Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Lexical salience and inference boundaries in Kompas.com COVID-19 reporting during 2021

Delfan Putra
delfaneko@unib.ac.id
University of Bengkulu
Meiselina Irmayanti
University of Bengkulu
Abdul Muktadir
University of Bengkulu
Nur Najla Zainal Anuar
Rabdan Academy
Taufik Hasyim
University of Bengkulu

Article

Keywords: crisis communication, lexical salience, Indonesian news, computational text analysis, moral panic

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11552803/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Lexical prominence in a news archive does not by itself establish moral panic. This study reassesses the documented analytical record of an earlier corpus analysis of Kompas.com COVID-19 reporting during 2021. It distinguishes reproducible arithmetic from reported lexical measurements and unverified descriptions of software outputs. The accessible empirical materials are the manuscript's tables, narrative and four embedded figures; the raw and cleaned news records, processing scripts and eight separately cited Voyant exports were not available in the revision materials. Monthly entries sum to 22,531 headline records and 649 article records, compared with stated totals of 22,491 and 740. These sums verify internal discrepancies rather than the underlying corpus size. Three embedded count labels rank COVID at 3,160, cases at 1,298 and health at 805, but their counting unit remains unresolved. The documented associations foreground pandemic, prevention and institutional vocabulary without establishing hostility, disproportionate concern or audience anxiety. Descriptions of segmented or non-chronologically ordered Voyant displays cannot support calendar trends, and the underlying displays require recovery before independent inspection. Unresolved selection rules, translation procedures, denominators and co-occurrence settings prevent reproduction of the central corpus findings. The contribution is a case-specific account of evidence requirements connecting lexical description, contextual moral attribution and social consequences, rather than an empirically validated model of moral panic. Record reconciliation and recovery of the original analytical materials remain necessary before corpus-level claims can be verified.

Introduction

News coverage of a public-health emergency makes a material threat publicly intelligible: it names the disease, enumerates infections, explains preventive measures and reports decisions that reorganise everyday life. Yet the vocabulary necessary to describe an emergency can also appear in narratives that exaggerate danger, assign blame or legitimise punitive intervention. The analytical difficulty is to establish which of these functions a text performs. Repeated references to COVID-19 demonstrate attention to the disease; they do not, by themselves, establish moral panic. For computational communication research, this distinction concerns the validity of the inference connecting an observable linguistic feature to a sociological explanation.

Indonesia in 2021 provides a consequential setting in which to examine that problem. In its situation report of 21 July, the World Health Organization recorded 2,983,830 cumulative confirmed cases and 77,583 deaths, including 33,772 newly reported cases and 1,383 newly reported deaths that day. During 12–18 July, 32 of the country's 34 provinces reported increasing case numbers (World Health Organization, 2021). These surveillance figures describe reported disease burden, rather than the full extent of infection or mortality. They nevertheless establish why intensive reporting could have substantial public value. A claim that coverage was disproportionate must therefore examine what journalists asserted against credible information available at publication, rather than take the prominence of disease vocabulary as evidence of exaggeration.

Communication research identifies several reasons why such attention matters without assigning it a uniform effect. Malecki et al. (2021) situate risk perception within the social conditions of crisis communication. Chu et al. (2022) report differentiated relationships between pandemic news exposure, psychological responses, disease concern and preventive behaviour. Garfin et al. (2020) identify pathways through which repeated media exposure may amplify distress, while Pitol et al. (2023) examine media consumption and mental health among young adults in Bangladesh. These studies justify investigating communication during emergencies, but their exposure and response measures cannot be replaced by counts of words in a news archive. The present study consequently distinguishes the subjects made visible in reporting from the experiences of the people who encounter it.

Cross-country surveys also show why apparently similar health vocabulary can address different practical concerns. In June 2020, Lazarus et al. (2021) surveyed 13,426 respondents in 19 countries and found that 71.5% expressed potential acceptance of a COVID-19 vaccine. A subsequent survey of 23,000 respondents across 23 countries in June 2021 reported acceptance of 75.2% (Lazarus et al., 2022). These figures should not be read as a directly comparable global trend: the country coverage changed, and the 2021 definition included receipt of at least one dose as well as willingness among unvaccinated respondents, whereas the earlier survey measured hypothetical acceptance. Their relevance here is that trust, perceived safety and preventive decisions require explicit measurement. The occurrence of vaccine or vaccination in news cannot disclose whether a passage promotes vaccination, questions its safety, reports access barriers or corrects misinformation.

An Indonesian survey provides a more situated context. The Ministry of Health, NITAG, UNICEF and WHO (2020) conducted an online vaccine acceptance survey on 19–30 September 2020, obtaining responses from more than 115,000 people across all 34 provinces. For the acceptance item, the report identifies 112,888 respondents: approximately 65% indicated willingness to receive a government-provided vaccine, nearly 8% refusal, and more than 27% hesitation. These are findings among online respondents before the 2021 reporting period, with question-specific denominators and participation limitations; they are not population estimates for Kompas.com readers. They nonetheless identify an information environment in which uncertainty and prevention coexisted. Research on Indonesian governmental efficacy information and online responses to restrictions further demonstrates that institutional messaging and public reactions are distinct empirical objects (Rosemary et al., 2022; Tri Sakti et al., 2021).

Moral panic and the evidence for social attribution

Moral panic supplies a demanding framework for analysing these relationships. Cohen (1972) connects representations of a threat to social values with the construction of deviant figures and processes of societal reaction. Goode and Ben-Yehuda (2009) distinguish concern, hostility, consensus, disproportionality and volatility. These dimensions extend beyond the frequency with which a danger is named. A disease can attract intense, justified concern while particular groups are simultaneously blamed or stigmatised. Identifying the latter process requires evidence about who is represented as

threatening, how responsibility is attributed, whose moral judgement is voiced and what response is advocated.

Earlier epidemic research underscores the importance of examining the forms that threat narratives take. Washer (2004) analyses representations of SARS in British newspapers, including the construction of social difference, while Ungar (2008) examines both alarming and reassuring coverage of bird flu. Neither line of inquiry supports treating a disease designation as a self-evident indicator of panic. More recent COVID-19 research offers specific examples of moral attribution: Capurro et al. (2022) examine the construction of “covidiots” in Canadian newspapers, and Cârstocea (2022) analyses representations of Roma communities as public-health threats. These studies make actors, evaluations and social boundaries visible. They provide comparative questions for Indonesian reporting, rather than evidence that the same processes occurred in the present collections.

The possibility of inadequate concern must also remain open. Hier (2021) examines denial in pre-crisis risk communication, while Cottingham (2024) distinguishes representations of fear in Ebola news from emotions expressed in social media responses. Together, these contributions caution against interpreting urgency as inherently excessive or assuming that a text and its reception share the same emotional character. Disproportionality requires a specified comparison, and audience fear requires evidence about audiences. Neither can be established through a graph of lexical prominence alone.

Digital circulation introduces further distinctions. Walsh (2020) examines the changing relationship between social media and moral panic, whereas Cinelli et al. (2020) and Gallotti et al. (2020) investigate information diffusion and infodemic risks through platform-related evidence. A publisher's archive records available texts; it does not measure recommendation systems, readership, exposure or interpersonal transmission. Accordingly, the presence of repeated terms in Kompas.com reporting is analytically separable from claims about algorithmic amplification, commercial intention or the spread of fear. Those mechanisms require observations beyond the archive.

Lexical salience across headlines and news texts

Computational text analysis can help locate the passages and patterns through which these questions become tractable, provided its measurements are aligned with the intended claim. Grimmer et al. (2021) distinguish concept discovery, measurement, prediction and causal assessment, each with different evidentiary requirements. Nelson (2020) connects computational pattern detection with interpretive refinement and confirmation; Baden et al. (2022) identify gaps between computational methods and social-scientific concepts. These approaches support using frequent terms to guide contextual inquiry while retaining a clear boundary between an exploratory pattern and a validated interpretation.

The earlier analytical account describes Voyant Tools Cirrus, Links, Trends and StreamGraph outputs for collections labelled News and Headlines. It reports disease, prevention, governmental and geographical vocabulary, alongside police and political terminology in Headlines and attribution or publication terms in News. The eight exports cited in that account are absent from the materials available for this revision.

Their labels, relative type sizes and settings therefore remain reported observations rather than independently inspected evidence. Those descriptions motivate questions about topic inclusion and retained publication material, but cannot establish how much of either collection concerns COVID-19 or support quantitative comparison between independently generated word clouds.

Headlines and article texts also perform different communicative functions. Headlines condense and select information; article texts can include attribution, qualifications, explanatory detail and competing voices. Their comparison can therefore help identify where lexical visibility warrants closer examination. However, the reported headline retrieval and first-Monday article sampling do not establish matched headline–article pairs. Differences between the exports may reflect sampling, document length, preprocessing or source aggregation as well as editorial presentation. This study treats the collections separately and does not interpret an unmatched comparison as evidence that headlines amplify fear relative to their articles.

The meaning of distributional graphics depends on their units. The earlier account describes ten segments of a single document in Headlines Trends and 18 non-chronologically ordered document labels in News Trends. If accurately described, those positions cannot establish daily or monthly change without publication-date mapping. Because the exports are unavailable here, those specifications also require verification. Described Links connections similarly cannot establish intentions, hostility or causal relationships from their topology alone.

Language is a further source of interpretive uncertainty. The reported source material is Indonesian, while the documented labels and the embedded count figure use English lexical forms. Translation timing, tokenisation, case folding, stopwords and the handling of COVID-19 variants can affect both frequencies and co-occurrences. The term positive illustrates the problem: in a health report it can designate a diagnostic result rather than a favourable attitude. The validation concerns raised by van Atteveldt et al. (2021) are therefore relevant even where sentiment is not formally classified. Kania's (2022) corpus-assisted analysis of disease naming offers a more specific model for connecting lexical choices with contextual interpretation. A valid Indonesian application would retain original expressions and distinguish an outlet's endorsement from quotation, correction or rejection.

Research questions and contribution

The problem addressed here is how to specify the evidentiary reach of lexical analysis in a severe public-health emergency. Existing work explains why media exposure, institutional trust, moral attribution and digital circulation matter, but does not validate an inference from disease-word frequency to moral panic in a particular archive. The Kompas.com case is examined through the documented collection procedures, numerical descriptions, tables and four embedded figures in the earlier manuscript. This revision reassesses that analytical record; it does not reconstruct the annual news corpus or verify the separately cited Voyant exports. The distinction is made explicit so that the original collection remains the study's historical basis without being represented as a newly reproduced dataset.

Three research questions guide the study. RQ1: What corpus construction and measurement procedures are documented, and which unresolved discrepancies constrain the interpretation of lexical salience? RQ2: What lexical prominence and distributional patterns are visible in the reported headline and news outputs, and how do their units and settings limit comparison? RQ3: Which claims about moralising crisis communication can these materials support, and what additional contextual or audience evidence would be needed to assess moral panic? These questions retain the Indonesian case while avoiding assumptions about the direction or magnitude of media effects.

The contribution is an explicit account of the passage from lexical observation to social inference. It distinguishes the visibility of an emergency from the moral evaluation of actors, shows why corpus provenance and display units affect interpretation, and identifies the evidence needed before claims can extend to audience consequences or disproportionate concern. This approach supports a situated reading of Kompas.com reporting while keeping the conclusions proportionate to the available materials. It also provides a basis for a subsequent reproducible analysis in which source-language passages, documented transformations and verified dates can test competing interpretations of the same vocabulary.

Methods

Design and evidence status

This is a documentary reassessment of the study's own reported corpus analysis. The accessible evidence consists of the earlier manuscript's narrative, four editable tables and four embedded figures. The editorial letter and journal instructions govern the revision but are not news data. The manuscript describes eight additional Voyant exports; neither those files nor raw or cleaned article records, retrieval logs, processing macros or settings exports were available for independent examination. No new news collection, participant recruitment, contextual coding or statistical testing was performed. The reassessment checks arithmetic, compares printed labels with prose, and evaluates whether documented units support the claims attached to them. It does not authenticate those labels against source articles. Findings below distinguish directly checked documentation from statements inherited from the earlier account and from proposed future analyses. This scope is a limitation of the present revision and requires author and editorial confirmation before resubmission as an original research article.

The reported collection design

The manuscript describes headline collection from https://indeks.kompas.com/headline/?site over 1 January to 31 December 2021. It reports 22,491 headline records. Article collection is described separately, using https://indeks.kompas.com/?site on the first Monday of each month, retaining material from the first five index pages. It reports 740 articles. Data were reportedly retrieved through R/RStudio using HTTP requests, parsed and stored in CSV files. RStudio is an environment rather than a collection

algorithm, so the description identifies a tool setting without identifying the packages, selectors or functions used.

The inclusion rule is unresolved. Some passages describe the headline material as the year's headline collection; others refer to COVID-related records or moral-panic-related content. No retrieval query, explicit inclusion decision, exclusion log or topic-filter expression establishes whether all indexed headlines were retained. Nor is it clear whether article collection was restricted to entries corresponding to collected headlines. The two collections are therefore treated as related by source and period, but their overlap, nesting and topic scope are unknown. Table 1 preserves these uncertainties alongside the conflicting totals.

Table 1
Corpus provenance and reconciliation

Corpus	Source	Period and selection	Reported total	Monthly sum	Inclusion criteria	Unresolved discrepancy
Headlines	Kompas.com headline index	January–December 2021; year-long collection reported	22,491 records	22,531 records	All indexed versus COVID-related headlines unresolved; query and deduplication rule absent	Monthly sum exceeds stated total by 40; underlying total not resolved
Articles	Kompas.com general index	First Monday monthly; first five pages; 2021	740 articles	649 records	Topic filter, record identity and linkage to headlines unresolved	Monthly sum is 91 below stated total; January 222 requires log check
Cross-period material	Source described as Kompas.com	COVID in July 2021 versus Jokowi in February 2024	Not established	Not available	Different topics, periods and unknown corpus composition	Excluded from principal argument; common denominators absent

Sums recalculated from the monthly values documented in the analytical account and reproduced in Fig. 3. These are checks of the manuscript record, not verified counts of unique source documents. Monthly headline entries: 1,836; 1,731; 1,948; 1,905; 1,797; 1,904; 1,926; 1,922; 1,777; 1,853; 1,905; 2,027. Article entries: 222; 24; 23; 35; 49; 20; 89; 57; 47; 32; 29; 22. No corrected corpus size is asserted.

The first-Monday procedure selects twelve calendar occasions, not continuous article coverage. If applied literally in 2021, those dates would be 4 January, 1 February, 1 March, 5 April, 3 May, 7 June, 5 July, 2 August, 6 September, 4 October, 1 November and 6 December. These are calendar-derived expectations,

not verified retrieval dates. Collection logs must establish whether the procedure followed these dates and whether publication dates, retrieval dates or website date filters determined eligibility. The current revision does not substitute this calculated calendar for missing metadata.

Selecting the first five pages also requires an account of what a page represents. Chronological ordering, relevance ranking, editorial prioritisation and technical pagination would each produce a different sampling frame. The number of records per page and the treatment of updated or repeated entries are not documented. Index position cannot be equated with homepage exposure or public visibility without evidence. The reported strategy is therefore described as a fixed weekday and page-position sample, with unknown inclusion probabilities, rather than a representative sample of annual coverage.

No sampling justification establishes that Mondays represent journalism across the week or that five index pages capture the relevant population. Weekday restriction can omit weekend reporting, and a single occasion per month can miss major developments. Page limits can exclude stories outside the selected position range. The direction and magnitude of resulting selection bias cannot be estimated from the accessible documentation. Even accurate token counts would describe the retained material rather than automatically characterise the outlet's annual article population.

Processing and measurement documentation

The source manuscript states that raw texts were collected in Indonesian, manually checked for errors and HTML artefacts, and processed through VBA. It mentions term correction, special-character removal and combining material. It also lists capabilities attributed to Voyant Tools, including stopword filtering, punctuation handling and duplicate removal. A list of software capabilities does not show which operations were enabled. No processing script or settings export establishes the token boundary rules, case folding, numerical filtering, Indonesian stopword list, duplicate criteria or treatment of inflected forms.

Translation timing is particularly uncertain. The English labels cases, health, home and taskforce may represent translations of Indonesian terms, outputs from a translated corpus, or interpretive categories combining variants. Each possibility produces a different measurement. Translation before counting can collapse different expressions or introduce new lexical forms. Translation after counting can preserve an Indonesian measurement while obscuring its exact spelling to readers. In the current figures, labels follow the reported manuscript and are explicitly identified as such; they are not asserted to be original Indonesian tokens.

COVID, Covid-19 and COVID-19 require a declared normalisation rule. Case folding alone would not necessarily merge punctuation and number variants. Removing punctuation might split a hyphenated expression into separate components, while removing numbers could transform it again. Stemming or lemmatisation could alter other terms, but neither operation is documented sufficiently to be treated as performed. Combining multiple records into one file could also alter document boundaries and collocation counts. Table 2 separates each documented step from the settings needed for reproduction.

Table 2
Processing and measurement audit

Step	Documented procedure	Required for reproduction	Unit or denominator	Limitation
Retrieval	R/RStudio; HTTP requests; CSV storage	Scripts, packages, selectors, query, time stamps and failure log	Retrieved records	Completeness and inclusion unknown
Cleaning	Manual checks and VBA processing reported	Macros; change log; raw-to-clean mapping	Characters and records	Specific replacements unavailable
Language	Indonesian collection; English output labels	Original forms; translation timing and mapping	Source-language lexical forms	Analysis language unresolved
Tokenisation	General software functions described	Token rules; case and COVID-19 variants; stopwords; stemming decision	Tokens after declared processing	Enabled settings not established
Duplicates	Removal claimed	Identity rule; URLs; exact and near-duplicate logs	Unique documents versus records	No verified unique total
Frequency	Counts and percentages displayed	Full export and percentage base	Tokens, documents or selected-term total	Percentage denominator unspecified
Collocation	Associations and bigrams reported	Direction; window; boundaries; filters; association metric	Pairs within declared context	Reverse-direction values differ
Trends	Daily VBA trend claimed; monthly caption	Date-level counts and coverage; interval denominator	Counts per day; prevalence if verified	No daily input; no event test
Validation	Cross-tool correlation claimed	Aligned vectors; statistic; n; settings; output	Matched terms or dates	Unverified; cannot validate moral panic

Requirements are proposed recovery steps, not completed analyses. No new sentiment, topic-model or reliability result is reported.

Several units must remain distinct. A record is an entry obtained by the retrieval process; a unique article is a deduplicated item under an explicit identity rule; a token is an occurrence after processing; document frequency counts documents containing a term; and a co-occurrence count depends on its context definition. The phrase mentions does not identify which unit was counted. Headline records cannot be used as token denominators, and article totals cannot establish the frequency of a headline term without

a defined relationship between collections. This revision therefore avoids new proportions based on unresolved units.

For a future verified reanalysis, token frequency could be normalised per 1,000 processed tokens, using the total number of tokens in the relevant collection or time interval. Document prevalence could instead use unique eligible documents as its denominator. These are distinct statistics answering distinct questions. No such measure has been calculated for the missing corpus. A term's percentage among a selected list of frequent terms is also different from its proportion of all tokens or all documents; figures must name the denominator rather than rely on the generic phrase total dataset.

Collocation is equally dependent on specification. An adjacent ordered bigram, a symmetric token window and same-document co-occurrence are not interchangeable. The source manuscript uses association, bigram and collocation strength without supplying the window size, direction, boundary handling, minimum frequency or association statistic. A high raw co-occurrence count can simply reflect high marginal frequency. Neither a network edge nor a line chart establishes statistical association strength unless the metric is defined. Directional differences are recorded as unresolved rather than repaired through an assumed software default.

Temporal interpretation and validation

The earlier account describes daily trend construction through Visual Basic for Applications (VBA), but also records a monthly caption and month-grouped labels. Date-level inputs and the original lexical time-series display are unavailable for independent inspection. The plotted units, retrieval gaps, duplicate updates and consistency of coverage therefore cannot be determined. Separately, descriptions of the unavailable Voyant exports identify document segments and non-chronological document ordering; these descriptions are insufficient to establish a calendar series.

Figure 3 replaces the proposed lexical time series with a plot of monthly collection volumes printed in the table. Those twelve observations per collection have a clear documentary source and a defined unit: reported records per month. They do not measure COVID prevalence, continuous article coverage or population exposure. No epidemic-event annotations are added because an annotation would invite a relationship the available inputs cannot test. A future temporal analysis would need aligned publication dates, coverage logs, term counts and denominators before independently verified events could be compared with news patterns.

The earlier analysis claims correlation analysis and visual comparison between VBA and Voyant outputs, but the available manuscript contains no aligned term-frequency vectors, coefficient, sample size or validation output. Cross-tool agreement remains unverified. Even a documented correlation would address agreement under specified processing choices, not moral panic or audience effects. A future check should report exact discrepancies alongside any correlation; it has not been performed here.

Contextual interpretation and research integrity

The earlier account refers to translated example headlines associated with six thematic categories, but provides no accessible URL-linked Indonesian passages or documented selection rule for their authentication. No such example is reproduced here as a verified quotation. The categories remain proposed interpretive labels rather than validated codes; no inter-coder reliability statistic, thematic prevalence or sentiment score is claimed.

The new contribution concerns inference boundaries: which observations support lexical description, which require contextual interpretation, and which require evidence from outside the text collection. Figure 1 differentiates reported procedures, unresolved steps and proposed work. Figure 4 distinguishes the resulting levels of inference. Their arrows organise analytical requirements and do not model estimated causal relationships. The ethical and data-access implications of further analysis are considered explicitly, recognising that vulnerability and identifiable information can matter in digital research even when material is accessible online (Massey et al., 2023).

Results

Corpus accounting and verification status

The twelve monthly headline values reproduced in Figure 3 sum to 22,531, exceeding the stated 22,491 by 40 records. The twelve article values sum to 649, which is 91 below the stated 740. The embedded figure and narrative agree on the monthly values, including January's 222 articles. Table 1 preserves both the stated totals and calculated sums. This independently repeated calculation verifies an internal discrepancy, not the underlying corpus size.

There are several possible sources of disagreement: a transcription mistake, use of different corpus versions, inclusion of unlisted records, removal of duplicates or confusion between retrieved and retained material. The record does not distinguish them. Selecting the monthly sum as the true corpus total would therefore be as unjustified as repeating the stated total without qualification. Both remain visible in Table 1. Unique-record counts and reconciliation logs are needed before any population denominator is adopted.

The discrepancy matters differently for different claims. A printed frequency rank may remain unchanged after reconciliation, but its relative frequency and scope can change. A claim that the corpus covers the complete annual headline population requires evidence of retrieval completeness independently of its total. A claim that the article sample validates headline interpretations requires a traceable connection between the two collections. Neither requirement is met merely because a large number appears in a Methods section. The two totals consequently cannot be combined into one sample size.

Monthly collection volumes are themselves informative about the documented sampling conditions. The printed headline counts range from 1,731 in February to 2,027 in December. The article entries range from

20 in June to 222 in January, with much smaller values in most other months. These are ranges of reported records, not uncertainty estimates. January's article value merits checking against the page limit, records per page and retrieval log, but its magnitude alone does not justify changing it to another number. Figure 3 makes that asymmetry visible without converting it into a claim about pandemic intensity.

Lexical prominence within the available outputs

The earlier account transcribes eleven English labels from a percentage display: COVID, cases, health, vaccine, PPKM, taskforce, vaccination, home, pandemic, viral and emergency. It also reports that the original caption described ten terms. The eleven documented percentages are 33.54, 13.74, 8.54, 8.45, 7.86, 6.73, 5.54, 4.10, 4.04, 3.76 and 3.65; they sum to 99.95%. The sum and number of transcribed entries can be checked, but the original percentage display and its denominator are not available here. Rounding over selected terms is one possible explanation of the near-total, not a verified measurement rule.

The documented percentages cannot establish that approximately one-third of all headlines concerned COVID or that one-third of all tokens were COVID. Those are different measures requiring explicit denominators. The COVID label of 3,160 cannot be divided by either disputed headline total as document prevalence without establishing that it counts unique eligible headlines. No percentage chart is reproduced because its denominator and units remain unresolved.

Figure 2 contains the labels COVID = 3,160, cases = 1,298 and health = 805; these values agree with the accompanying manuscript narrative. Their descending order is directly observable in this retained figure. Their counting unit and derivation cannot be authenticated against the source corpus. They are therefore reported count labels, not verified token frequencies or document counts. The earlier account's claim that they also agree with an original collocation display cannot be independently checked because that display is absent.

Figure 2. Reported count labels for three central terms. Bars reproduce COVID = 3,160, cases = 1,298 and health = 805 from the earlier manuscript. The horizontal axis represents reported labels with an unresolved counting unit; bars are not independently verified token frequencies, document counts or statistical estimates. No error bars are applicable to this documentary transcription.

At this descriptive level, the documented labels identify infection tracking, health, preventive action and institutional response as subjects in the earlier analysis. Vaccine, vaccination, PPKM and taskforce remain labels reported in that account, not a newly verified distribution of article frames. Without authenticated passages, a codebook and counted coding units, the record cannot establish the prevalence of moral or emotional positions, or the stance adopted towards these subjects.

Co-occurrence patterns and measurement uncertainty

The earlier narrative records COVID→cases = 1,298 and cases→COVID = 987, together with COVID→health = 805 and health→COVID = 264. These are documented directional values, not newly verified co-occurrence measurements. The unequal reverse-direction values preclude treating them as symmetric pair counts without clarification. Ordered windows, different queries, aggregation rules, labelling inconsistency or transcription could account for the differences; none is established. Averaging the values or selecting one as the true association strength would introduce an unsupported correction.

The earlier account also describes an original collocation chart joining separate lexical groups with a continuous line. The original chart is not present for independent inspection. Even if its description is accurate, a list of words on a horizontal axis is not time, and a high point cannot identify a news-cycle surge. The documented relationships therefore remain measurements with unresolved units rather than evidence of a chronology or escalating panic.

The association of cases with highest, reach, spikes and increase is consistent with numerical reporting of infection changes. It could also occur in emphatic or alarmist language. Distinguishing those uses requires the surrounding sentence, information source and comparison being made. Highest may report an accurate record; increase may describe a verified trend; reach may simply connect a count to its numerical value. None independently identifies a threatening social group, moral condemnation or exaggerated assessment of harm. The same limitation applies to positive when it denotes a test result.

The association of health with workers, protocols, vaccine and vaccination is consistent with protective-information reporting. It could also accompany criticism of policy, controversy over mandates or concerns about resources. Because passages are unavailable, the revision does not assign reassuring or anxiety-producing effects to these associations. Table 3 identifies the inferential claims and the narrower statements that remain defensible. The result is an account of evidence insufficiency, not evidence that the coverage was necessarily neutral or beneficial.

Table 3. Evidence and claim audit

Claim	Available evidence	Defensible revised claim	Further evidence needed
COVID dominates all news	Earlier narrative transcribes percentages; retained Figure 2 contains three reported labels	COVID has the highest of three audited reported count labels	Complete eligible population; defined counts and denominator
Words induce fear	Frequent disease and health terms	Pandemic-related vocabulary is prominent in reported outputs	Contextual evaluation plus exposure and audience measures
Coverage creates moral panic	Interpretive prose; frequency and association displays	Moral panic is not demonstrated by available measurements	Actor attribution; hostility; moral judgement; control demands
July peak follows pandemic events	Earlier narrative describes an unavailable temporal chart	A July peak is described in the source; causal association untested	Dated counts; coverage logs; verified event series; justified model
Jokowi is a non-crisis control	Different-topic charts from different years	Cross-period comparison cannot test moral panic	Comparable frame, exposure period and common denominators
High validation and reliability	Claim of correlation; absent outputs	Cross-tool agreement remains unverified	Inputs; coefficient; n; code; discrepancy assessment
Commercial or algorithmic amplification	No relevant measurements	Mechanisms cannot be inferred from frequency	Production, platform, ranking or engagement evidence

Claims are paraphrased. The revised claims concern the supplied record, not independently reproduced corpus findings.

Temporal displays and limits of chronological inference

The earlier account describes a July peak in an annual COVID lexical chart. That original chart and its date-level inputs are unavailable here, so the peak remains an unverified description. No exact daily values, date assignments or interval denominators can be authenticated. First-Monday article sampling and the reported year-long headline collection also have different temporal coverage. No correlation with infections, restrictions or vaccination events is documented.

Figure 3 shows only monthly record volumes documented in the manuscript. There are twelve reported observations for each collection. These are collection counts, not a lexical time series. December has the highest reported headline collection count, whereas January has the highest article entry; neither establishes the timing of COVID word prevalence. Determining changes in attention would require dated term numerators, consistent denominators and evidence of retrieval completeness.

Figure 3. Monthly reported collection volumes in 2021. Panel a reproduces twelve headline entries; panel b reproduces twelve article entries from the manuscript record. Monthly sums are 22,531 and 649, respectively, compared with stated totals of 22,491 and 740. Values describe reported records and do not measure COVID prevalence, daily coverage, exposure or epidemiological change. January's 222 article records remain unverified against retrieval logs. The figure is retained without altering the documented values.

The comparison between COVID in July 2021 and Jokowi in February 2024 is excluded from the main evidentiary argument. It compares a disease designation with a political actor's name in different years and reporting environments. The calendar months have different lengths, and corpus composition, retrieval completeness, term variants and normalisation are unspecified. A lower count for one topic does not constitute a control condition for the other, nor does political coverage establish the absence of a crisis. The comparison remains an issue documented in the audit, rather than a test supporting moral panic.

Evidence for moral panic remains undetermined

No authenticated corpus passages establish hostility towards an identified group, demands for punitive control or a consensual moral evaluation. The record contains interpretive statements that these processes occurred, but their evidence is not supplied in a form permitting examination. Audience fear, anxiety and behavioural consequences were not measured through surveys, experiments, interviews, exposure records or linked response data. The presence of relevant vocabulary therefore cannot be used to estimate these outcomes.

Disproportionality is similarly unresolved. The manuscript does not define an epidemiological benchmark, distinguish accurate emergency reporting from exaggeration, or compare reported claims with knowledge available at publication. A peak in news vocabulary does not reveal whether concern exceeded the actual threat. Figure 4 identifies these boundaries. Moral panic remains a theoretical possibility requiring contextual and external evidence, rather than a confirmed result. Table 4 summarises rival interpretations and the observations needed to discriminate among them.

Table 4. Rival interpretations

Observed reported pattern	Competing explanation	Available support	Counterevidence or limitation	Discriminating evidence
COVID and cases prominence	Routine emergency statistics	Case-related words and count labels	No source passages or surveillance alignment	Templates; dated case bulletins; verified headlines
PPKM and taskforce labels	Official announcements	Institutional vocabulary appears	Sources and voices not coded	Attribution to officials; announcement links; context
Vaccine and protocol terms	Public-service and efficacy reporting	Response vocabulary reported	Promotion, criticism and uncertainty indistinguishable	Indonesian extracts; advice and efficacy coding
Recurrent vocabulary	Repeated or syndicated material	Possible given archive collection; not demonstrated	Duplicate and update logs absent	Exact and near-duplicate audit; sensitivity checks
Directional pair differences	Processing or query artefacts	COVID–cases 1,298; cases–COVID 987	Window and direction unknown	Export settings; paired reruns; unit definitions
Increase and emergency terms	Alarmist or moralising framing	Potential candidate terms; no validated framing evidence	Accurate danger reporting remains plausible	Blame, exaggeration and negative-case coding
Temporal variability	Changing attention or coverage gaps	Source describes peaks; monthly volumes vary	Date-level counts and coverage unavailable	Aligned daily numerator, denominator and collection log

Competing explanations are hypotheses, not established mechanisms. No contextual passages or audience outcomes have been newly coded.

Verification status of separately described Voyant exports

The earlier account describes eight Voyant exports labelled Cirrus, Trends, StreamGraph and Links for News and Headlines. Those files are not embedded in the accessible manuscript and were not separately provided for this revision. References to Supplementary Figures S1–S8 have therefore been removed rather than imply that inspected supplements accompany the present document. The following observations preserve the documented descriptions while withholding independent verification.

According to that account, Headlines Cirrus foregrounds police, Jokowi, Indonesia and Jakarta alongside COVID-related vocabulary; News includes said, cases, health and publication-related terms. Headlines Trends is described as ten segments within one document whose label begins SINGLE1YEAR, while Headlines StreamGraph is described as approximately 50 positions. News distributional views reportedly display 18 document labels in non-chronological order. If confirmed, these properties would prevent direct

interpretation as calendar time and would raise questions about topic filtering and preprocessing. They do not establish duplicate records, numerical ranks, cross-corpus effect sizes or the extent of pandemic coverage.

The reported Links descriptions include COVID connections to vaccination, cases or patients and a separate police neighbourhood containing arrest and chief. The account also describes links involving said, monday and health. Without the original exports, weighted edge data, context settings and source passages, these descriptions cannot authenticate a semantic network or demonstrate pandemic enforcement, punitive framing or statistically adjusted association.

Recovery of the eight exports is necessary before their labels, axes, term selections and connections can be independently checked. Recovery alone would not reproduce their underlying measurements: corpus manifests, language mappings, exported values and analytical settings are also needed. No additional token count, sentiment score, temporal correlation or cross-tool validation is inferred from the earlier descriptions.

Discussion

From lexical salience to contextual interpretation

The reported outputs are most usefully interpreted as a provisional account of how pandemic-related subjects became visible in a news collection. Disease identification, case enumeration, health provision, vaccination and restrictions represent different domains of practical knowledge. Their co-presence suggests a vocabulary linking the emergency to institutional action and everyday conditions. This interpretation is deliberately limited: it concerns what the reported labels make available for further investigation, rather than what audiences thought or what editors intended.

The Indonesian case matters because institutional terms can identify locally specific forms of governance within a globally shared emergency. Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), the Indonesian designation for restrictions on community activities, cannot be replaced analytically by a generic English term such as lockdown without losing distinctions in policy designation. Similarly, taskforce may refer to a formal institution, a local coordinating body or translated terminology covering several expressions. Recovering Indonesian forms would allow the study to examine how administrative authority was named, which sources were quoted and whether responsibilities were attributed to institutions or individuals. At present, these are substantive questions opened by the lexical record, not findings about authority or compliance.

Research on governmental response in Indonesia offers a contextual basis for asking such questions. Mahendradhata et al. (2022) examine health academics' assessments using a modified Delphi design, while Rosemary et al. (2022) examine efficacy information in media coverage. Neither study supplies missing evidence about the present corpus, but both demonstrate that institutional performance and efficacy are researchable through defined units and procedures. Tri Sakti et al. (2021) examine public

sentiment and emotion through online opinion material, underscoring that audience expression requires a different dataset from publishers' headlines.

The interpretive distinction is between reporting that names a problem and reporting that attributes culpability. A headline about cases reaching a numerical level can organise knowledge of the emergency without identifying a deviant actor. A headline blaming a community for transmission would perform a different social action, especially if it generalised responsibility or advocated punishment. Establishing either action requires an authenticated passage and attention to whose voice is represented. The present reassessment identifies this distinction without claiming to have coded its distribution in the corpus.

Rival interpretations of repeated emergency vocabulary

Routine reporting of an emergency is a credible explanation for the prominence of COVID, cases and health. Recurring statistical updates can reproduce a small vocabulary because the information being conveyed is repetitive. Official announcements can make administrative terms frequent. Public-service journalism can generate vaccine and protocol vocabulary. These mechanisms can occur without sensationalism and can coexist with isolated moralising stories. The presence of a credible rival explanation means that frequency alone does not discriminate the interpretation favoured by the study.

Studies of journalism in health emergencies show why communication cannot be presumed to function only as alarm. Winters et al. (2020) investigate the roles of Sierra Leonean journalists during Ebola, placing journalistic practice within an emergency response. Sleigh et al. (2021) examine visual risk communication used to promote prevention messages. These studies involve different countries, periods and media forms, but they establish the analytical relevance of practical information and prevention. They should inform the questions asked of the Kompas.com texts, rather than be used to assert an identical effect.

Repeated content and processing artefacts are additional possibilities. An archive can contain updates, reused headlines, syndicated items or recurring statistical templates. Exact duplicates and near-duplicates have different implications: the former can inflate a count through retrieval repetition, while the latter can represent genuine repeated publication. Combining files can alter document boundaries, and translated category labels can merge distinct source expressions. These mechanisms require technical examination alongside contextual reading. Removing duplicates without a documented rule could itself erase meaningful repetition, so sensitivity analyses should retain both record-level and deduplicated descriptions where justified.

The evidence cannot rule out alarmist or moralising framing. Rather, it locates the observations that would support that explanation: exaggerated factual claims, essentialising group descriptions, moral condemnation, demands for punishment or dismissal of disconfirming evidence. Negative cases would include accurate numerical reports, practical explainers, corrections, balanced accounts of uncertainty and stories resisting stigma. A contextual analysis capable of distinguishing them would be stronger than assigning a panic interpretation to every term associated with illness or public-health control.

Evidence thresholds for moral panic and audience consequences

Media exposure can be associated with psychological outcomes, but the relationship is heterogeneous and cannot be read directly from publication frequency. Lee et al. (2020) examine media usage, infection knowledge and anxiety through structural analysis. Holman et al. (2020) use a probability-based sample to investigate mental health during the unfolding pandemic. Jones et al. (2021) discuss emotional responses to news and social media information. These sources support the importance of studying audience experience while showing that it needs observations of people, exposure or expression.

The distinction is especially consequential when public concern motivates protective action. Van Bavel et al. (2020) discuss social and behavioural science contributions to pandemic response, and Chu et al. (2022) report differentiated exposure associations with disease concern and preventive measures. Thus, worry cannot simply be equated with irrational panic, and a story that makes risk salient may inform behaviour as well as distress. No direction or size of either outcome is estimated for Kompas.com in this revision.

Information production and information demand also need separate measurement. Gravino et al. (2022) examine both news supply and demand, illustrating that the amount published is not the same as public attention. A headline archive records what was retained from a publisher's output under a collection procedure. It does not show which items were encountered, read, believed or shared. Inferring public fear from that archive would aggregate several unobserved steps into a single claim. Those steps should instead become explicit research questions.

Vaccination and misinformation are distinct constructs

Vaccination vocabulary demonstrates the presence of an important subject, not the position taken towards it. Larson (2022) addresses the problem of defining and measuring vaccine hesitancy, and Nuwarda et al. (2022) discuss its varied influences. Lazarus et al. (2021) investigate potential acceptance through a survey rather than infer it from media vocabulary. These approaches clarify why frequent vaccine references cannot establish reassurance, fear, distrust or hesitancy without the relevant unit of observation. The present outputs leave these positions undetermined.

Misinformation also requires identification of false or misleading claims, not merely the occurrence of terms such as hoax or claims. Sell et al. (2020) analyse message veracity and content during the Ebola communication crisis. Eysenbach (2020) distinguishes information monitoring, literacy, quality refinement and knowledge translation in infodemic management. The frequency of misinformation-related vocabulary in a news collection could indicate false information, reporting about it or attempts to correct it. These are substantively different communicative acts.

Experiments and surveys further illustrate the measurement gap. Pennycook et al. (2020) test an accuracy intervention, Roozenbeek et al. (2020) investigate susceptibility to misinformation across

national samples, and Loomba et al. (2021) examine effects on vaccination intent. Ngai et al. (2022) analyse vaccine misinformation themes and writing strategies in relation to social-media virality. Such findings cannot be transferred to the present outlet without matching evidence. They support specifying truth status, message form, exposure and response before making claims about informational harm.

What would make the contextual analysis defensible

Further work should begin with corpus reconciliation rather than additional visualisation. The minimum record should identify each retrieved item, its source URL, publication and retrieval times, collection membership, eligibility decision and duplicate status. A processing log should connect raw text to retained tokens and analysis exports. This would establish whether the reported collections contain all indexed records, a pandemic-specific subset or another selection. Until then, even an apparently clear frequency rank remains conditional on an unresolved sampling frame.

Contextual interpretation should preserve Indonesian text and attach each extract to an identifiable article. Selection could combine a transparent spread across months and term families with deliberately examined candidate moralising and negative cases. A purposive diagnostic sample would test interpretations, while a probability sample from a verified frame could support estimates of prevalence. The two purposes should not be confused. A sample chosen because passages appear alarming cannot establish how common alarmist reporting was across the collection.

Coding should distinguish threat information, efficacy information, actor attribution, moral judgement, stigmatising generalisation, punitive demands and explicit uncertainty. It should identify whose voice is represented: a journalist, an official, a quoted individual or a rebutted claim. Translations should preserve evaluative force and negation, with ambiguities recorded. If several coders participate, training, independent assessment and disagreements should be reported; no reliability statistic should be invented for a single analyst or for retrospective agreement without independent coding.

The broader analytical sequence should follow the principle of returning computational patterns to contextual evidence, rather than replacing interpretation with a more sophisticated model (Nelson, 2020). Automated sentiment or topic analysis might assist exploration, but it would need its own validation and could still fail to measure moral attribution or disproportionality. The suitability of a method depends on the question it answers (Grimmer et al., 2021). More complex software is not, by itself, a solution to a mismatch between a lexical measure and a sociological claim.

Assessing proportionality without denying material harm

A defensible assessment of disproportionate concern would first specify what is being compared. The volume of news, the factual severity of claims, the level of public worry and the restrictiveness of policy are different outcomes. A claim that a numerical report exaggerates infections would require comparison with the corresponding surveillance information and its limitations. A claim that audiences overestimate personal risk would require measurements of perceived risk and an appropriate benchmark. A claim that

punitive responses exceed the threat posed by a group would require evidence about that group's attributed and documented conduct. The present lexical outputs provide none of these comparisons.

Timing would also matter. A later estimate of disease severity is not automatically the correct standard for judging reporting published under earlier uncertainty. Journalists may have relied on provisional information, changing testing practices or incomplete official records. An analysis should compare a story with credible information available when it appeared and record what was uncertain. This does not absolve misleading reporting; it makes the standard for identifying it explicit. An urgent report may be accurate, an understated report may omit important danger, and an alarming claim may be unsupported. These positions require examination of claims rather than an emotional judgement about a word.

Moral evaluation can nevertheless occur during a real emergency. Evidence that the virus posed substantial harm would not preclude evidence that some communities were scapegoated. Conversely, documentation of scapegoating would not establish that all concern about the pandemic was excessive. These are compatible observations at different levels. A future analysis should identify the specific actor, attributed responsibility, evaluative language and response demanded, then assess whether evidence supported that attribution. This approach avoids treating a whole year of heterogeneous coverage as one undifferentiated moral-panic episode.

The same discipline applies to consensus and volatility. Agreement among editorials, agreement among quoted officials and agreement among audiences represent different populations. A short spike in an archive's word counts establishes neither shared moral judgement nor a volatile public response. Repeated publication by one outlet could reflect a recurring beat, while a diverse set of stories could disagree about policy despite using similar vocabulary. For the present study, these dimensions remain questions requiring defined units and additional material, rather than latent properties assumed to accompany lexical salience.

These distinctions provide a practical stopping rule for interpretation. When the available record establishes only frequency, the conclusion should concern frequency. When authenticated passages establish moral attribution, the conclusion can concern that attribution within the examined material. When a credible comparison and audience evidence are also available, stronger social claims may become warranted. Progress through these levels is conditional, not inevitable. Figure 4 represents the evidentiary thresholds precisely so that the reader can see where this study currently stops and what would be needed to move further.

Contribution within the study’s evidentiary scope

The contribution is a transparent reassessment of the relation between a documented lexical output and the claim it is used to support. Internally reproducible arithmetic exposes unresolved corpus accounting, while unit-specific qualifications prevent displayed labels from becoming estimates of social response. The case also specifies the additional evidence needed to distinguish routine emergency reporting, moral attribution and audience consequences. Figure 4 organises these requirements; it has not been validated

as a measurement instrument, causal model or general theory. Its usefulness beyond this case remains to be evaluated using authenticated corpora and independent contextual or audience observations.

Limitations and implications for interpretation

Corpus integrity and research identity constrain the central findings. This reassessment checks the documented analytical record; it does not independently reproduce the original news analysis. Raw and cleaned records, retrieval logs and executable processing materials were unavailable, and eight separately described Voyant exports could not be inspected. The conflicting headline and article totals prevent confirmation of corpus size, uniqueness and completeness. Internal arithmetic and the labels in the four embedded figures remain checkable; corpus-level frequencies do not. Recovering versioned source records, URL-based identifiers and a retrieval-to-retention reconciliation is necessary to restore that inference. These are central verification gaps rather than minor caveats that can be solved by qualified wording.

Sampling and coverage restrict annual and cross-collection inference. First-Monday sampling and the first-five-page rule cover selected occasions and positions; archive completeness, page ordering and eligibility criteria are undocumented. January's 222 article entries require checking against actual page capacity and retrieval logs. The reported headline and article collections are not established as matched pairs. Observed differences therefore cannot isolate editorial presentation from topic selection, length or sampling coverage. The reported procedures can be described, but neither annual representativeness nor the direction or magnitude of selection bias can be estimated. An explicit sampling frame, inclusion log, coverage audit and verified headline–article linkage would address these limitations.

Language, preprocessing and measurement limit lexical interpretation. Indonesian source texts are described alongside English outputs, but translation timing and a form-to-form mapping are absent. Tokenisation, case and COVID-19 normalisation, stopwords, duplicate rules and recurring publication metadata remain undocumented. Translation may merge or separate expressions, but its actual effect is unknown. Unresolved count units, percentage denominators, co-occurrence windows, directionality and visual settings prevent comparable frequency or association estimates. The defensible result is the ordering of reported labels, not prevalence or association strength. Recovering originals, transformation logs, complete exports and settings would allow sensitivity checks and unit-specific reanalysis.

Temporal validity is limited by absent date-level records and uninspected distributional exports. Segments within a document and non-chronologically ordered documents, as described in the earlier account, are not verified calendar intervals. Monthly collection totals remain reportable as documentary entries, but they cannot establish changes in COVID attention or relationships with infection waves, vaccination or restrictions. Publication-date mapping, consistent interval denominators and retrieval-gap checks are required before any temporal comparison. No direction or size of a temporal relationship is estimated here.

Construct validity, causal inference and transferability remain limited. Lexical repetition does not discriminate routine emergency reporting from blame, hostility, consensus, disproportionality or

moralising narratives. Authenticated Indonesian passages, source attribution, contextual coding and negative cases would be needed to examine these constructs. No audience exposure, anxiety, behaviour, editorial intention, commercial strategy or algorithmic mechanism was measured, so neither harmful nor beneficial effects are established. Audience or production evidence and an appropriate design would be needed for those claims. Failure to demonstrate moral panic does not establish its absence. Finally, one outlet in one pandemic year cannot represent Indonesian journalism generally; comparison across verified outlets, periods and crisis settings would be needed to assess transferability.

Conclusion

The accessible analytical record supports a restricted conclusion: the earlier manuscript reports pandemic, prevention and institutional vocabulary, and its embedded count-label figure ranks COVID above cases and health. These are documentary observations rather than independently reproduced corpus findings. The available evidence does not establish moral panic, audience fear or disproportionate concern; it also does not establish that such phenomena were absent or that reporting was neutral or beneficial.

Recalculation confirms two internal accounting discrepancies: monthly headline entries total 22,531 against a stated 22,491, while article entries total 649 against a stated 740. Neither sum is adopted as the corrected corpus size. Count units, percentage denominators, translation and directional association settings remain unresolved. Descriptions of unavailable Voyant exports cannot establish calendar trends, and the July 2021 COVID–February 2024 Jokowi comparison supplies no valid control.

The case clarifies the evidence needed to connect lexical salience with sociological interpretation. Authenticated passages are required for moral attribution; credible contemporary benchmarks and appropriate audience evidence are required for stronger social claims. The proposed evidence requirements are a conceptual contribution, not a validated model. Recovery and reconciliation of the original research materials remain necessary before the study can substantiate corpus-level results or proceed on the basis of a verified original corpus analysis.

Statements and Declarations

Funding:

No funding was received for conducting this study.

Competing interests:

The authors have no competing interest to declare that are relevant to the consent of this article.

Ethics approval:

Not applicable.

Consent to participate:
Not applicable.

Consent for publication:

The authors have read and given their consent for publication.

Data, materials and code availability:
Supplementary.

Generative AI use:

Generative AI assistance was used for substantive drafting, restructuring, methodological critique and code-assisted display preparation in this revision.

References

1. Baden C, Pipal C, Schoonvelde M et al (2022) Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. Communication Methods Measures 16(1):1–18. https://doi.org/10.1080/19312458.2021.2015574
2. Capurro G, Jardine CG, Tustin J et al (2022) Moral panic about covidiots in Canadian newspaper coverage of COVID-19. PLoS ONE 17(1):e0261942. https://doi.org/10.1371/journal.pone.0261942
3. Chu TH, Yeo TED, Su Y (2022) Effects of Exposure to COVID-19 News and Information: A Meta-Analysis of Media Use and Uncertainty-Related Responses During the Pandemic. Journalism Mass Communication Q 99(1):89–112. https://doi.org/10.1177/10776990211068857
4. Cinelli M, Quattrociocchi W, Galeazzi A et al (2020) The COVID-19 social media infodemic. Sci Rep 10(1):16598. https://doi.org/10.1038/s41598-020-73510-5
5. Cohen S (1972) Folk devils and moral panics: The creation of the Mods and Rockers. MacGibbon and Kee, London
6. Cottingham MD (2024) Viral Fear or Panic Myth? Emotions in Ebola News and Social Media Responses. Symbolic Interact 47(3):389–409. https://doi.org/10.1002/symb.681
7. Cârstocea A (2022) Going Viral: The Moral Panic Constructing the Roma as a Threat to Public Health During the First Wave of the Covid-19 Pandemic. J Ethnopolitics Minor Issues Europe 21(2):57–80. https://doi.org/10.53779/niqa9963
8. Eysenbach G (2020) How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 22(6):e21820. https://doi.org/10.2196/21820
9. Gallotti R, Valle F, Castaldo N et al (2020) Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 4(12):1285–1293. https://doi.org/10.1038/s41562-020-00994-6
10. Garfin DR, Silver RC, Holman EA (2020) The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 39(5):355–357. https://doi.org/10.1037/hea0000875

11. Goode E, Ben-Yehuda N (2009) Moral panics: The social construction of deviance, 2nd edn. Wiley-Blackwell. https://doi.org/10.1002/9781444307924
12. Gravino P, Prevedello G, Galletti M et al (2022) The supply and demand of news during COVID-19 and assessment of questionable sources production. Nat Hum Behav 6(8):1069–1078. https://doi.org/10.1038/s41562-022-01353-3
13. Grimmer J, Roberts ME, Stewart BM (2021) Machine Learning for Social Science: An Agnostic Approach. Annu Rev Polit Sci 24(1):395–419. https://doi.org/10.1146/annurev-polisci-053119-015921
14. Hier SP (2021) A Moral Panic in Reverse? Implicatory Denial and COVID-19 Pre-Crisis Risk Communication in Canada. Can J Communication 46(3):505–521. https://doi.org/10.22230/cjc.2021v46n3a3981
15. Holman EA, Thompson RR, Garfin DR et al (2020) The unfolding COVID-19 pandemic: A probability-based, nationally representative study of mental health in the United States. Sci Adv 6(42):eabd5390. https://doi.org/10.1126/sciadv.abd5390
16. Jones R, Mougouei D, Evans SL (2021) Understanding the emotional response to COVID-19 information in news and social media: A mental health perspective. Hum Behav Emerg Technol 3(5):832–842. https://doi.org/10.1002/hbe2.304
17. Kania U (2022) Snake flu, killer bug, and Chinese virus: A corpus-assisted critical discourse analysis of lexical choices in early UK press coverage of the COVID-19 pandemic. Front Artif Intell 5:970972. https://doi.org/10.3389/frai.2022.970972
18. Larson HJ (2022) Defining and measuring vaccine hesitancy. Nat Hum Behav 6(12):1609–1610. https://doi.org/10.1038/s41562-022-01484-7
19. Lazarus JV, Ratzan SC, Palayew A et al (2021) A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 27(2):225–228. https://doi.org/10.1038/s41591-020-1124-9
20. Lazarus JV, Wyka K, White TM et al (2022) Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. Nat Commun 13:3801. https://doi.org/10.1038/s41467-022-31441-x
21. Lee YR, Lee JY, Park IH et al (2020) The Relationships among Media Usage regarding COVID-19, Knowledge about Infection, and Anxiety: Structural Model Analysis. J Korean Med Sci 35(48):e426. https://doi.org/10.3346/jkms.2020.35.e426
22. Loomba S, de Figueiredo A, Piatek SJ et al (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 5(3):337–348. https://doi.org/10.1038/s41562-021-01056-1
23. Mahendradhata Y, Lestari T, Djalante R (2022) Strengthening government's response to COVID-19 in Indonesia: A modified Delphi study of medical and health academics. PLoS ONE 17(9):e0275153. https://doi.org/10.1371/journal.pone.0275153
24. Malecki KMC, Keating JA, Safdar N (2021) Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. Clin Infect Dis 72(4):697–702. https://doi.org/10.1093/cid/ciaa758

25. Massey PM, Murray RM, Chiang SC et al (2023) Social Media, Public Health Research, and Vulnerability: Considerations to Advance Ethical Guidelines and Strengthen Future Research. JMIR Public Health Surveillance 9:e49881. https://doi.org/10.2196/49881
26. Ministry of Health, NITAG, UNICEF and WHO (2020) COVID-19 Vaccine Acceptance Survey in Indonesia. November 2020. https://www.unicef.org/indonesia/coronavirus/reports/covid-19-vaccine-acceptance-survey-indonesia
27. Nelson LK (2020) Computational Grounded Theory: A Methodological Framework. Sociol Methods Res 49(1):3–42. https://doi.org/10.1177/0049124117729703
28. Ngai CSB, Singh RG, Yao L (2022) Impact of COVID-19 Vaccine Misinformation on Social Media Virality: Content Analysis of Message Themes and Writing Strategies. J Med Internet Res 24(7):e37806. https://doi.org/10.2196/37806
29. Nuwarda RF, Ramzan I, Weekes L et al (2022) Vaccine Hesitancy: Contemporary Issues and Historical Background. Vaccines 10(10):1595. https://doi.org/10.3390/vaccines10101595
30. Pennycook G, McPhetres J, Zhang Y et al (2020) Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychol Sci 31(7):770–780. https://doi.org/10.1177/0956797620939054
31. Pitol MNS, Patwary MM, Aurnob S et al (2023) Exploring media consumption and mental health among young adults during the second wave of COVID-19 in Bangladesh. Heliyon 9(10):e20371. https://doi.org/10.1016/j.heliyon.2023.e20371
32. Roozenbeek J, Schneider CR, Dryhurst S et al (2020) Susceptibility to misinformation about COVID-19 around the world. Royal Soc Open Sci 7(10):201199. https://doi.org/10.1098/rsos.201199
33. Rosemary R, Nur Rochimah TH, Susilawati N (2022) Efficacy information in Government's initial responses to Covid-19 pandemic: A content analysis of the media coverage in Indonesia. Int J Disaster Risk Reduct 77:103076. https://doi.org/10.1016/j.ijdrr.2022.103076
34. Sell TK, Hosangadi D, Trotochaud M (2020) Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 20(1):550. https://doi.org/10.1186/s12889-020-08697-3
35. Sleigh J, Amann J, Schneider M et al (2021) Qualitative analysis of visual risk communication on twitter during the Covid-19 pandemic. BMC Public Health 21(1):810. https://doi.org/10.1186/s12889-021-10851-4
36. Tri Sakti AM, Mohamad E, Azlan AA (2021) Mining of Opinions on COVID-19 Large-Scale Social Restrictions in Indonesia: Public Sentiment and Emotion Analysis on Online Media. J Med Internet Res 23(8):e28249. https://doi.org/10.2196/28249
37. Ungar S (2008) Global Bird Flu Communication: Hot Crisis and Media Reassurance. Sci Communication 29(4):472–497. https://doi.org/10.1177/1075547008316219
38. van Atteveldt W, van der Velden MACG, Boukes M (2021) The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning

Algorithms. Communication Methods Measures 15(2):121–140. https://doi.org/10.1080/19312458.2020.1869198

39. Van Bavel JJ, Baicker K, Boggio PS et al (2020) Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 4(5):460–471. https://doi.org/10.1038/s41562-020-0884-z
40. Voyant Tools (n d.) Trends and StreamGraph documentation. https://docs.voyant-tools.org/docs/Tools.Trends.html; https://voyant-tools.org/docs/Tools.StreamGraph.html. Accessed 15 October 2026
41. Walsh JP (2020) Social media and moral panics: Assessing the effects of technological change on societal reaction. Int J Cult Stud 23(6):840–859. https://doi.org/10.1177/1367877920912257
42. Washer P (2004) Representations of SARS in the British newspapers. Soc Sci Med 59(12):2561–2571. https://doi.org/10.1016/j.socscimed.2004.03.038
43. Winters M, Nordenstedt H, Mölsted Alvesson H (2020) Reporting in a health emergency: The roles of Sierra Leonean journalists during the 2014–2015 Ebola outbreak. PLoS Negl Trop Dis 14(5):e0008256. https://doi.org/10.1371/journal.pntd.0008256
44. World Health Organization (2021) Coronavirus Disease 2019 (COVID-19): Indonesia Situation Report 64, 21 July 2021. https://cdn.who.int/media/docs/default-source/searo/indonesia/covid19/external-situation-report-64_21-july-2021.pdf

Figures

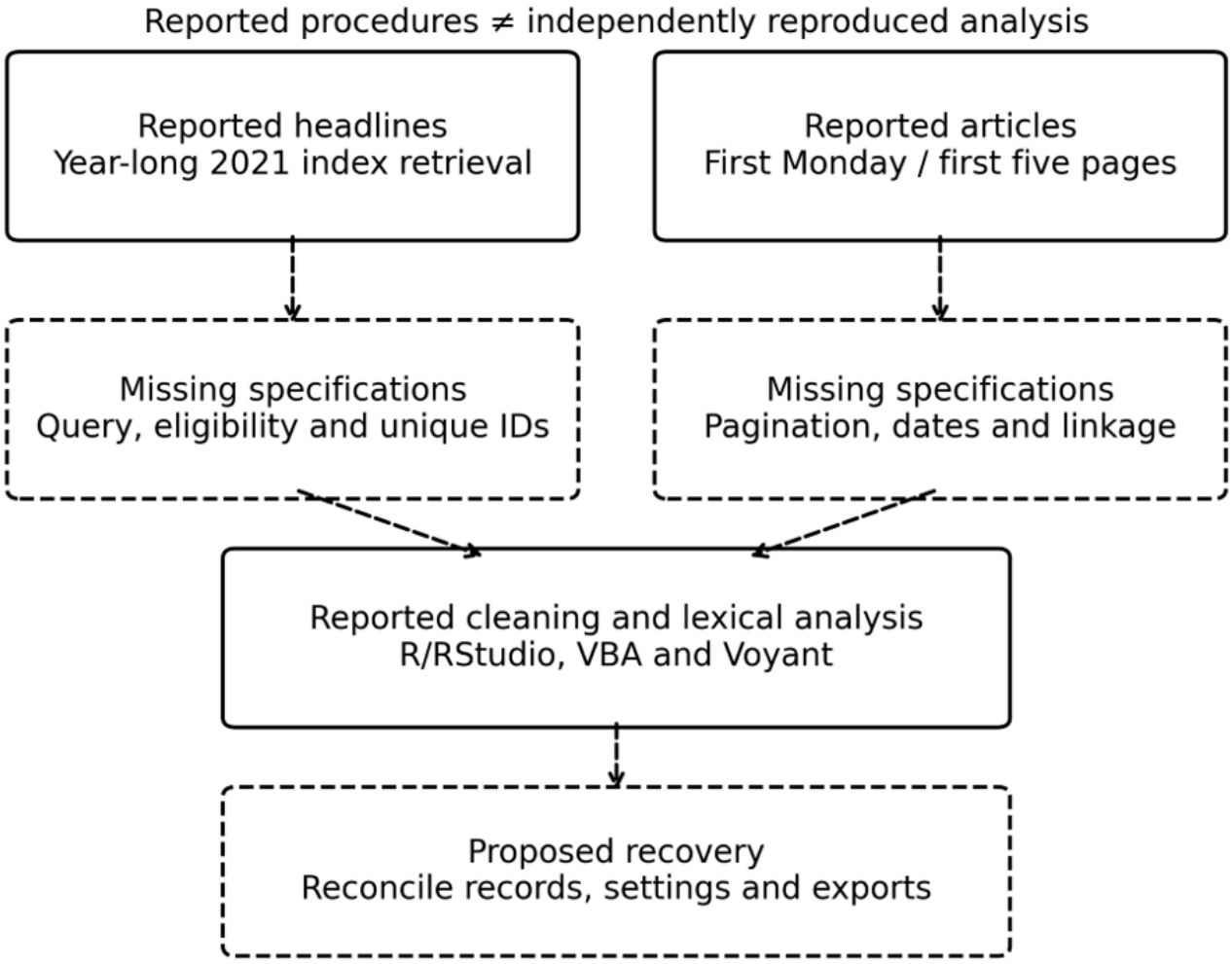


Figure 1

Reported corpus workflow and reproducibility gaps. Solid boxes summarise procedures described in the analytical record; dashed boxes identify missing specifications or proposed recovery work. Arrows organise documentation requirements and do not represent observed causal effects. Source: figure retained from the earlier manuscript; no source-corpus reanalysis.

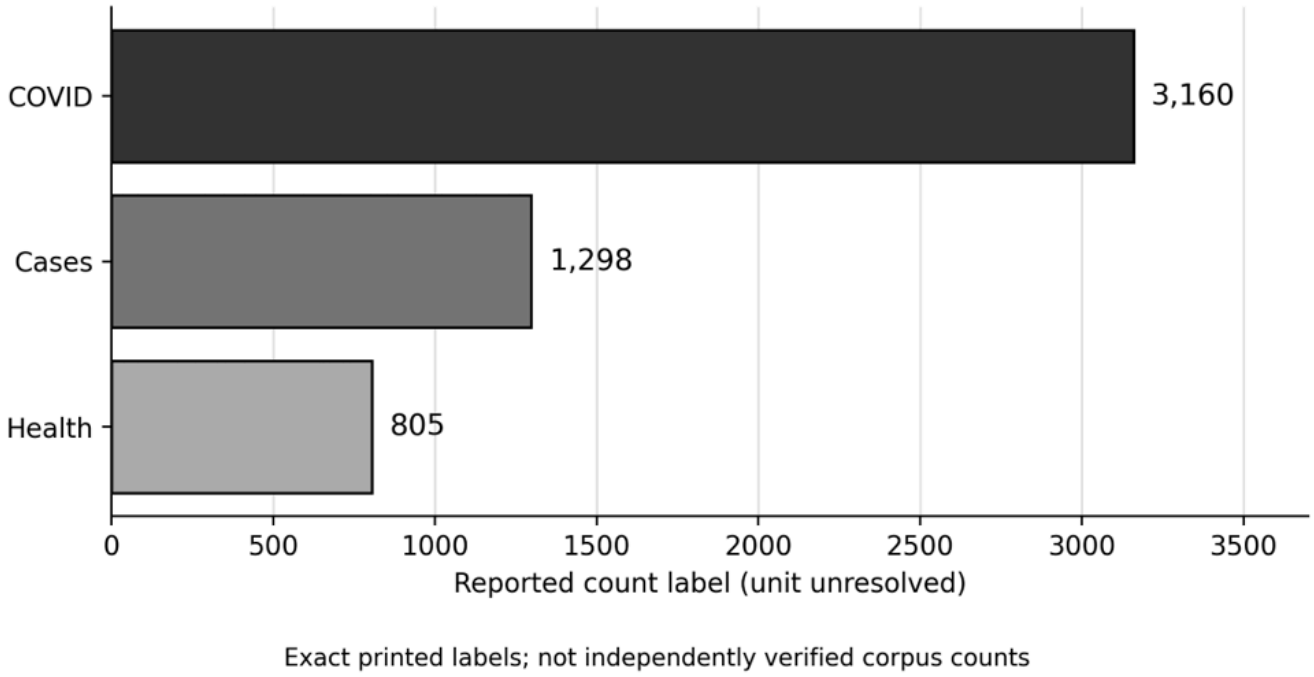


Figure 2

Reported count labels for three central terms. Bars reproduce COVID = 3,160, cases = 1,298 and health = 805 from the earlier manuscript. The horizontal axis represents reported labels with an unresolved counting unit; bars are not independently verified token frequencies, document counts or statistical estimates. No error bars are applicable to this documentary transcription.

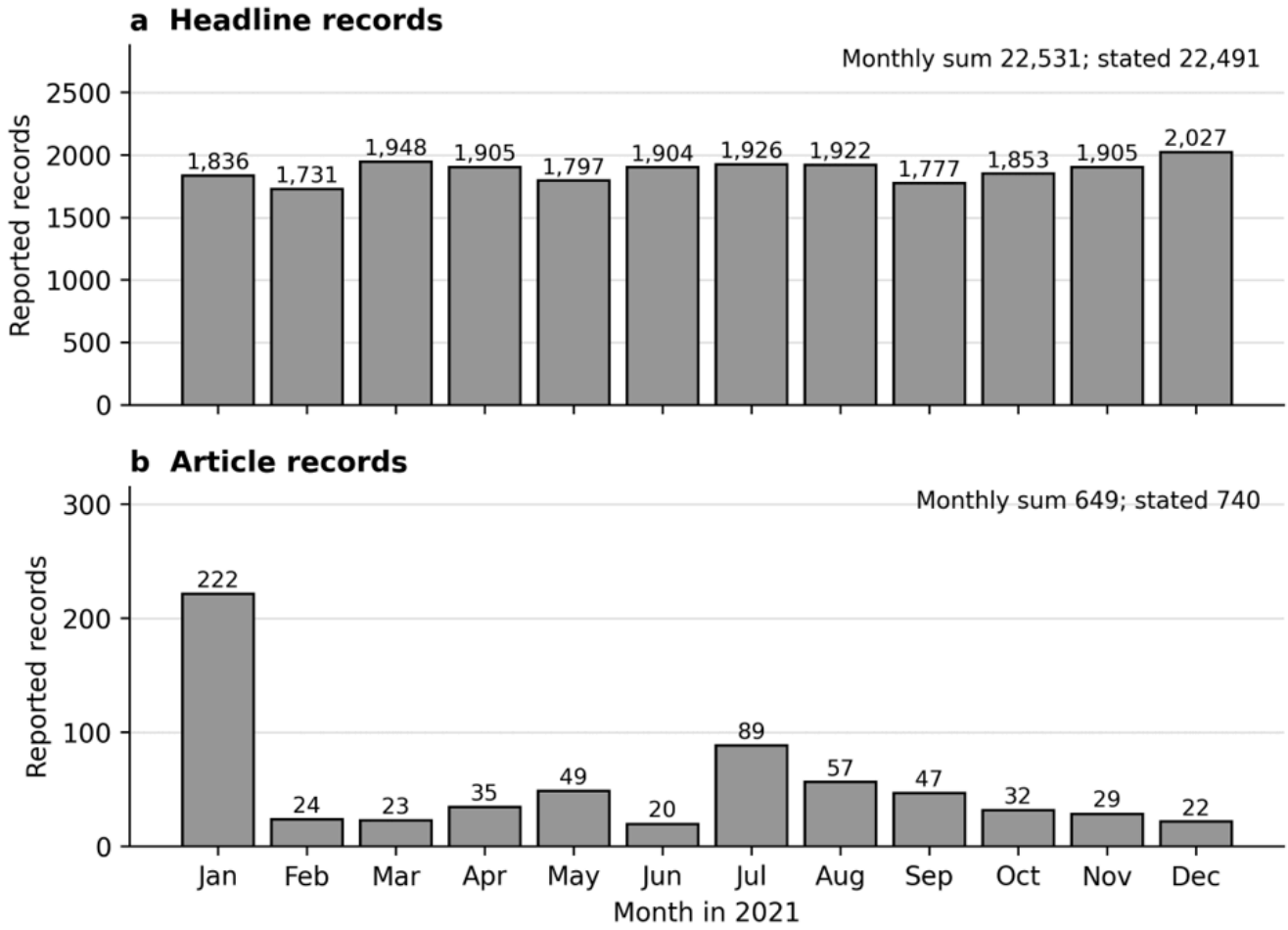


Figure 3

Monthly reported collection volumes in 2021. Panel a reproduces twelve headline entries; panel b reproduces twelve article entries from the manuscript record. Monthly sums are 22,531 and 649, respectively, compared with stated totals of 22,491 and 740. Values describe reported records and do not measure COVID prevalence, daily coverage, exposure or epidemiological change. January's 222 article records remain unverified against retrieval logs. The figure is retained without altering the documented values.

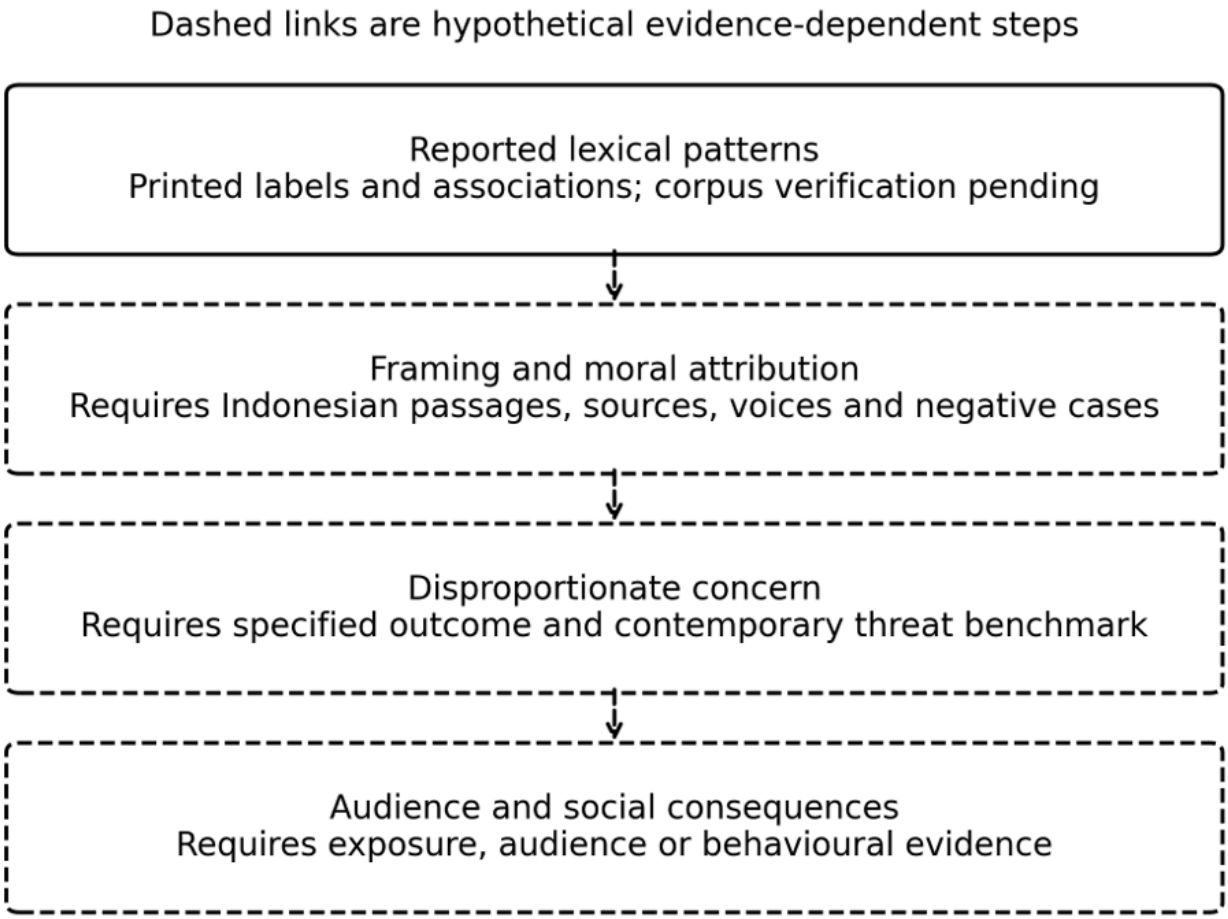


Figure 4

Evidence requirements from lexical description to social inference. Boxes distinguish reported lexical patterns, contextual moral attribution, disproportionate concern and audience consequences. Dashed arrows indicate possible analytical steps conditional on additional evidence, not estimated causal paths or a necessary sequence. This is a conceptual organising device, not an empirically validated model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureTable.docx
- RAWDATA.docx
- CombinedRScriptsandReferenceData.r
- NewsKompas2021CombinedChronological.docx
- HEADLINESKompas2021.xlsx