

In search of infodemics: US media amplification of risk

Cormac Bryce¹, Michael Dowling², Cheng Long³ and Jamie K. Wardman⁴

¹Bayes Business School (formerly Cass), City, University of London, UK

²DCU Business School, Dublin City University, Ireland, and Rennes School of Business, France

³Trinity Business School, Trinity College Dublin, Ireland

⁴Nottingham University Business School, University of Nottingham, UK

Abstract

Drawing on the Social Amplification of Risk Framework (SARF) we substantiate the occurrence of different forms and patterns of risk amplification within US news media. We apply keyword terms from the Society for Risk Analysis Glossary along with Latent Dirichlet Allocation topic modeling analysis to examine 271,854 news stories that contain explicit reference to risk over the last two decades. We find that risk news media coverage gravitates around seven core topics; health, business, environment, entertainment, domestic affairs, geopolitics, and technology. We also identify a series of discreet risk news events over this time period which exhibit an explosive growth in media coverage that can be distinguished as incidences of ‘risk amplification’, including key risks such as COVID-19 which demonstrate infodemic properties. We discuss the implications of these findings for policymakers and researchers regarding risk communication, infodemics, and risk perception.

Keywords: media risk communication; infodemics; risk perception; social amplification of risk; topic modelling.

1 Introduction

"We're not just fighting an epidemic; we're fighting an infodemic"

— Tedros Ghebreyesus, World Health Organization, 15 February 2020

"What exactly do I mean by the ‘infodemic’? A few facts, mixed with fear, speculation and rumor, amplified and relayed swiftly worldwide"

— David Rothkopf, The Washington Post, 11 May 2003

‘Once in a generation’ events such as the 9/11 attacks or the NASA Columbia disaster, have the propensity to define and reshape our experiential understandings, awareness, and appreciation of risk [1]. By capturing and focusing society’s attention, the mass circulation of risk-related information and images can place an indelible mark on how we perceive, interpret, and learn from such events [2, 3]. As a result, our heightened sensitivity to risk, and the need to avoid any associated harms, can subsequently become psychologically, socially, and politically internalized, providing an anchor that frames our response to future threats we encounter [4, 5, 6, 7].

The COVID-19 global health crisis may be considered as one such risk event [8, 1, 9]. The deluge of communication accompanying the pandemic has in turn been termed an *infodemic* by the World Health Organisation (WHO) to characterize the rapid surge and ‘over-abundance’ of both factual and false COVID-19 related information. It is feared that this may overwhelm people’s comprehension of the threat, along with their ability to ascertain trusted official health advice, if left ‘untreated’ [10, 11, 1, 12]. The COVID-19 pandemic is, thus, thought to represent a novel phenomenon that poses a number of significant risk communication challenges to health leaders trying to combat the adverse impacts of the coronavirus outbreak

[10, 13].

However, the conceptual underpinnings and applicability of the term *infodemic* to denote the rapid and widespread transmission of COVID-19 related information remain critically underdeveloped and unverified. While the infodemic metaphor was first coined almost two decades ago [14] to characterize the communication surrounding the emergence of severe acute respiratory syndrome (SARS) as a major public health concern, it has, until recently, simply not been widely used in risk research studies or figured as a major concern of risk managers. Nor, indeed, has the term been confirmed as distinctly unique from other characterisations of risk events. Yet, researchers and policymakers have largely accepted the term infodemic as a given, seemingly taking for granted that this problem diagnosis best fits the current communicative context of COVID-19 [15]. Meanwhile, questions abound concerning whether an infodemic has in fact occurred, how such a phenomenon might be understood to differ from more regular mass communication and reporting of risk events and crises, and the historic possibility of other infodemics. Consequently, without firmer conceptual and empirical validation, any corresponding policy treatments adopted in response face the possibility of being misjudged or misplaced, as well as being hard to justify and difficult to evaluate [16].

Addressing these issues, this paper aims to conceptually elaborate and substantiate a more refined understanding of the current infodemic. As indicated above, the COVID-19 pandemic is not the first major risk event to capture the world's attention and pose significant risk communication challenges. Nor is *infodemic* the only applicable concept for understanding and describing mass communication surrounding risk events. Instead, we propose that current articulations of the COVID-19 info-

demic share some key, yet unacknowledged, parallels with long-established conceptual processes and mechanisms outlined by the social amplification of risk framework (SARF) [17, 18]. We suggest that academic and policy discussions of the COVID-19 infodemic may accordingly benefit from capitalizing on the insights and understandings elaborated by SARF and subsequent research employing this framework. For instance, the original formulation of the infodemic concept [14] refers to the ‘amplification’ of facts, fear, and rumor. More recently, the WHO Information Network for Infodemics (EPI-WIN) has employed a series of ‘amplifiers’ to promote the spread of factual information from reliable sources to help fight the amplification of misinformation online and correct media reporting that ‘gets ahead of the evidence’ [12]. SARF studies have similarly historically been shaped by a shared interest in the societal communication of risk. They are often prompted by concerns about (dis-)proportionate risk reporting and behavior and continually aim to provide insights which will help improve official communication in response to new risk issues and events [17, 18]. Subsequently, we draw out and highlight these shared connections.

However, added to this, we also note that SARF and infodemic studies are both criticized for failing to set out plain criteria and methods by which to identify whether an instance of mass communication has arisen demonstrably in contradistinction to that which might otherwise be expected to fall within the normal parameters of everyday risk reporting. Addressing these issues, we employ a novel investigative approach that uses Latent Dirichlet Allocation topic modeling in conjunction with keyword search terms drawn from the Society for Risk Analysis (SRA) Risk Glossary. We apply this approach to US news coverage from an English-language media dataset of 271,854 news articles from 2000-2021. The approach provides an instruc-

tive method for uncovering the most prominent incidences of risk amplification, along with major risk topics and themes.

Our analysis yields a number of significant conceptual, methodological, and empirical contributions. First, we identify that US news media reporting of risk from the past two decades gravitates around seven main categories - namely: business, geopolitics, entertainment, domestic affairs, environment, technology, and health. We consider the concentration of US media reporting around these categories to reflect the main areas of everyday media attention and consumption through which risk is amplified.

Second, we are able to uniquely identify a series of major historic risk amplification events, their emergence from an underlying topic and prominence within these categories as they evolve. We further find that discussions relating to different risk categories involves diverse domain specific keywords, indicating that the lexicon of risk employed in reporting risk-related news is highly variable across contexts.

Finally, because of how the contours and dynamics of risk news reporting are structured, we are also able to further specify how media risk amplification can exhibit distinctive infodemic properties. These infodemic properties are reflected in terms of the speed and spread of risk information transmission surrounding certain events, and the wider systemic influence they exhibit within their own, and across other, risk news topics.

The remainder of the paper is structured as follows. In Section 2, we outline the major conceptual insights and research contributions of SARP, including recent developments regarding different forms of media attention and digitized communication in the ‘information age’. We use this overarching framework as a basis for

further contextualizing and conceptually elaborating understandings of infodemics vis-a-vis the mass communication and reporting of COVID-19. In Section 3, we introduce our methodology, along with justifications for our sampling and statistical approaches incorporating the SRA Risk Glossary. We also detail our dataset and descriptive analytical approach. Section 4 presents our study findings and analysis. Section 5 concludes our study by discussing these findings, along with significant implications, limitations, and possible future research directions.

2 Literature review

2.1 The social amplification of risk framework

Despite considerable progress in social scientific studies of risk, a systematic and comprehensive understanding of social risk experiences was yet to be realised until the late 1980s when the SARF was first developed to integrate this broad base of knowledge and expertise into an overarching framework [18]. Since its introduction, the SARF has attracted considerable research and policy interest following two major contributions. Firstly, it specifies a powerful metaphor with which to intuitively understand the social dynamics of risks and their societal impacts [19, 20]. Secondly, it outlines a comprehensive conceptual framework with which to instructively explore and examine many different mechanisms, components, and impacts contained within social responses to risk issues and crisis events [21, 22, 23].

In particular, the ‘social amplification and attenuation of risk’ metaphor characterises how individual, social, and, cultural processes interact in concert with the biological and physical properties of hazards in ways that can either intensify or

dampen the perception and communication of risks and their wider societal impacts [17, 24]. SARF then elaborates the respective components, mechanisms and communication pathways involved in this process, which it delineates into two main stages. In Stage 1, ‘risk amplification’ is typically understood to begin with a ‘signal event’, such as the emergence of a novel risk that characteristically embodies attributes that are of concern to observers [18]. If this initial risk event has strong ‘signal value’ it can gain social attention and intensify through interpersonal communication and the social exchange and sharing of corresponding messages and images that are taken up, filtered, and circulated by ‘amplification stations’, such as news media organisations which report about the risk in question [25].

In Stage 2, ‘ripple effects’ are observed when these risk signals resonate, grow, and reverberate throughout society. Specifically, as the circulation of risk signals becomes more widespread and permeates different areas of public discourse they impact on the awareness, understandings, and behaviours of actors, including the general population alongside government institutions and other organisations [17, 26]. These ripple effects may vary in form and magnitude, but in more extreme cases are commonly observed to result in significant secondary impacts, such as consumer boycotts and protests, the stigmatisation of persons, activities, or products, and the imposition of stricter regulations. This can lead to further risk avoidance and the instigation of protective measures, as well as produce economic losses - all of which have been observed in the current pandemic [9].

Researchers have applied the SARF quite extensively across a diverse range of risk related events and contexts, with studies addressing natural hazards, food-borne diseases, radiation scares, pandemics, and nuclear disasters, to name but some amongst

myriad topics [27, 28, 29, 30, 31]. Researchers have also critiqued and built upon original conceptual underpinnings of SARF and extended theoretical considerations to include more fine grained understandings of the contributions of such variables as culture, emotion, and politics to amplification and attenuation, as well as their impacts on risk perceptions and institutional processes before, during, and after major incidents and scares [32, 26].

2.2 Risk amplification: The role of news media

One key component of SARF that has received considerable research attention is the pivotal role played by the media – including newspapers, TV, film, radio and latterly social media [25, 17]. As people don't typically encounter exposure all risks immediately or directly, SARF originally gave traditional news media prominence as a key amplification station due to its central role in filtering, transferring, and circulating risk signals through public communication channels around society [25, 17, 33, 34, 35]. The importance of the media is underscored by study findings showing that a more rapid increase in media messaging is associated with ‘high dread’ risks, which are known to arouse substantial fear and anxiety [31]. At the other end of the spectrum, research also shows that people’s media use is associated with the perceived severity of health hazards [36]. Further studies have similarly found that those risk stories which highly resonate with the concerns of news consumers are also more likely to be shared with others, especially if they contain elements which create strong emotional arousal - awe, anger, anxiety, fear [37, 20].

News media reports may as a consequence be understood to help cultivate and frame an audience’s understandings and appraisals of risk in ways that can serve to

raise awareness and understanding as well as give reassurance or conversely cause alarm [38, 39, 40]. Corporations, NGOs, government agencies, or indeed concerned individuals, thus often seek to increase awareness and gain sympathy by aiming to publicize their version of events with the help of the media [26]. Consequently, contra to some criticisms, it has been remarked that the amplification of public risk stories is not a passive transfer of publicly derived facts from producers to wider audiences [26].

SARF has subsequently been applied to explain the volume of news over time with respect to differences in news framing and content between media channels and platforms stories [41, 33]. As the ‘gatekeepers’ of news, journalists, editors and/or media owners determine what stories to publish and take into consideration how the characteristics of a risk story will shape its ‘newsworthiness’ in light of their own ‘news agendas’ [35, 42]. Newsworthiness is conventionally assumed to correspond to the features of reports and media stories that people find intrinsically interesting and wish to hear about. This is respectively enhanced by attributes such as the prominence and importance of events, whether there is a human interest angle, and if there is any conflict and controversy attached to the issue, along with the timeliness and proximity of the events covered [43]. To take one illustrative example, news media coverage of the A/H1N1 ‘swine flu’ virus outbreak in Spring 2009, was underpinned by ‘risk-amplifying frames’ that placed emphasis on conflict and damage, thereby attempting to draw out the ‘drama’ and emotion of the situation to gain public attention [33].

Research has subsequently observed that, aside from some deviations, the traditional ‘yardsticks of newsworthiness’ broadly equate to the ‘signal value’ of risk

stories, and by extension the selection, extent, duration, and half-life of risk news attention and coverage [25]. In this way, the newsworthiness of risk stories contributes to the overall levels and patterns of media coverage, known as ‘issue attention cycles’, given over to reporting risk [25, 44]. Research to date has generally indicated patterns of risk related media coverage broadly exhibit an initial increase over a short period at the beginning, which subsequently decreases over time. However, this can vary by context and duration, with dramatic ‘risk events’ characteristically the focus of explosive episodic coverage that soon drops off, and long term or continuous risks being the focus of fluctuating news coverage which may ebb and flow [25, 44]. Risk stories may be particularly enduring and have a larger news cycle if they have a strong narrative, which is to say a ‘good storyline’ that resonates across different social, political and policy arenas [45], especially if it is high in both newsworthiness and risk signal value [25], as evidenced in cases such as the controversial sinking of the Brent Spa oil platform [46]

However, while the resonance of risk news stories is argued to be key to achieving amplification [47], the relationship between media coverage and levels of public concern is not always considered to be directly proportionate to the severity of the risk in question [48, 49, 50, 51]. Further, these associations do not always represent a linear causal relationship [52, 53]. For instance, the filtering of risk information by the media may also have an attenuating, or dampening effect, as when news coverage downplays the risk in question or simply does not report on it at all or in any great detail [24, 18]. News media outlets may for example choose not to run reports on a risk story if the ‘scoop’ was obtained by a rival outlet.

Meanwhile, studies into natural language uses of the word ‘risk’ also show that

it can be variably incorporated into news stories to convey different meanings. One early research study [54] found that ‘risk’ can be employed both as a noun or a verb, and can incorporate positive or negative semantic prosodies to convey particular ideas, understandings, and perspectives that are reflective of different contexts of everyday life and activity. Interpretations of specific risk terms and concepts may also reflect particular associations and carry different meanings between cultural groups, and their wider significance may respectively be independent of what originally occurred or was intended [17]. Consequently, this means that the use and interpretation of risk terms and concepts in news media is multidimensional and not simply be confined to describing major risk events and their characteristics. For these reasons, the ways in which the media actively selects, interprets, filters and relays risk information, whether simply to report things factually, or to amplify or dampen concerns, is of especial interest [18, 41, 52]. Albeit, as long forewarned by the authors of SARF, conclusions based solely on the volume of media coverage should be treated with some caution [17, 24].

SARF studies have also made important strides towards understanding how the risk signals proliferate through digital and social media platforms [48, 55], including with respects to COVID-19 outbreak [56, 57]. As noted above, the infodemic metaphor was first coined to conceptualise the rapid explosion and widespread distribution of both factual and erroneous information [14]. The global news-media was accordingly quickly mobilised in response to the outbreak and became a vital source of information and opinion once this novel ‘high dread’ infectious disease took hold [58]. Research indicates that few people trust social media platforms for COVID-19 related information, whereas those relying on news organisations as a source of in-

formation feel better informed [59]. This is thought to be due in part to factual news media generally being considered to provide more timely, accessible, reliable, and independent information compared to many other information sources [60]. This is not to say that factual news media is infallible or has been free from misreporting during the pandemic; numerous studies have focused on aspects such as misinformation in social media [61, 10, 12]. Traditional media can also find itself be the source of incorrect information, whether through problems of bias or news reports that subsequently turned out to be false in light of fast changing scientific knowledge [9, 12]. This does however underscore the significant contribution of news media to public understanding of the emergence, risks, and complexities of the COVID-19 pandemic.

2.3 This study

Trends towards the digitization of news media have afforded fresh opportunities for studying ‘media explosions’ of risk news. Earlier research was often subject to practical constraints and sample size limitations when gaining access to news repositories [62]. Together with increased computer processing power, studies have more recently been able to fully capitalize on the greater availability of vast amounts of digitised media data to examine key trends, topics, and dynamics in risk news [63, 64]. This research has incorporated computer assisted methods for all manner of purposes, such as to: monitor media in ‘real time’ [65]; collect and analyse historic media archives [66]; examine news coverage of acute ‘events’ [56]; track enduring ‘issues’ [67]; provide focused analyses of single risks [68]; and make comparative observations derived from news coverage of multiple risks [69]. In the process, this work has provided a more nuanced understanding of the focus, dynamics, and sentiment of

media attention to risk whilst also further demonstrating the applicability and value of computer assisted methods and analytical advances in corpus linguistics [66, 64].

Regarding COVID-19, a recent study employed data drawn from 3 different news sources in the US and China to show that media attention to the pandemic varied in content and over time between each country [44]. The US news cycles were characterised by reports of conflict, in stark contrast to the more consensual tone of Chinese news media. Another study employed five different news sources, along with data drawn from social media, showing COVID-19 coverage reflected diverse themes that were associated not only with conventional health topics, but also politics and the economy [70]. This work underscores the broader newsworthiness and reach of COVID-19 whilst also finding evidence of misinformation being shared by ostensibly factual media at the beginning of the pandemic.

While these studies have generated important insights into news reporting of COVID-19, to date current understandings of the infodemic remain loosely defined and poorly grounded in existing research literature [15] and the intersection between media amplification of risk and the infodemic has been elided. In view of this lacuna, we suggest that when put into a historical context, the current pandemic arguably does not fully warrant the ‘unprecedented’ characterisation it is often ascribed. While it might be said that the ‘modern disaster’ of COVID-19 does in some sense present a ‘new species of trouble’ [71], it must first be acknowledged that infectious diseases (including coronaviruses) have been causing health problems, and attracting substantial news coverage, throughout human history [72]. Against this backdrop, the task of examining the dynamics and impacts of dramatic risk events has been a long preoccupation of research inquiry within the social and decision sciences [6, 18]. Yet,

the possible contributions of SARF, have with few notable exceptions, been largely overlooked in studies of COVID-19. In this study, we are therefore interested in key questions concerning whether the explosion in news coverage described as an infodemic may be understood in SARF terms as a conventional case risk amplification, or if it might instead represent a distinct phenomenon evidenced by unique qualities that warrant special characterisation and definition.

3 Methodology

3.1 Determining a relevant news dataset

At the heart of our study, we collected US news articles that include risk-related discussion, covering the period from January 2000 to July 2021, from both print and online platforms. We made an initial decision to focus on identifying articles that have a high chance of being related to risk, rather than identifying all possible risk articles. The former approach requires refined search terms that might exclude articles at the margins of discussing with risk. In contrast, the latter approach captures more articles, but can potentially introduce articles that are unrelated to risk. For our topic modeling technique, as detailed in Section 3.2 that follows, it is necessary to exclude articles outside the core area of focus, so the refined approach was preferred.

3.1.1 Searching for risk discussions in US news

The *LexisNexis* news database has previously been used to identify news stories relevant to understanding risk perceptions, including the social amplification of A/H1N1

(‘swine flu’) [33] and the risks of nanotechnology [73]. For that reason, we employed the *LexisNexis* database as the external source of news stories for this current study. In order to determine which *LexisNexis* articles were relevant to our analysis, we screen all available articles for keywords about risk using the 2018 *Society for Risk Analysis* (SRA) Glossary of base risk terms¹. This glossary of risk terms provides an ability to scan articles based on "well-defined and universally understood terms and concepts" regarding risk (SRA Glossary, pg 1). However, some of the SRA keywords have dual meanings in English-language usage. To address this, from the glossary we remove terms where the primary meaning, according to Merriam-Webster, is not risk-related. This left us with a reduced set of terms: *risk; uncertainty; harm; damage; hazard; safe; safety; security; secure; threat; resilience; vulnerability*. To define an article as being related to risk, we required at least one mention of the term *risk*, and at least one additional SRA reduced-glossary term in the article text.

The initial search results from LexisNexis contained 520,374 print and online news articles. After filtering out close duplicates (a common feature of news due to syndication) and other verification checks, we obtained 283,165 news for topic modeling. Table 1 above summarises the text search parameters. One notable feature of the table is the increase in the volume of risk-related news coverage over time.

3.2 Topic modeling approach

3.2.1 Processing textual data

Our starting point is the dataset of 283,165 risk-related news articles, which we processed in order to run our desired tests. We employed Natural Language Processing

¹The SRA Glossary is publicly available at: www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf

(NLP) techniques for this purpose to convert the content of our news articles into an analyzable format [74].

In the first step, we identified the relevant part of each news article that referred to risk. We chose this over using the full article, in order to reduce the amount of non-risk discussion of news stories included in our final text dataset. We, therefore, retained, for each article, only the sentences that have a risk term in them, and the sentence before and after the key sentence. This allowed us to retain both the risk-relevant discussion as well as the most pertinent context.

Second, working in a *Python* coding environment, the news article was tokenized, or separated, into individual words and other writing features, such as numbers and punctuation. We removed text features such as punctuation, uppercase letters, extra white spaces in the text, numbers, and special characters from these individual terms. This also included high-frequency terms known as ‘stop words’, which are functional words used so frequently so as not to be informative as to context and meaning. Common stop words include: *and*, *of*, *the*. We use the *Python* stopwords lists from *Natural Language Toolkit (NLTK)*, and *spaCy*.

In step three, we adjusted the remaining words in our dataset by converting them to the lemma (the dictionary base) of the word. We used a word’s lemma to match different forms of words to their common lemma. For example, using this approach, the words finance, finances, financial would all be mapped to the single word: finance. To implement this, we employed the *Python* package: *spaCy Lemmatizer*². We also inspected the remaining terms to see if any need to be manually adjusted. This included the most frequent terms and manual removal of terms that are unrelated

²spaCy package: <https://spacy.io/api/lemmatizer>

to risk or risk discussions, such as terms related to the day of the week or other time signifiers. This manual inspection is labor-intensive, but significantly improves the quality of the eventual analysis [75].

In step four, we prepared the remaining terms for testing. We generated a Term Document Frequency matrix, which represents news articles (documents) in matrix form in which the rows correspond to all remaining terms across all documents, columns correspond to all documents in the corpus, and cells correspond to the weights of the terms per document. We adopted the usual choice, which is to consider one word (one-gram), or two consecutive word (bi-grams) terms. After building the corpus of terms, we found some risk-related words (for example *risk*) to be present in a large number of articles, as well as a long-tail of rarely used terms. We, therefore, excluded terms that appear in more than 10 percent of all news (on the grounds that these words are too common to be useful in separating risk sub-topics), as well excluding terms that appear less than 10 times (on the grounds that these words are too rare to be useful in separating risk sub-topics). The latter adjustment removed most bi-grams as bi-grams are, by their construction, combinations of terms.

3.2.2 Identifying risk topics with topic modeling

Topics in our news dataset were identified using a Latent Dirichlet Allocation (LDA) model [76, 77]. We applied the LDA implementation of the *Python Gensim* package alongside the *Mallet* [78] package. *Mallet* has an efficient implementation of the LDA which is known to run faster and give clearer topic segregation.

LDA starts with a base assumption that the words chosen by the writer reflects the topics they wish to write about. By inspecting the observable words, word co-

occurrences, and word similarities across related documents, we can therefore uncover the topics which are hidden (or 'latent'). LDA uses a probabilistic technique based on Dirichlet priors and a machine-learning based expectation-maximisation (EM) algorithm to identify these hidden topics across documents.

Our initial step in LDA topic modeling estimated the number of topics we think best described our body of articles, this is domain-specific and therefore benefited from domain expertise. Dirichlet priors then allocated words across this number of topics, initially mechanically, and later incorporating probabilistic learning. The expectation-maximization algorithm produces a definitive 'best' allocation of groups of words to topics. The key decision for the researcher, therefore, is the specification of the number of hidden topics. In practice, we tested across a range of potential total number of topics (between 5 and 55), using *coherence scores* and inspection of the topics we concluded that a 35-topic model best represented our data.

The output from LDA for each topic, is a topic number and a list of terms that best describe that topic. Each term has a *beta* value that quantifies its importance to the topic. To assign labels to each topic, two researchers independently coded, and then collectively agreed each topic label based on the highest beta terms per topic, along with a random sample of 20 news articles pertaining to that topic, and assigned a descriptive topic label. We followed standard methodological protocols common in data-driven thematic analysis to converge on topic labels [79, 80].

After labeling, we excluded three topics and any news articles which have those topics as their primary focus as they were considered irrelevant to risk. The presence of irrelevant topics is a common feature of topic modeling [81]. In this way, our total corpus of news articles fell from 283,165 articles after initial filtering, to our final

dataset of 271,854 articles. The remaining 32 topics derived from our LDA analysis provided the basis for analysis and are reported in the Appendix, Tables A1 and A2. These tables are presented in the order in which the topics are numbered by the LDA process, and report the topic labels, the top terms per topic, the term beta values, along with the proportion of all news articles where the primary topic of the news article is related to that topic.

After labelling the 32 topics, we began the process of labelling the seven broad thematic categories of US News discussion, by employing the same methodological protocols outlined above. The underpinning (as determined by LDA) first order topics labels were used to develop consensus on labeling the second order thematic categories of risk discussion. The composition of the second order categories and first order topics are reported overleaf in Table 2, with Figure 2 showing the distribution of the category themes over time, it is here that our first results are presented.

4 Results

4.1 The topics and categories of US risk news

Figure 1 shows a word cloud of the most popular terms in the final dataset, and their relative usage within the dataset based on their size within the cloud. From this we can see a wide range of risk-related news coverage, including prominent health related risk terms such as *disease*, *injury*, and *life*. Also evident are geopolitical terms such as *war*, *security*, and *attack*, and economic risk: *investor*, *investment*, and *money*. When underlying topics were grouped together the following categories of risk discussion emerged: *business*, *domestic affairs*, *entertainment*, *environment*, *geopolitics*, *health*,

and *technology*.

We see from both Table 2 and Figure 2 that the category of domestic affairs takes up approximately one-third of risk-related news reported. This comprises daily news about the US and the operation of its institutions, including issues such as crime and policing, politics, welfare, and local issues. It is clear from the timeline in Figure 2 that news media reporting of risk has been impacted by the COVID-19 in the final 18 months of our data-set, with each of our categories displaying clear increases in news coverage when compared to prior years.

4.2 SRA Glossary terms in US news risk discussions

Figure 3 presents the presence and influence of the SRA 2018 reduced-glossary terms across the categories. When we remove the term *risk* from this analysis (as it is required to be present in all articles as a search criteria), the three most frequent terms are *safe / safety*³, *security*, and *threat*.

As observed in Figure 3, we can garner further insights into the use of SRA terms by US news media when reporting risk. *Security* and *vulnerability* are used extensively in technological risk communication, while the key health risk terms are *harm* and *safe / safety*. Geopolitical risk news reporting is concerned with *security* and *threat(s)*, and environmental risks are communicated as *hazard(s)*, *damage*, and *resilience*. The entertainment and domestic affairs categories have a reasonable distribution of all the risk terms, with *resilience* and *vulnerability* figuring most prominently for the entertainment category and *safe* and *security* used most extensively for domestic affairs respectively. Lastly, business risk news reporting is the

³Safe incorporates the term *safety* due to our data processing rules, which reduces terms to their common core. See Section 3.2.1 for further details.

main source of the *uncertainty* term, which is to be expected given its focus on the financial aspects of risk news reporting.

4.3 Risk amplification within and across risk categories

The word clouds for each of the seven risk categories are presented in Figure 4, and the category and sub-topic timelines are provided in Figure 5. The word clouds represent the focal risk interests of each category, while the timelines allow us to identify critical signal events and amplification across the time period. A particular strength of our topic modeling analysis is that it allows for a more granular view of the individual topics and timelines of each category. We discuss some notable events in each category's timeline now. At this stage of the analysis, we do not yet consider COVID-19 (which has an evident impact in many categories), as this is addressed in the subsequent section.

For example, for business risk (Figure 5, Chart (a)) spikes in the volume of news reported in the ‘financial markets’ topic respectively demonstrate the impacts of the dotcom crash in 2001-02, the global financial crisis of 2007-2009, and the 2018 mini-crash (when the S&P500 fell 6.24% during the year) in the financial markets topic.

By contrast the domestic affairs category contains 11 underlying topics, meaning individual trends within categories are more difficult to discern at the aggregate level. However, once again, the topics tell a tale of the history of domestic signal events in the US. For instance, we see the large growth in the ‘air safety’ topic relating to 9/11 and the subsequent risks surrounding air travel. An increase in news coverage can be seen regarding the election of President Trump, and subsequently his loss

of the next election. Heightened differences between the political parties, and the risks this entails for political stability, are also evident in more recent years in the ‘party politics’ topic. Two other interesting features are the rise in the discussion of ‘policing risks’ in 2020, as the Black Lives Matter protests took place, and also a notable spike in regulation risk reported in 2010 as the major Dodd–Frank Wall Street Reform and Consumer Protection Act passed congress.

The entertainment category is characterized by a reasonably even distribution of risk news discussion across the topics of ‘arts’, ‘human interest’ and ‘sports’. However, the word cloud for this category (Figure 4) indicates that sports news stories are highly prominent, with news coverage focusing on such issues as player injuries and the risk-related decisions of players on the sports field. Arts discussions revolve more around the risk-taking of characters within productions, and the risks involved in forming and maintaining relationships.

When we examine news media reporting of environmental risks, and the two component topics, we can see strong variability over the time series brought about by natural disaster signal events. The 2010/11 peak primarily concerns the Fukushima nuclear disaster, which was a global news story, a major domestic hurricane season (particularly Hurricane Irene), and the Deepwater Horizon oil spill. The news stories of the time show that Fukushima led to discussions about the risks of nuclear energy, and Deepwater engendered a discussion about the risks and responsibilities of major oil exploration. Domestic weather disasters are evident across the timeline, with the 2005 Hurricane Katrina which devastated New Orleans, and 2017’s Hurricane Harvey. These domestic weather events are a natural focus in this category given the direct relevance of these risks to population safety. Other stories captured in

these timelines include wildfires and growing awareness of the risk of climate change.

In the technology category, we see an overall growth inconsistent reporting of risks associated with *cyber-security*, *R&D* and *technology failure* over the time period. In the first half of our sample news reporting consists of event-driven spikes, while in the second half there is a constant volume of news as new technology embedded itself firmly in the lives of US citizens. Our sample period starts by showing a decline in the reporting of technology risk news immediately after the Y2K problem (Millennium bug), which had taken up so much attention in the pre-2000 time period and has been argued to be a social amplification event [82]. In 2003 there is a large volume of news coverage associated with the Space Shuttle Columbia disaster linked to technology failure. The timeline of articles from this period shows a quick media shift from the specific shuttle failure to the systematic risk failings of NASA that allowed the disaster to occur. Unfortunately, the smaller spike in 2005 is also related to another NASA space shuttle flight, that while not fatal, illustrated some of the same design flaws as the Columbia shuttle. In the second period of our sample we see the large rise of cyber-security-related news associated with a long list of well-covered data breaches concerning such companies as Target, Yahoo!, Equifax, and Sony. These cyber-security risk cases (amongst many others), are also matched by growing reporting around general data security and advice on reducing those new risks.

The geopolitics category shows significant rises in the post 9/11 years and the President Trump presidency. Of note, we see the reduction of war discussion over the timeline, after the elevated period sustained around the wars in Afghanistan, and particularly Iraq. We also see the rise in discussion around less-direct conflicts,

including the rise of risk discussion related to conflicts with Russia, Syria, and Iran. Of interest, also, is the recent rise in discussion around trade, especially with China, and this can be clearly seen in the timelines. Trade, once a minor topic, has grown to compete with war as a geopolitics issue, as US media news reporting appears to converge around a belief that conflict with China has a strong trade element.

Within the health category, the infectious disease topic clearly highlights the emergence of the Ebola virus in 2014, and H1N1 influenza (swine flu) in 2009 - a signal event previously the focus of research investigation [33]. In the medicine topic we see the out-sized growth and influence of the 2004 Vioxx drug removal from the market. In the health category, more than any other, we also see the reporting of risks in order to increase public awareness evident in the news surrounding female health and medicine safety.

4.4 The COVID-19 pandemic as a major instance of risk amplification and ripple effects

We now proceed to analyze the period of COVID-19. This is driven by our interest in this as a potential instance of an infodemic, by the dramatic rise in health risk reporting over the pandemic time period ⁴, and by the clear spillover to other risk categories, especially business, domestic affairs, and entertainment risk.

Figure 2 shows the overall timeline of risk reporting, where it is evident that COVID-19 has dominated risk news unlike any other issue. COVID-19 started off as an infectious disease signal event that grew by 2368%, vastly more any other infectious disease events within our data. This lead to a 72% increase in all risk

⁴We only consider 2020 in this period at it was the only year with complete annual data for comparison to other years. As such we compare data from 2019 to 2020 within this section as a temporal point of reference

discussions across all topics within US news media, totalling 23427 articles in all categories. Simultaneously, the reporting of COVID-19 also evidenced Stage 2 ripple effects as other risk news within the health category began to be crowded out, leading to a new narrative within health topics. As seen in the health chart of Figure 5 food safety was the only topic to see a notable increase in coverage during this period with a rise of 243% in 2020. However, this increase primarily relates to news reporting around the safety of food establishments in light of COVID-19, indicating the spill-over of COVID-19 into other categories.

Evidence of ripple effects arising from COVID-19 were not exclusive to the health category, as categories typically unrelated to infectious disease, or health more generally, began to experience an immediate spillover of this risk signal event into their own risk categories, crowding out all other news in a way not observed any time before in our data. Figure 6 demonstrates this effect, in which the word cloud of the most popular terms in all of 2020 either speaks directly to the infectious disease, or about the consequences of it. For example, news relating to job security (due to temporary unemployment) grew 255%, and child education (due to closing of schools) grew 216%. These increases reflected discussions across society of enforced lock-downs as way to bring the disease under control along with their consequences.

5 Discussion

The central research question driving this study primarily concerns whether the COVID-19 infodemic may be understood to constitute a novel form of risk amplification. In order to answer this question, we formulated an innovative research

strategy, which incorporated risk-related search terms drawn from the SRA Glossary along with computer assisted methods, in order to generate a corpus of US risk focused news coverage. This corpus spanned across all manner of risk-related stories reported in the US media over the past two decades. By doing so, we provide a newly documented recent history of risk amplification against which it might then be possible to identify the so called COVID-19 infodemic as a distinctive phenomenon.

Several standout aspects of our analysis and findings warrant further discussion. Firstly, our approach demonstrates that the SRA Glossary provides a valuable resource of "universally understood terms and concepts" (SRA, 2018) that can be usefully applied to the task of collating and filtering a corpus of risk focused news coverage. The novel use of the SRA Glossary employed by this study helped us to generate a corpus constituting an expansive and comprehensive risk news dataset. Subsequently, we were then able to track general increases and decreases in risk reporting over our time series, as well as identify the waxing and waning prominence of individual risk events and topics, along with the broader risk categories they helped to form. Using a broad set of risk related terms in this way notably allowed for a more linguistically nuanced longitudinal and thematic comparative analysis of our time series than might otherwise have been the case had we simply selected articles using the word 'risk' alone. The broad scope of our corpus thus stands in contrast to previous studies which have typically either attempted to gather news in relation to one specific risk, or used a much narrower range of newspaper sources. This could be considered as the difference between taking a narrow 'spotlight approach' to researching risk phenomena versus adopting a more broadly encompassing 'floodlight approach' that includes all risk phenomena reported.

Secondly, our study findings indicate that US media attention to risk is not confined solely to reporting extraordinary events and crises, but also that discussion of risk in the news reflects a wide cross-section of different segments and features of everyday life as found in early work by [54]. This complexity was linguistically distinguishable according to the varied use and prevalence of different risk-related terms with regard to certain risk contexts. Aside from the term *risk*, the most prevalent SRA Glossary terms in our corpus were *safe*, *security*, *threat* and *harm*. The overall prevalence of these terms was evidenced both within the wordcloud of our complete time series in Figure 1 and more specifically in relation to individual risk categories, particularly *domestic affairs*, *geopolitics*, and *health* (Figure 3). The terms *uncertainty*, *vulnerability* and *resilience* were, by contrast, much less prevalent and more evenly spread across individual risk categories. This suggests that risks described in conjunction with the former terms might be considered comparatively more ‘newsworthy’ than those described using the latter terms [34, 42]. From a SARF perspective, we might also interpret risks characterised in accordance with the former terms as achieving higher resonance as they were able to achieve greater levels of media amplification [47].

However, these broad patterns of linguistic association do raise an interesting question concerning whether the use of certain terms would help a particular risk to achieve higher resonance and greater amplification by the media. For instance, practically speaking, some risk related terms are used quite interchangeably, but if a government spokesperson described a particular risk in terms of safety, security, threat, and harm, would the risk in question gain more media attention than if the terms uncertainty, vulnerability, and resilience were used instead? There is as of yet

no definitive answer to this question suggesting that the relationship between linguistic associations and resulting levels of risk amplification warrants closer inspection.

Another key finding concerns the variable distribution of US risk news coverage across the different risk categories. The most prominent risk news category was *domestic affairs*, which obtained the highest news coverage both in terms of article count per year and the normalised percentage of articles across all of the years in our time series. This might be said to quite naturally reflect the US focus of the national news sample we obtained. However, there may also be other contributing factors. For instance, the high volume of news in this category may also reflect the fact that domestic affairs incorporates a high number of risk topics ($n=11$), ranging from *air traffic safety* through to *crime* and *policing*, as well as *education* and *elections*. Whilst these all reflect domestic concerns, it means that this category incorporates extraordinary events along with many everyday concerns such that domestic risk issues in one guise or another always have a high chance of being in the news. In our time series this risk category only dipped below 3000 articles once in the year 2009.

By contrast, the *environment* risk category contained the fewest topics and consistently received the lowest level of news coverage over the whole time series. The two environmental risk topics included *pollution* and *natural disasters*. These topics emerged in relation to acute incidents such as floods, storms and chemical spills (e.g. Deepwater Horizon oil spill, Katrina), as well as climate change. The comparatively small amount of media attention given to environment risk might be considered somewhat surprising for what might otherwise be considered an important risk category to society and that it contains both salient catastrophic risk events and the more enduring issue of climate change. Several possible explanations can be offered

for this finding. While it can be observed that major environmental disasters are typically significant new events, their occurrence also tends to be relatively isolated and infrequent. In this sense, environment risks do not arise as everyday events that might otherwise keep the environment regularly in the news, which might account for there being a smaller amount of news coverage for this risk category. In the case of climate change, while this is characteristically a longstanding systemic risk, it has not typically been considered to produce frequent newsworthy events with an everyday impact that might otherwise generate headline news. Indeed, there were no more than 5 major peaks in our time series for this category, with newsworthy events achieving more than 1000 articles only once in 2011. Lower news coverage may also be a facet of editorial decisions following a perception that climate change does not figure domestically in the US as a newsworthy concern warranting widespread media coverage. Journalists and editors across all news desks may as a consequence require greater education about different ways in which climate change can affect all arenas of life and how best to report this to their audiences.

Our results across risk categories clearly indicate the presence of SARP Stage 1 ‘signal events’ within US news media risk discussion [18, 25]. The amplification of these events is also evident from media reporting, with periods of sustained high-risk reporting within categories. Examples of this include the Afghanistan-Iraq war periods, the global financial crisis, the NASA shuttle disasters, and the recent China-US trade conflicts. Our topic modeling approach highlighting the rise of trade conflict (2018-2021) is matched by a UN Department of Economic and Social Affairs report [83] published around the time of the peak seen in our data for the trade topic, making a similar warning.

The additional benefit of our approach, compared to the UN report, is our ability to highlight the relative importance of trade conflicts compared to other geopolitical risk events at any given point in time, as well as to quantify time trends. Thus, we see how temporal events are well captured by the topic modeling approach, supporting our initial assertion that the SARF is still relevant as contemporary risk events emerge [17, 84].

COVID-19, as a SARF Stage 1 ‘signal event’, stands out as a sustained eruption of news reporting that led to immediate Stage 2 ‘ripple effects’ [17, 26] within, and importantly, between other categories of risk news reporting. Advice of what to do, what not to do, are evident in risk discussions within the environment category for upcoming hurricanes and storms. Further, when considered through the lens of road safety in the domestic affairs category, we can also evidence declines in news reports that raise public awareness. This is possibly tied to the significant fall in road deaths over the two decades of our timeline (from 15.8 deaths per 100,000 people in 2000, to 11.9 per 100,000 in 2019 according to UNHTSA data). These demonstrate how the reporting of risk news can shift perceptions in response to changing public campaigns, seasons, technologies, and the environment.

However, these examples of ‘signal events’ and their transferal of risk information and signals around society [25, 17, 33, 34, 35] are localised when compared to COVID-19. Originating within the ‘Infectious Disease’ topic of the ‘Health’ category, it began to rapidly crowd out some ‘usual’ forms of health risk news reporting particularly around female health and regular communications related to medicine safety. This phenomenon has been noted in news media discussion of the side-effects of COVID-19 [85], but not yet demonstrated in the academic literature until now. This is indicative

of COVID-19 being an instance of a ‘new species of trouble’ [71], not necessarily because it’s a coronavirus (it isn’t the first), but because of the wide-ranging, trans-boundary impacts of its existence in the fabric of society. Given that the work of [47] suggests that resonance is key to achieving amplification, one of the unintended consequences of ‘drama sells’ US editorial policies is that in the COVID era the SARF has been more important than ever. The volume of news media reporting (see Figure ??), particularly early on in the pandemic provided an important early risk information amplification station for the US population far beyond any other signal in the last 20 years. This confirms and extends prior findings examining a smaller subsection of US newspapers over the first 3 months of the COVID-19 crisis [44].

When this volume and dynamic movement across categories of news reporting is taken in the context of news story content as seen in Figure 6, it becomes clear that COVID-19 displays infodemic properties never seen before. So much so, that it now re-calibrates what it means for a signal event to be ‘amplified’ relative to all other societal risk, making it incumbent on future research to consider actual as opposed to perceived Stage 2 ‘Ripple effects’ when SARF is applied to future research studies.

6 Conclusion, limitations, and recommendations

Given the importance of news-media in generating awareness about risk and typically providing a more credible and trustworthy source of risk communication in many nations than social media, the results presented by this study are timely and significant. In this study, we are able to make a number of observations regarding

the both the association of the US news media to the categories of risk discussion, the COVID-19 infodemic, and its role in risk amplification and attenuation processes more broadly.

In particular, this is the first academic study that, to our knowledge, is able to provide confirmation of the existence of the infodemic claimed by the WHO to be associated with COVID-19. By drawing a comparative analysis with risk amplification events historically we are also able to map out and identify at a macro level the contours, dynamics, and impacts of major ‘signal events’ in US news coverage over the previous twenty years. Which is to say, the majority of previous SARP research might be considered to have adopted a ‘spotlight’ investigative approach to examine certain risk events or hazards in a particular place at a specific point in time, this study adopted a ‘floodlight’ investigative approach, which was able to illuminate the relative significance of ‘signal events’ and ‘ripple effects’ at a grander scale than previously achieved. This meant that it was also possible to identify widely varying types of amplification and attenuation processes that have taken place, as well as a variety of different types of Stage 2 ‘ripple effect’. This was exemplified by multiple cases of risk amplification in particular, including such events as 9/11. COVID-19 being the most emphatic case of risk amplification recorded to date and deserving of the status ‘infodemic’.

Future research in this area should consider textual analysis using other languages, which in itself is a research limitation of this current study. Further, this study attempted to define the main categories of risk as portrayed by the US news media, highlighting those events that have shaped public perceptions of risk. It was beyond the scope of this current research to micro-analyze all seven categories of risk and the

second-order risk events that have occurred within them over the last 20 years for framing and sentiment of the news stories. Now that our data clearly define these risk categories, the risk research community can address each of these categories individually, investigating the key events as signposted by this current study in more detail across an international sample of newspapers. Interestingly, the ability to compare news-media reporting of major risks (e.g., COVID-19) with more typical everyday risks (e.g. The common cold) via machine learning textual analysis could uncover new relationships that allow for more refined Public Health risk communication - a point highlighted by the World Health Organisation. Further, given the long time period under investigation, we did not include social media as a news source. Future research could employ a multi-level modeling approach to investigate how social and traditional media reflected each other. This could be extended to determine how effective risk communication interventions were at curb misinformation.

Finally, given the extensive nature of our analysis it may be worthwhile for institutional bodies such as the Society for Risk Analysis to offer media training to news agencies when reporting certain components of risk that our analysis suggests get less attention. This may aid the detection of emerging infodemics in the future using the methodology outlined by our study.

References

- [1] Paul Slovic. Risk perception and risk analysis in a hyperpartisan and virtuously violent world. *Risk Analysis*, 40(S1):2231–2239, 2020.
- [2] Andrew RA Conway, Linda J Skitka, Joshua A Hemmerich, and Trina C Kershaw. Flashbulb memory for 11 September 2001. *Applied Cognitive Psychology*, 23(5):605–623, 2009.
- [3] Bart Vyncke, Tanja Perko, and Baldwin Van Gorp. Information sources as explanatory variables for the Belgian health-related risk perception of the Fukushima nuclear accident. *Risk Analysis*, 37(3):570–582, 2017.
- [4] Helene Joffe and Nadia Bettega. Social representation of AIDS among Zambian adolescents. *Journal of Health Psychology*, 8(5):616–631, 2003.
- [5] Helene Joffe and Ny Louis Lee. Social representation of a food risk: The Hong Kong avian bird flu epidemic. *Journal of Health Psychology*, 9(4):517–533, 2004.
- [6] Paul Slovic. Perception of risk. *Science*, 236(4799):280–285, 1987.
- [7] Jamie K Wardman and Gabe Mythen. Risk communication: Against the Gods or against all odds? Problems and prospects of accounting for Black Swans. *Journal of Risk Research*, 19(10):1220–1230, 2016.
- [8] Charles Haas. Coronavirus and risk analysis. *Risk Analysis*, 40(4):660–661, 2020.
- [9] Jamie K Wardman and Ragnar Lofstedt. COVID-19: Confronting a new world risk. *Journal of Risk Research*, 23(7-8):833–837, 2020.
- [10] Nicole M Krause, Isabelle Freiling, Becca Beets, and Dominique Brossard. Fact-checking as risk communication: The multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research*, 23(7-8):1052–1059, 2020.
- [11] Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe van der Bles, and Sander van der Linden. Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, 7(10):201199, 2020.
- [12] John Zarocostas. How to fight an infodemic. *The Lancet*, 395(10225):676, 2020.
- [13] Jamie K Wardman. Recalibrating pandemic risk leadership: Thirteen crisis

- ready strategies for COVID-19. *Journal of Risk Research*, 23(7-8):1092–1120, 2020.
- [14] David J Rothkopf. When the buzz bites back. *The Washington Post*, 11:B1–B5, 2003.
 - [15] Felix M Simon and Chico Q Camargo. Autopsy of a metaphor: The origins, use and blind spots of the ‘infodemic’. *New Media & Society*, page 14614448211031908, 2021.
 - [16] Stephan Leitner, Bartosz Gula, Dietmar Jannach, Ulrike Krieg-Holz, and Friederike Wall. Understanding the dynamics emerging from infodemics: A call to action for interdisciplinary research. *SN Business & Economics*, 1(1):1–18, 2021.
 - [17] Roger E Kasperson, Ortwin Renn, Paul Slovic, Halina S Brown, Jacque Emel, Robert Goble, Jeanne X Kasperson, and Samuel Ratnick. The social amplification of risk: A conceptual framework. *Risk Analysis*, 8(2):177–187, 1988.
 - [18] Nick Pidgeon, Roger E Kasperson, and Paul Slovic. *The Social Amplification of Risk*. Cambridge University Press: Cambridge, 2003.
 - [19] Chris Fife-Schaw and Gene Rowe. Public perceptions of everyday food hazards: A psychometric study. *Risk Analysis*, 16(4):487–500, 1996.
 - [20] Lynn J Frewer, Susan Miles, and Roy Marsh. The media and genetically modified foods: Evidence in support of social amplification of risk. *Risk Analysis*, 22(4):701–711, 2002.
 - [21] Barbara Reynolds and Matthew Seeger. Crisis and emergency risk communication as an integrative model. *Journal of Health Communication*, 10(1):43–55, 2005.
 - [22] Kenneth A Lachlan, Patric R Spence, Xialing Lin, Kristy Najarian, and Maria Del Greco. Social media and crisis management: CERC, search strategies, and Twitter content. *Computers in Human Behavior*, 54:647–652, 2016.
 - [23] Xianlin Jin and Patric R. Spence. Understanding crisis communication on social media with CERC: Topic model analysis of tweets about Hurricane Maria. *Journal of Risk Research*, 2020.
 - [24] Roger E Kasperson and Jeanne X Kasperson. The social amplification and

- attenuation of risk. *The Annals of the American Academy of Political and Social Science*, 545(1):95–105, 1996.
- [25] Andrew R Binder, Michael A Cacciatore, Dietram A Scheufele, and Dominique Brossard. The role of news media in the social amplification of risk. *The SAGE Handbook of Risk Communication*, 69, 2014.
 - [26] Jamie K Wardman and Ragnar Löfstedt. Anticipating or accommodating to public concern? Risk amplification and the politics of precaution reexamined. *Risk Analysis*, 38(9):1802–1819, 2018.
 - [27] Ji Bum Chung and Gi Woong Yun. Media and social amplification of risk: BSE and H1N1 cases in South Korea. *Disaster Prevention and Management*, 22(2):148–159, 2013.
 - [28] James Flynn. Nuclear stigma. *The Social Amplification of Risk*, pages 326–352, 2003.
 - [29] Deborah C Glik. Risk communication for public health emergencies. *Annual Review of Public Health*, 28:33–54, 2007.
 - [30] Dominic Duckett and Jerry Busby. Risk amplification as social attribution. *Risk Management*, 15(2):132–153, 2013.
 - [31] Robert D Jagiello and Thomas T Hills. Bad news has wings: Dread risk mediates social amplification in risk communication. *Risk Analysis*, 38(10):2193–2207, 2018.
 - [32] Jeffrey R Masuda and Theresa Garvin. Place, culture, and the social amplification of risk. *Risk Analysis*, 26(2):437–454, 2006.
 - [33] Constanze Rossmann, Lisa Meyer, and Peter J Schulz. The mediated amplification of a crisis: Communicating the A/H1N1 pandemic in press releases and press coverage in Europe. *Risk Analysis*, 38(2):357–375, 2018.
 - [34] Peter LM Vasterman. Media-hype: Self-reinforcing news waves, journalistic standards and the construction of social problems. *European Journal of Communication*, 20(4):508–530, 2005.
 - [35] Peter Vasterman, C Joris Yzermans, and Anja JE Dirkzwager. The role of the media and media hypes in the aftermath of disasters. *Epidemiologic Reviews*, 27(1):107–114, 2005.

- [36] Jonah Berger and Katherine L Milkman. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205, 2012.
- [37] Meredith E Young, Geoffrey R Norman, and Karin R Humphreys. Medicine in the popular press: The influence of the media on perceptions of disease. *PLoS One*, 3(10):e3552, 2008.
- [38] Ralph Hertwig, Thorsten Pachur, and Stephanie Kurzenhäuser. Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4):621, 2005.
- [39] Thorsten Pachur, Ralph Hertwig, and Florian Steinmann. How do people judge risks: Availability heuristic, affect heuristic, or both? *Journal of Experimental Psychology: Applied*, 18(3):314, 2012.
- [40] Craig W Trumbo and Katherine A McComas. The function of credibility in information processing for risk perception. *Risk Analysis*, 23(2):343–353, 2003.
- [41] William R Freudenburg, Cynthia-Lou Coleman, James Gonzales, and Catherine Helgeland. Media coverage of hazard events: Analyzing the assumptions. *Risk Analysis*, 16(1):31–42, 1996.
- [42] David Manning White. The “gate keeper”: A case study in the selection of news. *Journalism Quarterly*, 27(4):383–390, 1950.
- [43] Pamela J Shoemaker and Stephen D Reese. *Mediating the message*. White Plains, NY: Longman, 1996.
- [44] Christopher D Wirz, Anqi Shao, Luye Bao, Emily L Howell, Hannah Monroe, and Kaiping Chen. Media systems and attention cycles: Volume and topics of news coverage on COVID-19 in the United States and China. *Journalism & Mass Communication Quarterly*, page 10776990211049455, 2021.
- [45] Claire McInerney, Nora Bird, and Mary Nucci. The flow of scientific knowledge from lab to the lay public: The case of genetically modified food. *Science Communication*, 26(1):44–74, 2004.
- [46] Vian Bakir. Greenpeace v. shell: Media exploitation and the social amplification of risk framework (SARF). *Journal of Risk Research*, 8(7-8):679–691, 2005.
- [47] Ortwin Renn. The social amplification/attenuation of risk framework: Application to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 2(2):154–169, 2011.

- [48] Ik Jae Chung. Social amplification of risk in the Internet environment. *Risk Analysis*, 31(12):1883–1896, 2011.
- [49] Barbara Combs and Paul Slovic. Newspaper coverage of causes of death. *Journalism Quarterly*, 56(4):837–849, 1979.
- [50] Judith Lichtenberg and Douglas MacLean. The role of the media in risk communication. In *Communicating Risks to the Public*, pages 157–173. Springer, 1991.
- [51] Cornelia Betsch, Lothar H Wieler, and Katrine Habersaat. Monitoring behavioural insights related to COVID-19. *The Lancet*, 395(10232):1255–1256, 2020.
- [52] Henry Jenkins, Sam Ford, and Joshua Green. *Spreadable Media: Creating Value and Meaning in a Networked Culture*, volume 15. NYU press: New York, 2018.
- [53] John Fellenor, Julie Barnett, Clive Potter, Julie Urquhart, John D Mumford, and Christopher P Quine. Ash dieback and other tree pests and pathogens: dispersed risk events and the social amplification of risk framework. *Journal of Risk Research*, 22(12):1459–1478, 2019.
- [54] Craig Hamilton, Svenja Adolphs, and Brigitte Nerlich. The meanings of ‘risk’: A view from corpus linguistics. *Discourse & Society*, 18(2):163–181, 2007.
- [55] Sheldon Krimsky. Risk communication in the internet age: The rise of disorganized skepticism. *Environmental Hazards*, 7(2):157–164, 2007.
- [56] Christopher D Wirz, Michael A Xenos, Dominique Brossard, Dietram Scheufele, Jennifer H Chung, and Luisa Massarani. Rethinking social amplification of risk: Social media and Zika in three languages. *Risk Analysis*, 38(12):2599–2624, 2018.
- [57] Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, SM Murshid Hasan, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, et al. COVID-19-related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4):1621–1629, 2020.
- [58] King-wa Fu and Yuner Zhu. Did the world overlook the media’s early warning of COVID-19? *Journal of Risk Research*, 23(7-8):1047–1051, 2020.

- [59] Rasmus Kleis Nielsen, Richard Fletcher, Nic Newman, Scott J Brennen, and Philip N Howard. Navigating the ‘infodemic’: How people in six countries access and rate news and information about coronavirus. Reuters Institute, 2020.
- [60] Rasmus Kleis Nielsen, Richard Fletcher, Antonis Kalogeropoulos, and Felix Simon. Communications in the coronavirus crisis: Lessons for the second wave. Reuters Institute for the Study of Journalism, 2020.
- [61] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The covid-19 social media infodemic. Scientific reports, 10(1):1–10, 2020.
- [62] Jeanne X Kasperson, Roger E Kasperson, Nick Pidgeon, and Paul Slovic. The social amplification of risk: Assessing 15 years of research and theory. In Social Contours of Risk, pages 217–245. Routledge, 2012.
- [63] Reiner Grundmann. Using large text news archives for the analysis of climate change discourse: Some methodological observations. Journal of Risk Research, pages 1–12, 2021.
- [64] Jens O Zinn and Marcus Müller. Understanding discourse and language of risk. Journal of Risk Research, pages 1–14, 2021.
- [65] Dražena Gašpar and Mirela Mabić. Social media: Towards open government. Economics of Digital Transformation, page 353, 2019.
- [66] Carina Jacobi, Wouter Van Atteveldt, and Kasper Welbers. Quantitative analysis of large amounts of journalistic texts using topic modelling. Digital Journalism, 4(1):89–106, 2016.
- [67] Valerie Hase, Daniela Mahl, Mike S Schäfer, and Tobias R Keller. Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018). Global Environmental Change, 70:102353, 2021.
- [68] Martin Rooke and Adam Burgess. Mapping and characterising changes to risk amplification within the British Press: 1985–2017. Journal of Risk Research, pages 1–14, 2021.
- [69] Jens O Zinn. The proliferation of ‘at risk’ in The Times: A corpus approach

- to historical social change, 1785-2009. *Historical Social Research/Historische Sozialforschung*, 43(2 (164)):313–364, 2018.
- [70] Wallace Chipidza, Elmira Akbaripourdibazar, Tendai Gwanzura, and Nicole M Gatto. Topic analysis of traditional and social media news coverage of the early COVID-19 pandemic and implications for public health communication. *Disaster Medicine and Public Health Preparedness*, pages 1–8, 2021.
- [71] Kai Erikson. *A New Species of Trouble: Explorations in Disaster, Trauma, and Community*. WW Norton & Company: New York, 1994.
- [72] Ann Bostrom, Gisela Böhm, Robert E O'Connor, Daniel Hanss, Otto Bodí-Fernandez, and Pradipta Halder. Comparative risk science for the coronavirus pandemic. *Journal of Risk Research*, 23(7-8):902–911, 2020.
- [73] Sharon M Friedman and Brenda P Egolf. A longitudinal study of newspaper and wire service coverage of nanotechnology risks. *Risk Analysis*, 31(11):1701–1717, 2011.
- [74] Kyle Hunt, Puneet Agarwal, and Jun Zhuang. Monitoring misinformation on Twitter during crisis events: A machine learning approach. *Risk Analysis*, 2020.
- [75] Charu C Aggarwal and ChengXiang Zhai. *Mining Text Data*. Springer Science & Business: New Jersey, 2012.
- [76] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.
- [77] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [78] Andrew Kachites McCallum. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>, 2002.
- [79] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [80] Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 222:248, 2015.
- [81] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems*, pages 288–296, 2009.

- [82] Deana Grobe. Media coverage of Y2K: Content analysis from a single newspaper, 1997 to 2000. Psychological reports, 90(2):412–414, 2002.
- [83] United Nations. World Economic Situation And Prospects: September 2019 Briefing. United Nations Department of Economic and Social Affairs: New York, 2019.
- [84] M-Elisabeth Paté-Cornell, Marshall Kuypers, Matthew Smith, and Philip Keller. Cyber risk management for critical infrastructure: A risk analysis model and three case studies. Risk Analysis, 38(2):226–241, 2018.
- [85] Nina Lakhani and Jessica Glenza. ‘menu of neglect’: The long-term health problems being ignored in US amid pandemic. The Guardian, 2021.

Table 1: Risk-related news dataset description

Data source		
LexisNexis: Major US Publications (print and online) of 17 notable newspapers, which make articles available in English, including: New York Times, Los Angeles Times, Daily News (New York), USA Today, The Baltimore Sun, The Philadelphia Enquirer, The Philadelphia Daily News, Tampa Bay Times, St-Louis Post -Dispatch, Pittsburgh Post Gazette, The Atlanta Journal-Constitution, The Jerusalem Post (Intl Ed), The Daily Oklahoman, The Tampa Tribune, Star Tribune, The Christian Science Monitor, The Hartford Courant.		
Initial article identification parameters		
<i>Society for Risk Analysis 2018 keywords (selection):</i> risk; uncertainty; harm; damage; hazard; safe; safety; security; secure; threat; resilience; vulnerability.		
Article search refinement		
1. At least one mention of ‘risk’ in the article, and; 2. Must contain one additional Society for Risk Analysis 2018 keyword		
Other search information		
<i>Date range:</i> Jan 2000 - July 2021 <i>Language:</i> English language only		
Descriptive statistics		
Time period	Annual average news (All)	Annual average news (Refined)
2000-2005	12,951	12,377
2006-2010	11,224	10,766
2011-2015	12,247	11,785
2016-2021	16,019	15,426
Total news articles	283,165	271,854

Table 2: Categories of US risk news (2000-2021)

	Business	Domestic affairs	Entertainment	Environment	Geopolitics	Health	Technology
<i>Panel A: Top contributing topic terms towards each category</i>							
1	investment	home	life	water	war	disease	technology
2	investor	law	player	climate_change	force	doctor	datum
3	business	charge	injury	region	attack	health	information
4	credit	lawyer	game	plant	troop	product	system
5	money	lawsuit	sport	damage	iraq	patient	internet
6	cost	prison	season	florida	isreal	woman	network
7	bank	issue	team	storm	soldier	virus	computer
8	firm	court	league	home	security	vaccine	access
9	loan	judge	fan	disaster	army	coronavirus	security
10	sale	attorney	field	area	afghanistan	disease_control	service
11	deal	trail	book	site	agreement	infection	damage
12	stock	jail	show	chemical	russia	mask	ship
13	economy	election	file	level	syria	prevention	mission
14	market	party	love	exposure	iran	study	space
15	growth	vote	man	air	region	drug	line
%	12.53%	31.51%	13.58%	5.93%	13.46%	17.42%	5.58%
<i>Panel B: Individual topics (topic numbers) in each category</i>							
Fin markets (4)	Air safety (3)	Human int (1)	Pollution (6)	War (5)	Female health (9)	R&D (2)	
Fin management (12)	Crime (15)	Arts (10)	Nat disasters (8)	Conflict (7)	Food safety (11)	Cyber sec (27)	
Banking (14)	Locality (19)	Sports (16)		European aff (23)	Illness (13)	Tech failure (28)	
Job security (18)	Compliance (20)			Trade (30)	Medicine (17)		
	Child welfare (21)				Infect disease (32)		
	Education (22)						
	Road safety (24)						
	Party politics (25)						
	Policing (26)						
	Regulation (29)						
	Elections (31)						

Table presents the seven second order categories of US risk news constructed from 32 individual first order topics. The individual topics are reported in the Appendix, Tables A1 and A2. Reported for each category topic is the label and two panels. Panel A reports the top 15 terms which most contribute to topics in that category, and the percentage of articles in the full dataset of 271,854 articles where the primary topic of the article belongs to that category topic. Panel B reports the contributing individual topics to each category.



Figure 1: Word cloud of terms from all risk-related news articles (2000-2021)

45

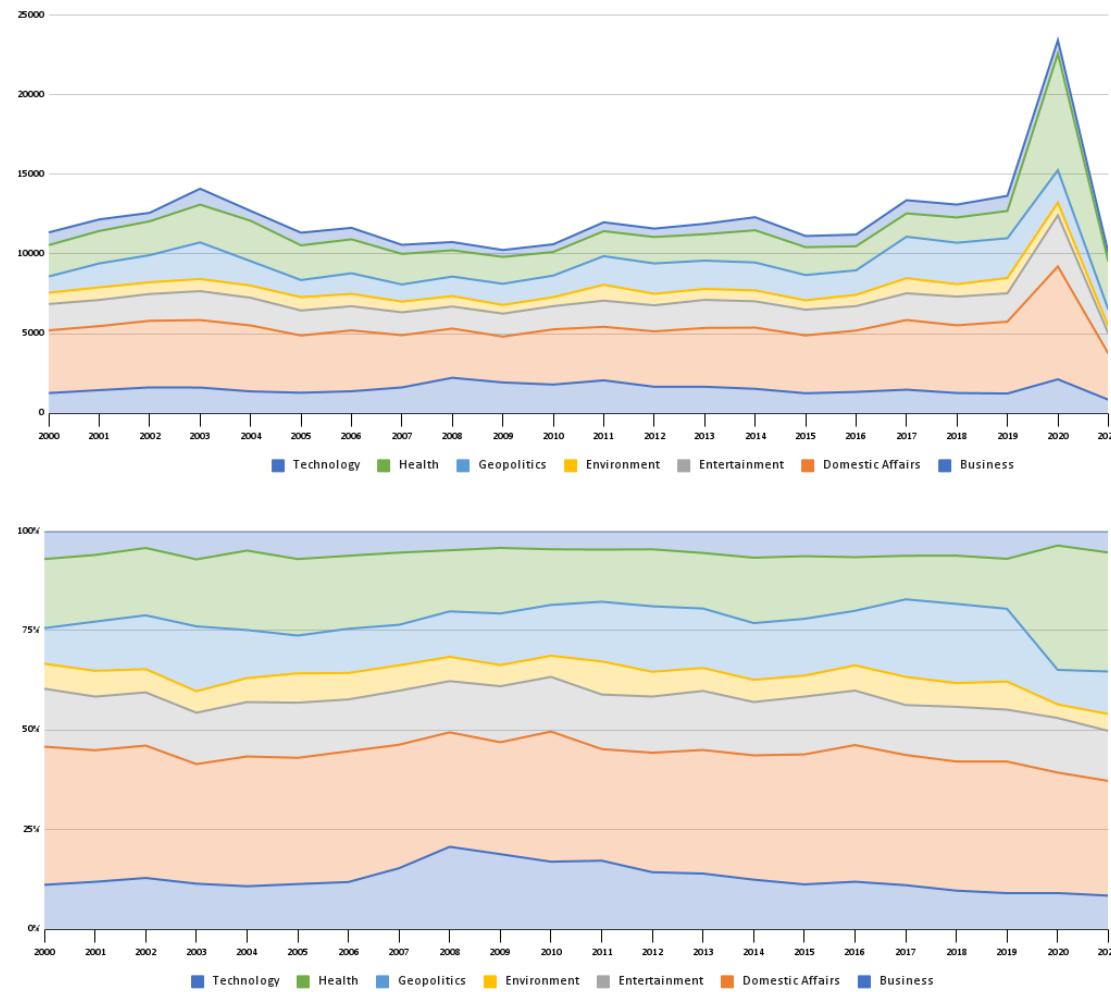


Figure 2: Distribution of category topics of US risk communication over time. Top figure: article count per year. Bottom figure: article count per year normalised by total articles per year.

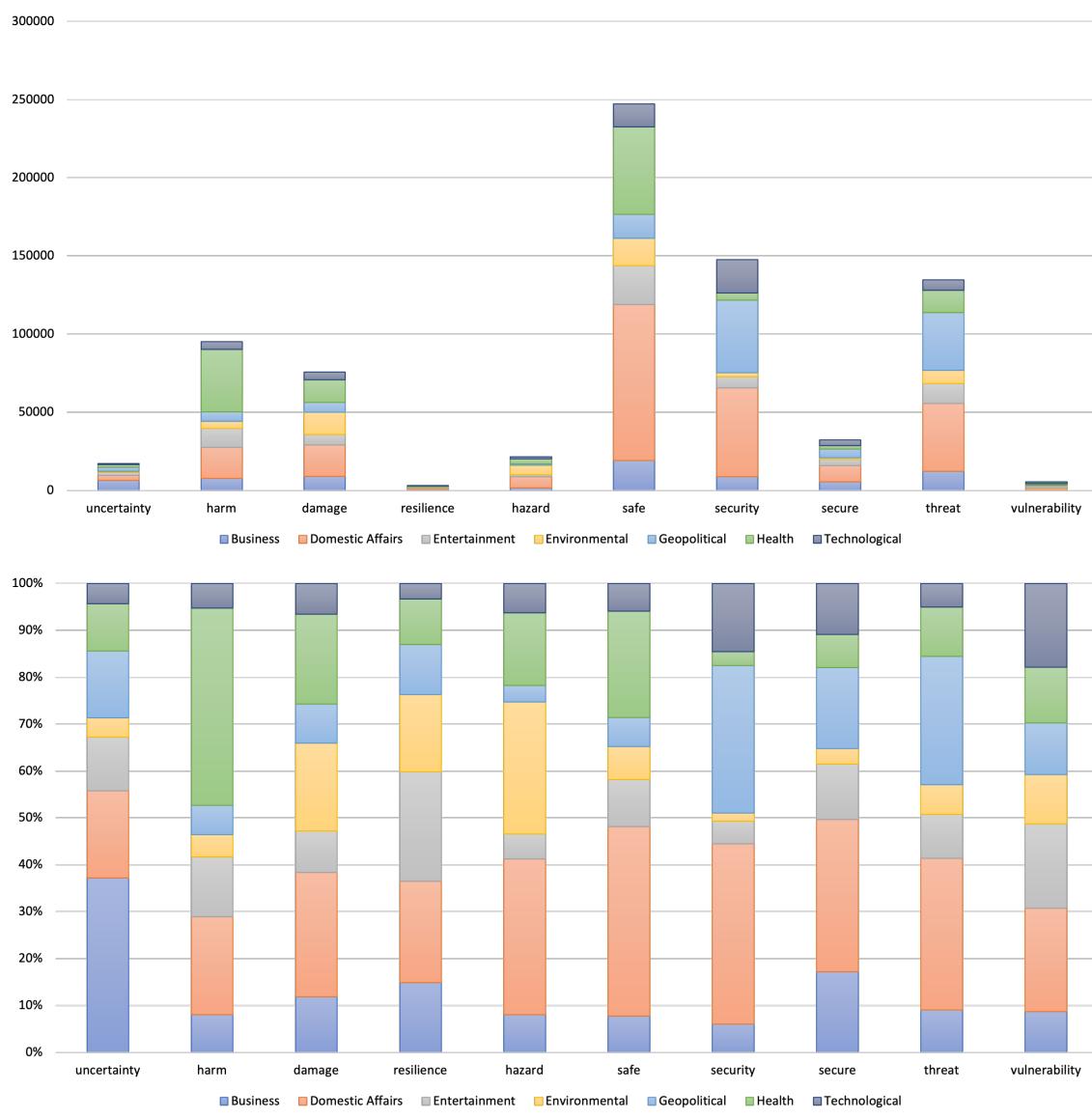


Figure 3: SRA 2018 (reduced) glossary terms propensity. Top figure is a stacked count of term occurrences. Bottom figure is a 100% stacked chart. The term *risk* is removed from these charts, as it is required to be present in all articles as per the search criteria.

Figure 4: Word clouds of news reporting of risk per category.

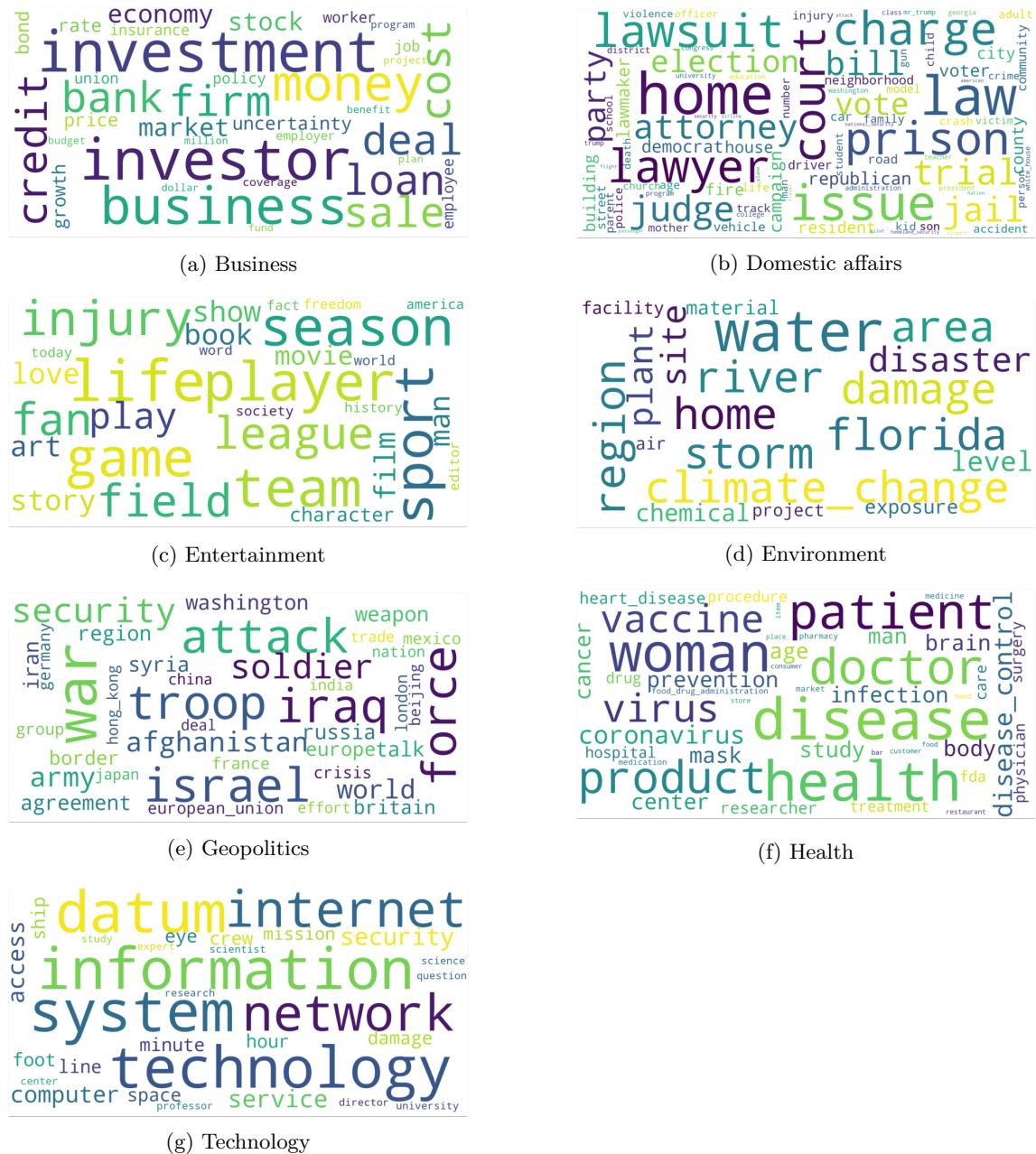
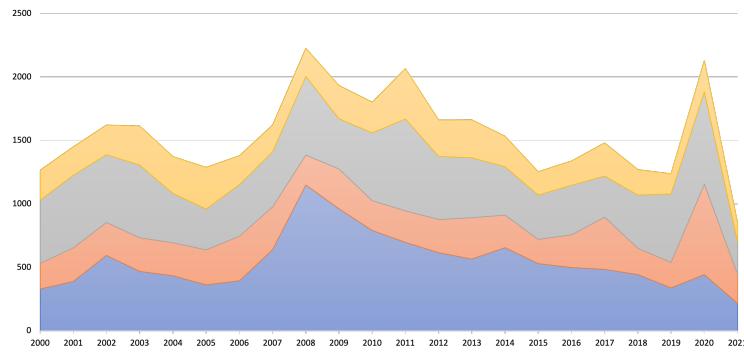
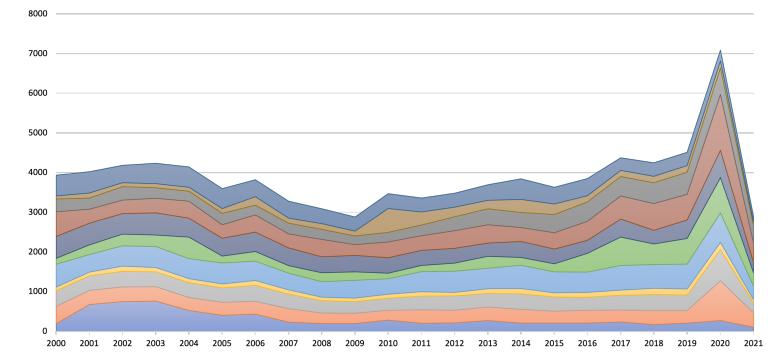


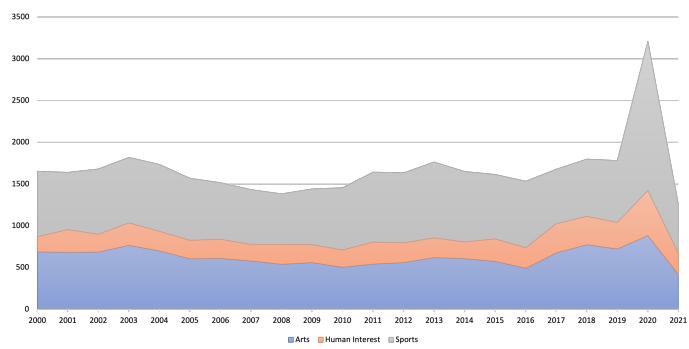
Figure 5: Category and sub-topic timelines (2000-2021) of news reporting of risk.



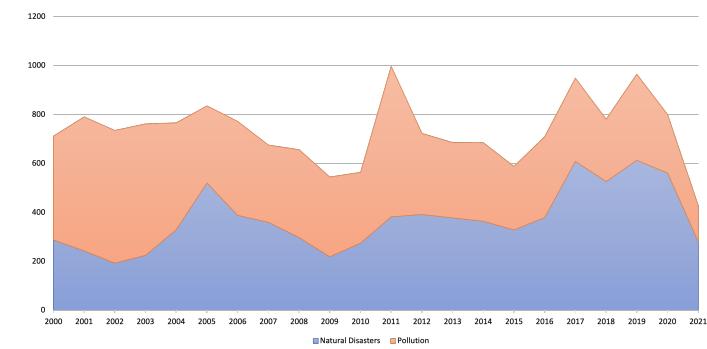
(a) Business



(b) Domestic affairs

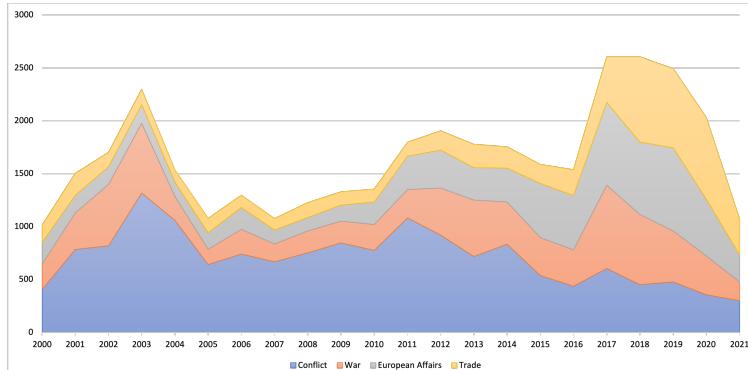


(c) Entertainment

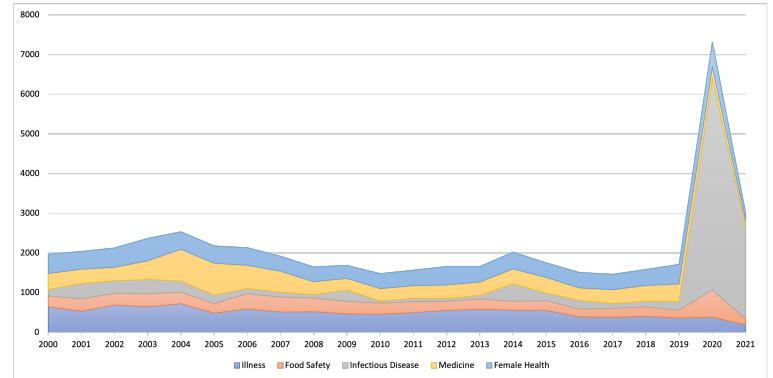


(d) Environment

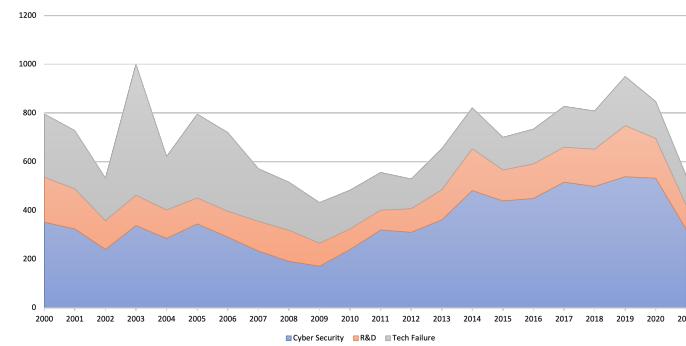
Figure 5: (Continued) Category and sub-topic timelines (2000-2021) of news reporting of risk.



(a) Geopolitics



(b) Health



(c) Technology



Figure 6: Word cloud of the most popular terms in all 2020 news media reporting of risk.

Appendix

Table A1: Topics of US risk communication (2000-2021)

	1: Human interest	2: R&D	3: Air safety	4: Financial markets	6: War	6: Pollution
1	world 0.018	university 0.030	security 0.034	economy 0.032	war 0.019	plant 0.022
2	life 0.016	research 0.020	attack 0.015	market 0.024	iraq 0.016	site 0.018
3	fact 0.015	study 0.018	airport 0.015	investor 0.015	attack 0.015	level 0.017
4	america 0.014	professor 0.016	airline 0.013	stock 0.015	israel 0.013	water 0.013
5	today 0.011	center 0.014	plane 0.012	uncertainty 0.015	force 0.011	chemical 0.012
6	freedom 0.011	scientist 0.014	flight 0.012	price 0.014	troop 0.010	exposure 0.009
7	word 0.011	expert 0.014	passenger 0.011	growth 0.012	security 0.009	material 0.009
8	society 0.010	science 0.011	travel 0.010	rate 0.012	soldier 0.009	facility 0.009
9	history 0.009	question 0.011	homeland_sec 0.009	bond 0.012	afghanistan 0.008	air 0.008
10	editor 0.009	director 0.011	pilot 0.009	investment 0.011	army 0.007	project 0.008
%	2.11%	1.05%	2.56%	3.81%	5.71%	2.93%
	7: Conflict	8: Natural disasters	9. Female health	10: Arts	11: Food safety	12: Financial management
1	iran 0.020	damage 0.024	woman 0.046	story 0.017	food 0.025	program 0.049
2	russia 0.016	area 0.023	patient 0.039	show 0.013	product 0.017	money 0.032
3	war 0.016	water 0.022	hospital 0.038	book 0.013	store 0.016	plan 0.028
4	world 0.013	home 0.020	doctor 0.028	life 0.009	restaurant 0.011	cost 0.017
5	weapon 0.011	storm 0.013	surgery 0.012	love 0.008	consumer 0.010	budget 0.012
6	washington 0.010	clim_change 0.012	procedure 0.012	film 0.008	customer 0.008	dollar 0.012
7	region 0.010	florida 0.011	treatment 0.011	man 0.007	place 0.007	fund 0.011
8	syria 0.009	region 0.009	care 0.011	art 0.007	item 0.006	project 0.011
9	agreement 0.007	river 0.008	health 0.010	movie 0.007	hand 0.006	benefit 0.010
10	talk 0.007	disaster 0.008	physician 0.009	character 0.006	bar 0.006	million 0.009
%	2.97%	3.00%	3.43%	5.07%	2.33%	1.98%
	13: Illness	14: Banking	15: Crime	16: Sports	17: Medicine	18: Job security
1	study 0.024	bank 0.042	court 0.028	team 0.035	drug 0.056	worker 0.038
2	cancer 0.016	business 0.019	law 0.025	game 0.030	patient 0.019	employee 0.037
3	woman 0.014	firm 0.013	lawyer 0.017	player 0.027	product 0.017	job 0.036
4	disease 0.014	loan 0.013	prison 0.015	season 0.020	fda 0.015	business 0.031
5	age 0.011	deal 0.013	lawsuit 0.014	sport 0.011	food_drug_ad 0.014	policy 0.022
6	man 0.010	money 0.011	charge 0.013	injury 0.010	market 0.012	cost 0.015
7	brain 0.009	investor 0.009	judge 0.013	league 0.009	medicine 0.011	insurance 0.014
8	body 0.008	sale 0.008	trial 0.011	field 0.009	medication 0.010	union 0.012
9	heart_disease 0.007	credit 0.008	attorney 0.010	fan 0.008	pharmacy 0.010	employer 0.011
10	researcher 0.007	investment 0.008	jail 0.010	play 0.008	doctor 0.009	coverage 0.011
%	4.10%	4.41%	4.19%	6.40%	3.26%	2.32%

Table presents the topics identified by the LDA topic modeling of risk-related news articles. See Section 3.2 for implementation details. The initial topic model reported 35 topics, but 3 topics were removed as non-risk topics. The table, therefore, reports the final 32 topics in the order in which the topic model reported them. The table reports the top 10 terms per topic, and their beta values - the statistical contribution of the term to the topic. The last row is the percentage of articles in the full dataset of 271,854 articles where the primary topic of the article belongs to that topic. Note: to allow the table to fit to page the following terms were shortened: Topic 3, Term 9 'homeland_security' to 'homeland_sec'; Topic 8, Term 6 'climate_change' to 'clim_change'; Topic 17, Term 5 'food_drug_administration' to 'food_drug_ad'.

Table A2: Topics of US risk communication (2000-2021) (contd)

19: Locality		20: Compliance		21: Child welfare		22: Education		23: European affairs		24: Road safety	
1 city	0.085	report	0.038	child	0.101	school	0.069	europe	0.022	car	0.033
2 building	0.024	department	0.030	family	0.062	student	0.046	britain	0.013	driver	0.023
3 home	0.022	agency	0.029	parent	0.037	district	0.021	border	0.013	road	0.022
4 fire	0.022	investigation	0.020	kid	0.022	university	0.019	european_un	0.011	vehicle	0.020
5 resident	0.021	office	0.017	home	0.022	program	0.017	france	0.010	accident	0.012
6 county	0.020	letter	0.014	mother	0.016	education	0.015	mexico	0.010	track	0.009
7 house	0.018	issue	0.012	life	0.014	georgia	0.014	london	0.009	crash	0.009
8 neighborhood	0.015	problem	0.012	adult	0.014	teacher	0.014	germany	0.008	injury	0.008
9 community	0.012	information	0.012	age	0.013	college	0.013	group	0.008	number	0.008
10 street	0.012	review	0.012	son	0.013	class	0.012	crisis	0.008	model	0.008
%	3.57%		0.91%		3.04%		2.92%		2.65%		3.26%
25: Party politics		26: Policing		27: Cyber security		28: Tech failure		29: Regulation		30: Trade	
1 bill	0.029	police	0.034	information	0.022	foot	0.011	industry	0.027	china	0.043
2 election	0.021	officer	0.028	system	0.022	damage	0.011	rule	0.025	world	0.034
3 vote	0.020	man	0.027	technology	0.018	space	0.011	regulation	0.016	japan	0.010
4 party	0.017	gun	0.015	datum	0.015	eye	0.010	oil	0.013	india	0.009
5 republican	0.015	violence	0.015	internet	0.013	ship	0.009	standard	0.011	hong_kong	0.008
6 democrat	0.014	victim	0.014	network	0.012	hour	0.008	change	0.011	effort	0.008
7 issue	0.013	death	0.012	security	0.012	line	0.008	requirement	0.009	trade	0.008
8 lawmaker	0.013	crime	0.011	service	0.012	crew	0.008	process	0.008	beijing	0.008
9 voter	0.012	church	0.011	computer	0.011	minute	0.007	law	0.008	deal	0.007
10 campaign	0.012	person	0.010	access	0.010	mission	0.007	pennsylvania	0.007	nation	0.007
%	3.91%		2.96%		2.86%		1.66%		1.51%		2.12%
31: Elections		32: Infectious disease									
1 president	0.061	virus	0.034								
2 administration	0.025	vaccine	0.029								
3 white_house	0.023	coronavirus	0.016								
4 trump	0.020	health	0.014								
5 washington	0.020	disease_control	0.011								
6 congress	0.017	disease	0.010								
7 mr_trump	0.016	prevention	0.010								
8 national_sec	0.014	infection	0.010								
9 nation	0.013	center	0.009								
10 american	0.013	mask	0.008								
%	2.69%		4.30%								

Table presents the topics identified by the LDA topic modeling of risk-related news articles. See Section 3.2 for implementation details. The initial topic model reported 35 topics, but 3 topics were removed as non-risk topics. The table, therefore, reports the final 32 topics in the order in which the topic model reported them. The table reports the top 10 terms per topic, and their beta values - the statistical contribution of the term to the topic. The last row is the percentage of articles in the full dataset of 271,854 articles where the primary topic of the article belongs to that topic. Note: to allow the table to fit to page the following terms were shortened: Topic 33, Term 4 'european_union' to 'european_un'; Topic 31, Term 8 'natural_security' to 'natural_sec'.