KS, Wang Q, Zhu Y, Li Y, Chu X, et al. Improving WHO's understanding of WHO guideline uptake and use in Member States: a scoping review. Health Res Policy Syst. 2022;20:98.

56. Breneol S, Curran JA, Marten R, Minocha K, Johnson C, Wong H, et al. Strategies to adapt and implement health system guidelines and recommendations: a scoping review. Health Res Policy Syst. 2022;20:64.

57. Erismann S, Pesantes MA, Beran D, Leuenberger A, Farnham A, Berger Gonzalez de White M, et al. How to bring research evidence into policy? Synthesizing strategies of five research projects in low-and middle-income countries. Health Res Policy Syst. 2021;19:29.

58. Norris SL, Sawin VI, Ferri M, Reques Sastre L, Porgo TV. An evaluation of emergency guidelines issued by the World Health Organization in response to four infectious disease outbreaks. PloS One. 2018;13:e0198125.

59. Te Brake H, Willems A, Steen C, Dückers M. Appraising Evidence-Based Mental Health and Psychosocial Support (MHPSS) Guidelines-PART I: A Systematic Review on Methodological Quality Using AGREE-HS. Int J Environ Res Public Health. 2022;19:3107.

60. Marshall-Andon T, Walsh S, Berger-Gillam T, Pari AAA. Systematic review of post-COVID-19 syndrome rehabilitation guidelines. Integr Healthc J. 2022;4:e000100.

61. Suhail MK, Hannis D, Rhodes A. An Evaluation of Kerala State's Emergency Response Plan Document Against Nipah Virus Using the AGREE-HS Tool. Outbreak Management and Response. 2025;1:2501391.

62. Zhang G, Li A, Bai W, Liu C, Chen X, Liu F, et al. Assessing the quality for integrated guidelines: systematic comparison between the AGREE II and AGREE-HS tools. ResearchSquare [preprint]. 2025 Jun 3 [cited 2025 May 18]. Available from: https://www.researchsquare.com/article/rs-6615827/v1.

63. Kirsh S, Ling M, Jassal T, Pitre T, Piggott T, Zeraatkar D. Values and preferences in COVID-19 public health guidelines: a systematic review. J Clin Epidemiol. 2024;174:111473.

64. AGREE-HS Research Team. Appraisal of Guidelines Research & Evaluation – Health Systems (AGREE-HS) Manual. Ontario, Canada: AGREE Enterprise; 2018 [cited 2024 Oct 25]. Available from: https://www.agreetrust.org/wp-content/uploads/2018/03/AGREE-HS-Manual-March-2018.pdf.

65. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10:39.

66. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22:276–82.

67. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.

68. Brouwers MC, Spithoff K, Lavis J, Kho ME, Makarski J, Florez ID. What to do with all the AGREEs? The AGREE portfolio of tools to support the guideline enterprise. J Clin Epidemiol. 2020;125:191–7.

69. Ako-Arrey DE, Brouwers MC, Lavis JN, Giacomini MK. Health system guidance appraisal--concept evaluation and usability testing. Implement Sci. 2016;11:3.

70. Ako-Arrey DE, Brouwers MC, Lavis JN, Giacomini MK. Health systems guidance appraisal--a critical interpretive synthesis. Implement Sci. 2016;11:9.

71. Brouwers MC, Ako-Arrey D, Spithoff K, Vukmirovic M, Florez ID, Lavis JN, et al. Validity and usability testing of a health systems guidance appraisal tool, the AGREE-HS. Health Res Policy Syst. 2018;16:51.

72. Ursić L, Žuljević MF, Vuković M, Bralić N, Roje R, Matas J, et al. Assessing the quality and completeness of reporting in health systems guidance for pandemics using the AGREE-HS tool. J Glob Health. 2023;13:06050.

73. Lotfi T, Hajizadeh A, Moja L, Akl EA, Piggott T, Kredo T, et al. A taxonomy and framework for identifying and developing actionable statements in guidelines suggests avoiding informal recommendations.J Clin Epidemiol. 2022;141:161–71.

74. Joanna Briggs Institute. Joanna Briggs Institute Levels of Evidence. Joanna Briggs Institute; 2014 [cited 2024 Oct 25]. Available from: https://jbi.global/sites/default/files/2019-05/JBI-Levels-of-evidence_2014_0.pdf.

75. Supporting Document for the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Joanna Briggs Institute; 2015 [cited 2024 Dec 12]. Available from: https://jbi.global/sites/default/files/2019-05/JBI%20Levels%20of%20Evidence%20Supporting%20Documents-v2.pdf.

76. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.

77. Ursić L, Bralić N, Spitale G, Germani F, Marušić A. Scientific discourse during health crises: A linguistic and content analysis of editorials on the response to the H1N1 and COVID-19 pandemics. Open Science Framework [preregistration]. 2024 Jul 3 [cited 2024 Oct 25]. Available from: https://doi.org/10.17605/OSF.IO/3F2NW.

78. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology. 2010;29:24–54.

79. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The Development and Psychometric Properties of LIWC-22. Austin, TX, USA: University of Texas at Austin; 2022 [cited 2024 Oct 25]. Available from: https://www.liwc.app/static/documents/LIWC-22%20Manual%20-%20Development%20and%20Psychometrics.pdf.

80. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC 2015. Austin, TX, USA: University of Texas at Austin; 2015 [cited 2024 Oct 25]. Available from:

https://www.liwc.app/static/documents/LIWC2015%20Manual%20-%20Development%20and%20Psychometrics.pdf.

81. Bahor Z, Liao J, Currie G, Ayder C, Macleod M, McCann SK, et al. Development and uptake of an online systematic review platform: the early years of the CAMARADES Systematic Review Facility (SyRF). BMJ Open Sci. 2021;5:e100103.

82. Alonso-Coello P, Irfan A, Solà I, Gich I, Delgado-Noguera M, Rigau D, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Qual Saf Health Care. 2010;19:e58.

83. Armstrong JJ, Goldfarb AM, Instrum RS, MacDermid JC. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. J Clin Epidemiol. 2017;81:13–21.

84. Olsen WM, Freeman C, Adewumi A, La Caze A. A scoping review of health system guidelines for pharmacist responsibilities when dispensing opioids. Explor Res Clin Soc Pharm. 2023;12:100382.

85. Kastner M, Bhattacharyya O, Hayden L, Makarski J, Estey E, Durocher L, et al. Guideline uptake is influenced by six implementability domains for creating and communicating guidelines: a realist review. J Clin Epidemiol. 2015;68:498–509.

86. Orangi S, Orangi T, Kabubei KM, Honda A. Understanding factors influencing the use of clinical guidelines in low-income and middle-income settings: a scoping review. BMJ Open. 2023;13(6):e070399.

87. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med. 2010;8:24.

88. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust. 2006;185:263–7.

89. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev. 2012;1:60.

90. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev. 2012;11:MR000030.

91. Vinkers CH, Lamberink HJ, Tijdink JK, Heus P, Bouter L, Glasziou P, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. PLoS Biol. 2021;19:e3001162.

92. Yao X, Xia J, Jin Y, Shen Q, Wang Q, Zhu Y, et al. Methodological approaches for developing, reporting, and assessing evidence-based clinical practice guidelines: a systematic survey. J Clin Epidemiol. 2022;146:77–85.

93. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol. 2019;19:132.

94. Norris SL. Meeting public health needs in emergencies-World Health Organization guidelines. J Evid Based Med. 2018;11:133–5.

95. Rehfuess EA, Akl EA. Current experience with applying the GRADE approach to public health interventions: an empirical study. BMC Public Health. 2013;13:9.

96. Ingold H, Gomez GB, Stuckler D, Vassall A, Gafos M. "Going into the black box": a policy analysis of how the World Health Organization uses evidence to inform guideline recommendations. Front Public Health. 2024;12:1292475.

97. Frieden TR. Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. N Engl J Med. 2017;377:465–75.

98. Cash-Goldwasser S, Reingold AL, Luby SP, Jackson LA, Frieden TR. Masks During Pandemics Caused by Respiratory Pathogens-Evidence and Implications for Action. JAMA Netw Open. 2023;6:e2339443.

99. Loeb M, Mertz D, Chagla Z, Conly J. Why Randomized Controlled Trials are Necessary to Address the Effectiveness of Non-Pharmacological Interventions During a Public Health Emergency Such as COVID-19. J Assoc Med Microbiol Infect Dis Can. 2024;9:197–9.

100. Duval D, Evans B, Sanders A, Hill J, Simbo A, Kavoi T, et al. Non-pharmaceutical interventions to reduce COVID-19 transmission in the UK: a rapid mapping review and interactive evidence gap map. J Public Health (Oxf). 2024;46:e279–e93.

101. Madhusudanan A, Iddon C, Cevik M, Naismith JH, Fitzgerald S. Non-pharmaceutical interventions for COVID-19: a systematic review on environmental control measures. Philos Trans A Math Phys Eng Sci. 2023;381:20230130.

102. Ravinetto R, Caillet C, Zaman MH, Singh JA, Guerin PJ, Ahmad A, et al. Preprints in times of COVID19: the time is ripe for agreeing on terminology and good practices. BMC Med Ethics. 2021;22:106.

103. Alperin JP, Shores K, Fleerackers A, Chtena N. Stark Decline in Journalists' Use of Preprints Postpandemic. Science Communication. 2024:10755470241285405.

104. Fraser N, Brierley L, Dey G, Polka JK, Pálfy M, Nanni F, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLOS Biol. 2021;19:e3000959.

105. Burki T. WHO ends the COVID-19 public health emergency. Lancet Respir Med. 2023;11:588.

106. Guyatt GH, Alonso-Coello P, Schünemann HJ, Djulbegovic B, Nothacker M, Lange S, et al. Guideline panels should seldom make good practice statements: guidance from the GRADE Working Group. J Clin Epidemiol. 2016;80:3–7.

107. Lavazza A, Farina M. The Role of Experts in the Covid-19 Pandemic and the Limits of Their Epistemic Authority in Democracy. Front Public Health. 2020;8:356.

108. Why Nature supports Joe Biden for US president. Nature. 2020;586:335.

109. Tollefson J. Scientists relieved as Joe Biden wins tight US presidential election. Nature. 2020;587:183–4.

110. US president-elect Joe Biden must quickly restore science to government. Nature. 2020;587:175–6.

111. Scally G, Jacobson B, Abbasi K. The UK's public health response to covid-19. BMJ. 2020;369:m1932.

112. Kepp KP, Cristea I, Muka T, Ioannidis JPA. COVID-19 advocacy bias in the BMJ: meta-research evaluation. BMJ Open Qual. 2025;14:e003131.

113. Mendez SR, Munoz-Najar S, Emmons KM, Viswanath K. US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study. J Med Internet Res. 2025;27:e59786.

114. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR public health and surveillance. 2020;6:e19447.

115. Chandrasekaran R, Desai R, Shah H, Kumar V, Moustakas E. Examining Public Sentiments and Attitudes Toward COVID-19 Vaccination: Infoveillance Study Using Twitter Posts. JMIR Infodemiology. 2022;2:e33909.

116. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. J Med Internet Res. 2020;22:e22624.

117. Spitale G, Biller-Andorno N, Germani F. Concerns Around Opposition to the Green Pass in Italy: Social Listening Analysis by Using a Mixed Methods Approach. J Med Internet Res. 2022;24:e34385.

118. Tang L, Liu W, Thomas B, Tran HTN, Zou W, Zhang X, et al. Texas Public Agencies' Tweets and Public Engagement During the COVID-19 Pandemic: Natural Language Processing Approach. JMIR Public Health Surveill. 2021;7:e26720.

119. Alhassan FM, AlDossary SA. The Saudi Ministry of Health's Twitter Communication Strategies and Public Engagement During the COVID-19 Pandemic: Content Analysis Study. JMIR Public Health Surveill. 2021;7:e27942.

120. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining Tweet Content and Engagement of Canadian Public Health Agencies and Decision Makers During COVID-19: Mixed Methods Analysis. J Med Internet Res. 2021;23:e24883.

121. Drescher LS, Roosen J, Aue K, Dressel K, Schär W, Götz A. The Spread of COVID-19 Crisis Communication by German Public Authorities and Experts on Twitter: Quantitative Content Analysis. JMIR Public Health Surveill. 2021;7:e31834.

122. Sule S, DaCosta MC, DeCou E, Gilson C, Wallace K, Goff SL. Communication of COVID-19 Misinformation on Social Media by Physicians in the US. JAMA Netw Open. 2023;6(8):e2328928.

123. Knudsen B, Høeg TB, Prasad V. Analysis of tweets discussing the risk of Mpox among children and young people in school (May-October 2022): a retrospective observational study. BMJ Paediatr Open. 2024;8:e002236.

124. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Committee on Responsible Science. Fostering Integrity in Research. Washington (DC): National Academies Press (US); 2017 Apr 11 [cited 2025 May 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK475953/.

125. 125. Algan Y, Cohen D, Davoine E, Foucault M, Stantcheva S. Trust in scientists in times of pandemic: Panel evidence from 12 countries. Proc Natl Acad Sci U S A. 2021;118:e2108576118.

126. McLaughlin DM, Mewhirter J, Sanders R. The belief that politics drive scientific research & its impact on COVID-19 risk assessment. PloS One. 2021;16:e0249937.

127. Kerr J, Panagopoulos C, van der Linden S. Political polarization on COVID-19 pandemic response in the United States. Pers Individ Dif. 2021;179:110892.

128. Bayram AB, Shields T. Who Trusts the WHO? Heuristics and Americans' Trust in the World Health Organization During the COVID-19 Pandemic. Soc Sci Q. 2021;102:2312–30.

129. Cologna V, Mede NG, Berger S, Besley J, Brick C, Joubert M, et al. Trust in scientists and their role in society across 68 countries. Nat Hum Behav. 2025;9:713–30.

130. Ioannidis JPA, Collins TA, Bendavid E, Baas J. Massive covidization and subsequent decovidization of the scientific literature involved 2 million authors. J Clin Epidemiol. 2025;180:111705.

12. CURRICULUM VITAE

Primary information

Name and surname: Luka Ursić
Address: Dubrovačka ulica 51, 21000 Split, Croatia
Contact e-mail: luka.ursic@mefst.hr
ORCID: https://orcid.org/0000-0003-2469-2948

Employment (by date, descending)

1. Research fellow (Croatian Science Foundation. Grant Number: IP-2019-04-4882). Department of Research in Biomedicine and Health, Centre for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia (21 July 2022–)
2. Owner. Empirica (self-employed/owner) (12 November 2022–)

Education (by date, descending)

1. PhD in biomedicine and health [ongoing]. Translational Research in Biomedicine (TRIBE) programme, University of Split School of Medicine, Split, Croatia (September 2022–): https://mefst.unist.hr/studies/graduate-school/tribe/138.
2. MA in History/English Language and Literature, Faculty of Humanities and Social Sciences Split, Croatia (25 September 2019 – 21 September 2021)
3. BA in History/English Language and Literature, Faculty of Humanities and Social Sciences Split, Croatia (25 September 2016 – 24 September 2019)

Positions and editorial experience (by date, descending)

1. Co-chair. Croatian Reproducibility Network (2023–): https://crorin.hr/
2. Member of European Association of Science Editors (2022–): https://ease.org.uk/member-profile/luka-ursic-5903/.
3. Associate Editor. ST-OPEN (2022–): https://st-open.unist.hr/index.php/st-open
4. Editorial Board Member. Carnival, Journal of the International Students of History Association (2019–20): https://ishainternational.wordpress.com/carnival/
5. Editor-in-Chief. Pleter, Journal of the Asociation of History Students "Toma Arhiđakon" (2016–20): https://hrcak.srce.hr/ojs/index.php/ishasplitpleter

List of recent publications (by date, descending)

1. Krstulović J, **Ursić L**, Hrgović Z, Šuljić N, Roje R, Znaor L, Marusic A. Barriers and facilitators for implementing WHO's Surgical Safety Checklist in a publicly funded hospital: a qualitative study from a tertiary-level public hospital in Croatia. BMJ Open.

2025 Jun 30;15(6):e095155. doi: 10.1136/bmjopen-2024-095155. PMID: 40588379; PMCID: PMC12211830.

2. Hrgović Z, **Ursić L**, Krstulović J, Viđak M, Znaor L, Marušić A. Perception of the ethical climate among hospital employees in a public healthcare system: a cross-sectional survey at the University Hospital of Split, Croatia. BMC Med Ethics. 2025 May 7;26(1):59. doi: 10.1186/s12910-025-01217-1. PMID: 40335974; PMCID: PMC12060318.
3. Kaliterna M, Žuljević MF, **Ursić L**, Krka J, Duplančić D. Testing the capacity of Bard and ChatGPT for writing essays on ethical dilemmas: A cross-sectional study. Sci Rep. 2024 Oct 30;14(1):26046. doi: 10.1038/s41598-024-77576-3. PMID: 39472544; PMCID: PMC11522523.
4. **Ursić L**, Bralić N, Žuljević MF, Puljak L, Buljan I. Exploring the understanding of reproducibility among stakeholders within academia and their expectations for a web-based education tool: A qualitative study. Account Res. 2024 May 5:1-30. doi: 10.1080/08989621.2024.2345723. Epub ahead of print. PMID: 38704659.
5. **Ursić L**, Žuljević MF, Vuković M, Bralić N, Roje R, Matas J, Mijatović A, Sapunar D, Marušić A. Assessing the quality and completeness of reporting in health systems guidance for pandemics using the AGREE-HS tool. J Glob Health. 2023 Oct 27;13:06050. doi: 10.7189/jogh.13.06050. PMID: 37883198; PMCID: PMC10602204.
6. Matas J, Tokalić R, **Ursić L**, Buljan I, Utrobicic A, Marusic A. Evidence base for recommendations for writing evidence-based syntheses. Cochrane Database of Systematic Reviews. 2023;10:MR000067. doi: 10.1002/14651858.MR000067.
7. **Ursić L**, Gudelj D, Tomić V, Marušić M, Marušić A. Analysing overlay journals: The state-of-the-art in 2021 and possible perspectives. Learned Publishing. 2022;35:640-649. doi: 10.1002/leap.1491.
8. **Ursić L**, Baldacchino G, Bašić Ž, Sainz AB, Buljan I, Hampel M, Kružić I, Majić M, Marušić A, Thetiot F, et al. Factors Influencing Interdisciplinary Research and Industry-Academia Collaborations at Six European Universities: A Qualitative Study. Sustainability. 2022;14(15):9306. doi: 10.3390/su14159306
9. Gudelj D, **Ursić L**, Tomić V, Marušić M. The first year of the ST-OPEN overlay+ journal. ST-OPEN. 2021;2:1-9. Available: https://st-open.unist.hr/index.php/st-open/article/view/63.
10. Bašić Ž, Kružić I, Jelenić M, **Ursić L**, Janković S, Mihanović F, Štambuk S, Kero D, Vilović K, Primorac D, Anđelinović Š. Anthropological individualization of relics from

sarcophagus stored in Vodnjan monastery, Vodnjan, Croatia. ST-OPEN. 2022;3:e2022.2219.3. doi: 10.48188/so.3.5

Conferences (by date, descending)

1. **Ursić L**, Vitlov N, Marušić SLJ, Marušić A. Athens, Greece. Correcting the published record: How authors react to editorial prompts about corrections to cited literature. 8th World Conference on Research Integrity 2024; 2-5 Jun 2024; Athens, Greece. Amsterdam, Netherlands: World Conferences on Research Integrity Foundation; 2024.
2. Kaliterna M, Žuljević MF, **Ursić L**, Krka J, Duplančić D. Testing the capacity of Bard and ChatGPT for writing essays on ethical dilemmas: A cross-sectional study. 8th World Conference on Research Integrity 2024; 2-5 Jun 2024; Athens, Greece. Amsterdam, Netherlands: World Conferences on Research Integrity Foundation; 2024.
3. Sapunar D, Puljak L, Bašić Ž, Gudelj D, Kero D, Tomić V, **Ursić L**, Gambiroža L, Orlandić IN, Shmatkova M, Marušić M, Marušić A. The role of scientific journals in times of war. PUBMET 2023; 13-15 Sep 2023; Zadar, Croatia. Zadar: University of Zadar; 2023. doi: 10.15291/pubmet.4269.
4. Mijatović A, **Ursić L**, Žuljević MF, Bralić N, Vuković M, Ercegović V, Roguljić M, Marušić A. How good are medical students in detecting duplications in digital images from research articles: A cross-sectional survey. ENRIO 2023; 7-8 Sep 2023; Paris, France.
5. Kaliterna M, Žuljević MF, **Ursić L**, Krka J, Duplančić D, Marušić A. Testing ChatGPT's capacity to write essays on ethical dilemmas: A cross-sectional study. ENRIO 2023; 7-8 Sep 2023; Paris, France.
6. Mijatović A, **Ursić L,** Buljan I, Marušić A. A Pretrained Language Model for Classification of Cochrane Plain Languages Summaries on Conclusiveness of Recommendation. Cochrane Colloquium 2023; 2023 Sep 4-6; London, UK. London: Cochrane UK; 2023. doi: 10.1002/14651858.CD202301.
7. **Ursić L**, Žuljević MF, Vuković M, Bralić N, Roje R, Matas J, Mijatović A, Duplančić D, Sapunar D, Marušić A. Assessing the quality and completeness of health system guidelines and recommendations for pandemics using the AGREE-HS tool. Cochrane Colloquium 2023; 2023 Sep 4-6; London, UK. London: Cochrane UK; 2023. doi: 10.1002/14651858.CD202301.
8. Žuljević MF, **Ursić L**, Bralić N, Puljak L, Buljan I. Exploring the concept of reproducibility and developing a needs-based tool for educating end-users about

research reproducibility: A qualitative study. Cochrane Colloquium 2023; 2023 Sep 4-6; London, UK. London: Cochrane UK; 2023. doi: 10.1002/14651858.CD202301.

9. **Ursić L**, Roje R, Mijatović A, Matas J, Buljan I, Duplačić D, Sapunar D, Marušić A. Evidence behind policies, guidelines, and recommendations for the 2009 H1N1 and the COVID-19 pandemics: A cross-sectional study. 14th Croatian Cochrane Symposium; 2022 Sep 1; Split, Croatia. Zagreb: Cochrane Croatia; 2022. Available: https://croatia.cochrane.org/sites/croatia.cochrane.org/files/uploads/2022%20CroCoS%201 4%20-%20Book%20of%20Abstracts.pdf.
10. **Ursić L**, Gudelj D, Tomić V. Promoting student research at ST-OPEN: the analysis of a university overlay journal's first two years of publishing. PUBMET 2022; 14-16 Sep 2022; Zadar, Croatia. Zadar: University of Zadar; 2022. doi: 10.152917pubmet.3957.

Training and schools (by date, descending)

1. CHANGER Pilot Workshop on new methods of ethical assessments for research projects, 23–24 June 2025. Split, Croatia.
2. Stehr-Boldt Fellow. University of Zürich, Institute of Biomedical Ethics and History of Medicine, Zürich, Switzerland (2024/09/20–2024/12/18)
3. Certified VIRT^{2}UE trainer, 08/12/2023. Amsterdam UMC, Amsterdam, The Netherlands.
4. Blended Intensive Program: Artificial Intelligence (AI) for Humanities: from Text Simplification to Automatic Humor Analysis, 20–24 March 2023. University of Brest, Brest, France.
5. Training activity – Capacity building on teaching-learning strategies to support students' transversal skills development, 6–10 March 2023. University of Split School of Medicine, Split, Croatia.
6. Training in teaching competence for students, 2 March 2023. Faculty of Humanities and Social Sciences Split, Split, Croatia.
7. Summer School of Responsible Research, 29 August – 2 September 2022. University of Split School of Medicine, Split, Croatia.
8. 4th Edition Summer School, Valencia 2022: Challenges in Data Science: Big Data, Biostatistics,
9. Artificial Intelligence and Communications, 27 June – 1 July 2022. University of Valencia, Valencia, Spain.

10. Autumn School of Qualitative Research, 22–26 November 2021. University of Split School of Medicine, Split Croatia.
11. EASE Editors School, 30 April 2022. European Association of Science Editors.
12. VIRT2UE Training Programme – Virtue based ethics and Integrity in Research: Train the Trainer program for Upholding the principles and practice of the European Code of Conduct for Research Integrity, 24 February 2021. University of Split School of Medicine, Split, Croatia

Awards/prizes (by date, descending)

1. Dean's Reward for Academic Excellence in 2022, Faculty of Humanities and Social Sciences Split, University of Split, Split, Croatia
2. Rector's Reward for Academic Excellence in 2022, University of Split, Split, Croatia
3. Rector's Reward for Excellence for Outstanding Contribution in 2022, University of Split, Split, Croatia
4. Stipend of the University of Split for Academic Excellence in 2022, University of Split, Split, Croatia
5. Stipend of the City of Split for Academic Excellence (2019-2022), Split, Croatia
6. Award of non-governmental organisation "Naš kvart", Split, Croatia, for contribution to the activities of youth within the city of Split in year 2019.

Invited lectures and workshops (by date, descending)

1. ECDC internal webinar: Assessing the transparency and completeness of reporting of health systems guidance for the H1N1 and COVID-19 pandemics - Findings and suggestions for future crises. 23 April 2024.
2. Preparing impactful charts, graphs and other illustrations; for the European Association of Science Editors (EASE) – 08/03/2023 and 11/09/2024
3. Writing a research manuscript; for the MedLaw 2024 conference at the University of Split School of Medicine, 19–21/04/2024.

13. SUPPLEMENT

Table S1. Full search strategy for study on the completeness of reporting of HSG

Keywords/search string	Modifiers	Date of search	Time span*	Results
CDC				
CDC Stacks (https://stacks.cdc.gov/)				
pandemic AND guideline	Document Full Text	29 November 2021		657
pandemic AND guidelines	Document Full Text	29 November 2021		2922
pandemics AND guideline	Document Full Text	29 November 2021		152
pandemics AND guidelines	Document Full Text	29 November 2021		539
pandemic AND recommendation	Document Full Text	29 November 2021		1508
pandemic AND recommendations	Document Full Text	30 November 2021		4089
pandemics AND recommendation	Document Full Text	30 November 2021		288
pandemics AND recommendations	Document Full Text	30 November 2021		652
pandemic AND policy	Document Full Text	30 November 2021		3023
pandemic AND policies	Document Full Text	30 November 2021		3149
pandemics AND policy	Document Full Text	30 November 2021		584
pandemics AND policies	Document Full Text	30 November 2021		458
H1N1 AND guideline	Document Full Text	1 December 2021		380
H1N1 AND guidelines	Document Full Text	1 December 2021		1530
H1N1 AND recommendation	Document Full Text	1 December 2021		971
H1N1 AND recommendations	Document Full Text	1 December 2021		2335
H1N1 AND policy	Document Full Text	1 December 2021		1452
H1N1 AND policies	Document Full Text	1 December 2021		1007
COVID-19 AND guideline	Document Full Text	1 December 2021		206
COVID-19 AND guidelines	Document Full Text	1 December 2021		1426
COVID-19 AND recommendation	Document Full Text	1 December 2021		669
COVID-19 AND recommendations	Document Full Text	1 December 2021		2721
COVID-19 AND policy	Document Full Text	1 December 2021		1486
COVID-19 AND policies	Document Full Text	1 December 2021		5795
Total				37999
Additional search to 17 March 2022				
pandemic AND guideline	Document Full Text	25 October 2022	29 November 2021 to 17 March 2022	6

pandemic AND guidelines	Document Full Text	25 October 2022	29 November 2021 to 17 March 2022	25
pandemics AND guideline	Document Full Text	25 October 2022	29 November 2021 to 17 March 2022	1
pandemics AND guidelines	Document Full Text	25 October 2022	29 November 2021 to 17 March 2022	9
pandemic AND recommendation	Document Full Text	25 October 2022	29 November 2021 to 17 March 2022	19
pandemic AND recommendations	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	51
pandemics AND recommendation	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	12
pandemics AND recommendations	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	50
pandemic AND policy	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	40
pandemic AND policies	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	33
pandemics AND policy	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	22
pandemics AND policies	Document Full Text	25 October 2022	30 November 2021 to 17 March 2022	16
H1N1 AND guideline	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	2
H1N1 AND guidelines	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	4
H1N1 AND recommendation	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	4
H1N1 AND recommendations	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	6
H1N1 AND policy	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	8

H1N1 AND policies	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	4
COVID-19 AND guideline	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	8
COVID-19 AND guidelines	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	39
COVID-19 AND recommendation	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	38
COVID-19 AND recommendations	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	134
COVID-19 AND policy	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	63
COVID-19 AND policies	Document Full Text	25 October 2022	1 December 2021 to 17 March 2022	32
Total				626
H1N1 page (https://www.cdc.gov/h1n1flu/)	N/A			413
MMWR (*via* PubMed)	N/A			150
TOTAL				39118
ECDC				
ECDC repository (https://www.ecdc.europa.eu/en/search?s=)				
COVID-19	Data, News, Publications	8 April 2022		911
COVID-19 recommendations	Data, News, Publications	8 April 2022		2150
COVID-19 guidelines	Data, News, Publications	8 April 2022		1507
H1N1	Data, News, Publications	8 April 2022		660

H1N1 recommendations	Data, News, Publications	8 April 2022	2262
H1N1 guidelines	Data, News, Publications	8 April 2022	1328
pandemic guidelines	Data, News, Publications	8 April 2022	1354
pandemic recommendations	Data, News, Publications	8 April 2022	2284
Topic filter search (Severe acute respiratory syndrome (SARS), SARS-CoV-2, SARS-CoV-2 variants, COVID-19, Coronavirus, Influenza A(H1N1), Influenza A(H1N1) 2009, Swine origin influenza, Influenza in humans, Swine origin, Influenza in humans, pandemic)	Data, News, Publications	15 April 2022	5909
H1N1 webpage (https://www.ecdc.europa.eu/en/seasonal-influenza/2009-influenza-h1n1)	N/A		204
TOTAL			18569
WHO			
WHO IRIS (https://apps.who.int/iris/)			
Title: recommendations CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	86
Title: recommendation CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	86
Title: guidelines CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	15
Title: guideline CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	15

Title: recommendations CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	1
Title: recommendation CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	1
Title: guidelines CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	2
Title: guideline CONTAINS Subject: H1N!	None/All of IRIS	17 March 2022	2
Subject: recommendations CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	0
Subject: recommendation CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	0
Subject: guidelines CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	538
Subject: guideline CONTAINS Subject: COVID-19	None/All of IRIS	17 March 2022	538
Subject: recommendations CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	0
Subject: recommendation CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	0
Subject: guidelines CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	6
Subject: guideline CONTAINS Subject: H1N1	None/All of IRIS	17 March 2022	6
Title: recommendations CONTAINS Subject: Pandemic	None/All of IRIS	17 March 2022	3
Title: recommendation CONTAINS Subject: Pandemic	None/All of IRIS	17 March 2022	3
Title: guidelines CONTAINS Subject: Pandemic	None/All of IRIS	17 March 2022	2
Title: guideline CONTAINS Subject: Pandemic	None/All of IRIS	17 March 2022	2

Subject: recommendations CONTAINS subject: pandemic	None/All of IRIS	17 March 2022	0
Subject: recommendation CONTAINS subject: pandemic	None/All of IRIS	17 March 2022	0
Subject: guidelines CONTAINS subject: pandemic	None/All of IRIS	17 March 2022	60
Subject: guideline CONTAINS subject: pandemic	None/All of IRIS	17 March 2022	60
H1N1 webpage (https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak)	N/A		121
WHO Committee Approved Guidelines (https://www.ncbi.nlm.nih.gov/books/NBK132015/)	N/A		337
TOTAL			1884

CDC – United States Centers for Disease Control and Prevention, ECDC – European Centres for Disease Control and Prevention, WHO – World Health Organization, MMWR – MMWR. Morbidity and Mortality Weekly Report, N/A – not applicable

*No limitation set to time span unless otherwise indicated.

Table S2. Interrater agreement for screening process in AGREE-HS analysis

	Total	Included	Excluded	Included	Excluded	Included	Excluded	IRR (% agreement)	Strength of agreement
ECDC*		**R1**	**R1**	**R2**	**R2**				
Title and abstract screening	5640	201	5439	191	5449			0.87 (99.10)	Strong
Full-text screening	219	54	165	55	164			0.81 (93.20)	Strong
WHO*		**R1**	**R1**	**R2**	**R2**				
Title and abstract screening	1057	266	791	277	780			0.93 (97.60)	Almost Perfect
Full-text screening	268	142	126	154	114			0.86 (93.30)	Strong
CDC*		**R1**	**R1**	**R2**	**R2**				
Title and abstract screening	17147	780	16367	791	16356			0.98 (99.90)	Almost Perfect
Full-text screening	817	189	628	171	646			0.70 (90)	Moderate
AGREE pilot†		**R1**	**R1**	**R2**	**R2**	**R3**	**R3**		
	195	111	84	106	89	115	80	0.70 (78.50)	Substantial

IRR – interrater agreement

*Cohen's kappa. Strength of agreement categorised following McHugh's interpretation (66).

†Fleiss' kappa for multiple raters. Strength of agreement categorised following Landis and Koch's interpretation (67).

Table S3. Overall results for the five domains assessed in the AGREE-HS analysis

	AGREE-HS score, MD (95% CI)
Topic	6.00 (5.00-6.00)
Participants	2.00 (1.00-2.00)
Methods	2.00 (2.00-3.00)
Recommendations	3.00 (3.00-4.00)
Implementability	3.00 (3.00-3.00)

CI – confidence interval, MD – median

Table S4. Model fit for ordinal regression used in AGREE-HS analysis

Model	Deviance	AIC	McFadden's R^2	Overall model test		
				χ^2	df	*P*-value
1	195	209	0.479	179	5	<0.001

df – degrees of freedom

Text S1. Details on the scraping and deduplication process for AGREE-HS analysis

Python: Using ECDC's search engine, we performed 18 queries applying the following keywords: "COVID-19", "COVID-19 recommendations", "COVID-19 guidelines", "H1N1", "H1N1 recommendations", "H1N1 guidelines", "pandemic recommendations", and "pandemic guidelines", and the following filters: "Severe acute respiratory syndrome (SARS)", "SARS-CoV-2", and "SARS-CoV-2 variants", "COVID-19", "Coronavirus", "Influenza A (H1N1)", "Influenza A (H1N1)2009", "Swine-origin influenza", "Influenza in Humans, Swine Origin", and "Influenza in humans, pandemic".

For each query we used its default URL address and concatenated the address with page numbers. For example, the default address for the "COVID-19 recommendations" section was concatenated with numbers ranging from 0 to 217, as 216 was the number of pages for "COVID-19 recommendations" query at the moment when the queries were performed. After generating query website URLs, we found 10 sections on each website that presented information on articles such as their title, short abstracts, and hyperlinks. In order to retrieve the hyperlinks, we scraped the obtained URL pages using requests and the "Beautiful Soup" and "HTTP" libraries for Python. We used the "fromkeys" function in Python to deduplicate the results. Articles' titles and hyperlinks were saved in a tabular pandas.DataFrame data structure and exported to a .xlsx file.

The process of querying was done automatically on 15 May 2022, for which we wrote a script in Python, version 3.8.8 (Python Software Foundation, Delaware, USA). A total of 18,363 articles were retrieved; 5,445 were left after deduplication across searches.

R: We used R, version 4.2.1 (R Core Team, Auckland, New Zealand) to scrape the "Publications" section of the H1N1 dedicated website using the "RSelenium" and "tidyverse" packages. We then selected the "Publications" subsection and iterated over the "Load More" button to reveal all possible articles, after which we scraped their titles and hyperlinks into a tabular format within R, which we then exported as a .csv. Through this method, we retrieved a total of 204 articles; 195 remained following deduplication.

Text S2. Full search strategy for the linguistic analysis of editorials

Web of Science – Web of Science Core Collection, MEDLINE (1 June 2024)

1. TS=(government* OR ministr* OR agenc* OR organi*ation* OR institut* OR department* OR nation* OR countr* OR state* OR glob*) – 19,079,902 results
2. TS=(guideline* OR recommend* OR law* OR regulation* OR legislat* OR polic* OR mandat* OR respons* OR prepar* OR manag* OR control* OR measur* OR surveillance OR prevent*) – 37,398,297 results
3. TS=(covid* OR corona* OR SARS-CoV-2 OR H1N1 OR "swine *flu*" OR pandemic) – 1,862,736 results
4. #1 AND #2 AND #3 – 415,230 results
5. #1 AND #2 AND #3 and Editorial Material (Document Types) – 14,052 results
6. #1 AND #2 and Editorial Material (Document Types) and 2024 or 2023 or 2022 or 2021 or 2020 or 2019 or 2009 or 2010 or 2011 or 2012 or 2013 or 2014 (Publication Years) – 11,980 results

Scopus (1 June 2024)

1. TITLE-ABS-KEY (government* OR ministr* OR agenc* OR organi*ation* OR institut* OR department* OR nation* OR countr* OR state* OR glob*) -18,795,909 results
2. TITLE-ABS-KEY (guideline* OR recommend* OR law* OR regulation* OR legislat* OR polic* OR mandat* OR respons* OR prepar* OR manag* OR control* OR measur* OR surveillance OR prevent*) – 43,248,631 results
3. TITLE-ABS-KEY (covid* OR corona* OR sars-cov-2 OR h1n1 OR "swine *flu*" OR pandemic) – 1,703,259 results
4. (TITLE-ABS-KEY(government* OR ministr* OR agenc* OR organi*ation* OR institut* OR department* OR nation* OR countr* OR state* OR glob*)) AND (TITLE-ABS-KEY(guideline* OR recommend* OR law* OR regulation* OR legislat* OR polic* OR mandat* OR respons* OR prepar* OR manag* OR control* OR measur*

OR surveillance OR prevent*)) AND (TITLE-ABS-KEY(covid* OR corona* OR sars-cov-2 OR h1n1 OR "swine *flu*")) – 371,323 results

5. (TITLE-ABS-KEY(government* OR ministr* OR agenc* OR organi*ation* OR institut* OR department* OR nation* OR countr* OR state* OR glob*)) AND (TITLE-ABS-KEY(guideline* OR recommend* OR law* OR regulation* OR legislat* OR polic* OR mandat* OR respons* OR prepar* OR manag* OR control* OR measur* OR surveillance OR prevent*)) AND (TITLE-ABS-KEY(covid* OR corona* OR sars-cov-2 OR h1n1 OR "swine *flu*")) AND (LIMIT-TO (DOCTYPE,"ed")) – 6,127 results
6. (TITLE-ABS-KEY(government* OR ministr* OR agenc* OR organi*ation* OR institut* OR department* OR nation* OR countr* OR state* OR glob*)) AND (TITLE-ABS-KEY(guideline* OR recommend* OR law* OR regulation* OR legislat* OR polic* OR mandat* OR respons* OR prepar* OR manag* OR control* OR measur* OR surveillance OR prevent*)) AND (TITLE-ABS-KEY(covid* OR corona* OR sars-cov-2 OR h1n1 OR "swine *flu*")) AND (LIMIT-TO (DOCTYPE,"ed")) AND (LIMIT-TO (PUBYEAR,2009) OR LIMIT-TO (PUBYEAR,2010) OR LIMIT-TO (PUBYEAR,2011) OR LIMIT-TO (PUBYEAR,2012) OR LIMIT-TO (PUBYEAR,2013) OR LIMIT-TO (PUBYEAR,2014) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2022) OR LIMIT-TO (PUBYEAR,2023) OR LIMIT-TO (PUBYEAR,2024)) – 5,170 results

Text S3. Inclusion and exclusion criteria for the linguistic analysis of editorials

Inclusion criteria

1. Must be an editorial, viewpoint, or similar type of opinion piece;
2. Must discuss the healthcare system response to either the H1N1 or COVID-19 pandemic at local, regional, national, or global levels, including, but not limited to;
    i) Strategies and approaches related to the implementation of nonpharmaceutical measures such as masking, social distancing, isolation/quarantine, *etc*., with a focus on population-/policy-level discussions of efficacy, effectiveness, adherence, sustainability, cost-effectiveness, *etc*. This includes any discussions on mandates, legal regulations, and similar enforcements of said interventions and measures for the general population or specific subgroups, including vaccine mandates, mask mandates, travel restrictions, lockdowns (inclusive of school lockdowns), *etc*.;
    ii) Impact of misinformation/disinformation on the implementation of said measures;
    iii) Healthcare resource allocation and management, including any aspects related to the management and distribution of human resources, vaccines, therapeutics or pharmaceuticals in general, respirators, personal protective equipment, and other resources;
    iv) Aspects of healthcare system preparedness, including monitoring and surveillance, response mechanisms, or regulations (*i.e.* International Health Regulations, "pandemic treaties", laws, *etc*.);
3. Discussions on policies, mandates, and guidance related to education, research, or technologies;
    i) *e.g.* school lockdowns, distance learning, research on H1N1 and COVID-19 specifically at the research ecosystem level and how it is being funded or conducted (*e.g.* need for more clinical trials, modelling *vs*. observational evidence), the application of new technologies in the development of therapeutics/vaccines for H1N1 and COVID, or the delivery of health services at the health system/global level (such as global changes or prospects of telemedicine), or surveillance/tracing;
    ii) "Post-COVID" or "post-H1N1" topics, as long as they reflect on the healthcare response during the pandemics and consider a "lessons learned" approach;
    iii) Specific political institutions or politicians and their work in the context of the healthcare response to either pandemic.

Exclusion criteria

1. Original scientific research and similar reports, including reviews of any kind, case studies, case reports, etc, as well as non-pandemic related editorials, audio-visual materials, textual interviews, reports on the findings of a single study, or guidelines/recommendations/frameworks;
2. Editorials, viewpoints, or similar types of opinion pieces discussing:
    i) Clinical aspects of H1N1, COVID-19, vaccination, treatment, or therapeutics at the individual level (*e.g.* discussing the efficacy of ivermectin), as well as discussions on the virological characteristics of SARS-CoV-2 (exclusive of aspects listed in inclusion criteria 2c, *e.g.* distribution of therapeutics among countries);
    ii) Impact of H1N1 and COVID-19 on healthcare services unrelated to the pandemics, as long as this impact is at the system level (*e.g.* impact of COVID-19 on malaria screening);
    iii) Discussions of the pandemic in relation to singular events, *e.g.* World Championships or Olympic Games.
    iv) The effect of H1N1 and COVID-19 on other health conditions, or considerations of H1N1 and COVID-19 in the context of comorbidities;
    v) The impact of COVID-19 on specific medical specialties outside of the health system-level context (*e.g.* the impact of the pandemic on cardiovascular surgeries at a specific hospital, or how it impacted dermatology as a discipline specifically, as well as discussions of testing of surgical patients for COVID-19 and so on);
    vi) The impact of the pandemic on education, research, and technology unrelated to policies, mandates, and guidance; outside of the national/global context; or not specific to the two pandemics (*e.g.* explanations of how a specific course at a single university was adapted during the pandemic; discussions on the impact of COVID-19 on non-COVID-19-related research; the use of novel technology due to H1N1/COVID-19 (such as the adoption of telemedicine) at a single hospital or within a single discipline);
    vii) A specific set of articles within a journal or a thematic issue, even if they are related to H1N1 or COVID-19;
    viii) Societal impacts of H1N1 or COVID-19 outside of healthcare (*e.g.* impact on the job market, on economies in general, on domestic violence and abuse).
3. Other (free input, anything outside of the criteria listed above).

Table S5. Characteristics of the studies cited in support of the recommendations found in the levels of evidence study, n (%)*

	First period	Second period	Third period	Fourth period
Type of publication				
Published in a journal	77 (26.2)	225 (58.3)	23 (53.5)	55 (61.8)
Preprint	19 (6.5)	40 (10.4)	6 (14.0)	16 (18.0)
Other guideline/HSG	52 (17.7)	41 (10.6)	6 (14.0)	4 (4.5)
Other literature†	146 (49.7)	80 (20.7)	8 (18.6)	14 (15.7)
Type of study				
Viewpoints and non-structured literature reviews	21 (21.9)	36 (13.6)	3 (10.3)	3 (4.2)
Bench research	4 (4.2)	17 (6.4)	1 (3.4)	3 (4.2)
Case study	6 (6.3)	13 (4.9)	2 (6.9)	4 (5.6)
Case series	15 (15.6)	32 (12.1)	1 (3.4)	7 (9.9)
Modelling study	15 (15.6)	40 (15.1)	7 (24.1)	6 (8.5)
Qualitative study	1 (1)	0 (0)	0 (0)	0 (0)
Cross-sectional study	11 (11.5)	55 (20.8)	4 (13.8)	8 (11.3)
Cross-sectional study with modelling component	1 (1)	5 (1.9)	0 (0)	1 (1.4)
Cohort study	9 (9.4)	31 (11.7)	4 (13.8)	23 (32.4)
Cohort study with modelling component	1 (1)	5 (1.9)	0 (0)	2 (2.8)
Case-control study	0 (0)	3 (1.1)	2 (6.9)	4 (5.6)
Randomised controlled trial	0 (0)	2 (0.8)	4 (13.8)	0 (0)
Systematic, rapid, narrative, and literature reviews without synthesis or meta-analysis	7 (7.3)	15 (5.7)	1 (3.4)	2 (2.8)
Systematic review with meta-analysis or synthesis	5 (5.2)	11 (4.2)	0 (0)	8 (11.3)

*Each period spanned 185 days from the publishing of the first HSG in the sample (7 May 2020) to the last one (28 January 2022), as this corresponded to the median update period for the HSG in our sample.

†Includes policy documents, risk assessments, risk reports, diagnostic test evaluations/standards, tools/checklists, news articles, etc.

Table S6. Full results of preliminary LIWC analysis of editorials, MD (IQR)

	COVID	H1N1	Statistic	*P*-value
Word count	1332 (947–1726)	838 (667–1207)	1104	<0.001
Analytical thinking	92.8 (89.1–95.4)	90.3 (85.3–95.1)	1651	0.083
Clout	43.2 (35.4–51.2)	41.9 (35.2–50.2)	2080	0.904
Authentic	32.9 (21.8–45.7)	31.3 (25.0–43.3)	2060	0.846
Tone	28.0 (17.1–37.2)	15.6 (7.97–25.4)	1336	0.004
Words per sentence	26.6 (23.6–30.3)	25.5 (23.3–27.8)	1694	0.116
Big words	36.6 (33.6–39.6)	31.7 (30.9–35.4)	1035	<0.001
Drives	5.53 (4.38–6.45)	5.20 (3.51–6.05)	1750	0.174
Affiliation	1.22 (0.780–1.79)	1.05 (0.662–1.45)	1775	0.205
Achieve	1.63 (1.17–2.24)	1.55 (0.992–2.40)	1901	0.428
Power	2.24 (1.61–2.97)	1.79 (1.55–2.69)	1769	0.198
Cognition	18.0 (16.2–19.6)	21.2 (19.3–22.5)	925	<0.001
All-or-none thinking	0.340 (0.210–0.560)	0.400 (0.150–0.622)	2111	0.997
Cognitive processes	10.5 (8.66–12.1)	11.1 (9.89–13.3)	1685	0.108
Insight	2.14 (1.57–2.73)	2.23 (1.67–3.29)	1956	0.558
Causation	2.13 (1.68–2.56)	2.04 (1.55–2.50)	1890	0.405
Discrepancy	1.22 (0.795–1.70)	1.79 (0.940–2.05)	1672	0.098
Tentativeness	1.62 (1.21–2.17)	1.85 (1.43–2.34)	1855	0.334
Certitude	0.230 (0.120–0.370)	0.390 (0.165–0.655)	1429	0.010
Differ	2.95 (2.29–3.59)	3.19 (2.64–3.49)	1915	0.459
Memory	0.00 (0.00–0.00)	0.00 (0.00–0.103)	1622	0.009
Affect	4.15 (3.37–5.03)	4.38 (3.39–5.46)	1920	0.472
Positive tone	2.32 (1.72–2.77)	2.09 (1.34–2.82)	1792	0.229

Negative tone	1.69 (1.22–2.31)	2.21 (1.74–3.07)	1391	0.007
Emotion	0.400 (0.240–0.660)	0.520 (0.368–0.835)	1632	0.071
Positive emotion	0.0850 (0.00–0.163)	0.145 (0.108–0.235)	1407	0.007
Negative emotion	0.220 (0.110–0.403)	0.155 (0.0525–0.34)	1732	0.153
Anxiety	0.0750 (0.00–0.170)	0.0800 (0.00–0.237)	2105	0.978
Anger	0.00 (0.00–0.0900)	0.00 (0.00–0.00)	1581	0.023
Sadness	0.00 (0.00–0.0700)	0.00 (0.00–0.00)	1746	0.108
Swear words	0.00 (0.00–0.00)	0.00 (0.00–0.00)	2048	0.257
Social processes	7.20 (6.12–8.66)	6.82 (5.69–10.4)	2064	0.857
Social behaviour	3.10 (2.28–4.23)	3.55 (2.34–4.68)	1928	0.490
Prosocial behaviour	0.810 (0.458–1.34)	0.800 (0.425–1.01)	1888	0.400
Politeness	0.00 (0.00–0.0800)	0.0200 (0.00–0.145)	1776	0.156
Interpersonal conflict	0.140 (0.060–0.263)	0.255 (0.00–0.455)	1846	0.315
Moralisation	0.120 (0.00–0.292)	0.155 (0.00–0.445)	2079	0.900
Communication	0.840 (0.508–1.24)	1.30 (0.722–1.87)	1464	0.015
Culture	1.40 (0.750–2.38)	0.935 (0.560–1.53)	1623	0.066
Politics	0.790 (0.378–1.53)	0.690 (0.465–1.30)	2048	0.810
Ethnicity	0.110 (0.00–0.330)	0.0350 (0.00–0.152)	1682	0.097
Technology	0.195 (0.080–0.388)	0.125 (0.0375–0.19)	1582	0.046
Need	0.945 (0.617–1.23)	0.615 (0.402–0.818)	1431	0.010
Want	0.00 (0.00–0.0625)	0.00 (0.00–0.128)	1884	0.311
Acquire	0.345 (0.190–0.532)	0.360 (0.228–0.637)	1932	0.498
Lack	0.240 (0.120–0.415)	0.165 (0.075–0.343)	1801	0.242
Fulfilment	0.125 (0.050–0.212)	0.0700 (0.00–0.290)	1891	0.402
Fatigue	0.00 (0.00–0.00)	0.00 (0.00–0.00)	1878	0.139

Reward	0.205 (0.090–0.352)	0.145 (0.00–0.300)	1665	0.092
Risk	0.980 (0.645–1.40)	0.860 (0.542–1.29)	1947	0.535
Curiosity	0.250 (0.117–0.480)	0.435 (0.165–0.958)	1436	0.011
Allure	2.82 (2.23–3.55)	3.10 (2.54–3.89)	1653	0.084

IQR – interquartile range, MD – median