COVERAGE OF EPIDEMICS IN AMERICAN NEWSPAPERS THROUGH THE LENS OF THE CRISIS AND EMERGENCY RISK COMMUNICATION FRAMEWORK

Yotam Ophir

The Centers for Disease Control and Prevention (CDC) Crisis and Emergency Risk Communication (CERC) framework has been used by the organization during recent outbreaks of infectious diseases. However, the dissemination of the organization's crisis messages depends largely on mass media coverage. This study analyzed 5,006 articles from leading American newspapers covering 3 epidemics: H1N1, Ebola, and Zika. Using a mixed method of automated and manual content analysis, it identified 3 distinct themes used to cover the diseases: pandemic, scientific, and social. Analysis of the themes based on CERC guidelines demonstrated substantial discrepancies between what CDC aims to communicate during epidemics and what the media actually disseminated to the public. Implications for public health organizations and communicators are discussed.

Keywords: Risk communication, Epidemic management/response

LARGE-SCALE OUTBREAKS of infectious diseases may have severe physical, psychological, economic, and social consequences.¹ An essential part of crisis management is the communication of timely and accurate information to the public, aimed to improve the understanding of risks and to inform decision making.² Public health organizations have tried to cope with the unexpected nature of crises by developing emergency communication models.³ The Centers for Disease Control and Prevention's (CDC's) model, the Crisis and Emergency Risk Communication framework,⁴ aims to mitigate risks, in part, by providing information that is expected to increase compliance with and support in the organization's coping efforts.⁵ However, even if the CDC adequately follows CERC guidelines—and even if they are indeed effective, as its authors argue—the successful implementation of the framework depends to a large degree on the compliance of the mass media.^{6,7}

This study examined American newspapers' coverage of 3 epidemics that threatened the United States in recent

years: H1N1 (2009-10), Ebola (2014-15), and Zika (2015-16). First, a corpus of 5,006 news articles was collected from 4 leading newspapers. An automated technique for text analysis, Structural Topic Modeling (STM), was used to understand the thematic content and structure of the corpus. Then, a community detection technique was employed on a network of the topics to identify common distinct themes. Lastly, human coders manually analyzed the content of articles from each theme, in light of the CERC framework. This method allowed the examination of discrepancies between CERC guidelines and what was disseminated to the public through the media.

CERC FRAMEWORK AND THE MEDIA

The CERC framework integrates elements from risk, crisis, and health communication theories.⁴ It asserts that trust, compliance, and support during crises can be achieved, in

Yotam Ophir, PhD, is a postdoctoral fellow, Annenberg School for Communication, University of Pennsylvania, Philadelphia.

part, by providing the public with the information it needs to make sense of the crisis.^{4,8} The CERC guidelines recommend different communications at different stages of the crisis; this study focuses on the initial and maintenance phases,⁶ as the pre-crisis, resolution, and evaluation stages focus more on internal procedures and preparations for potential crises. It also focuses on information types and not on message characteristics (eg, the use of empathy). According to the CERC lifecycle model, during an outbreak, CDC should provide information about risks and responses. Risk information aims to establish an understanding of the crisis circumstances, consequences, and anticipated health and social outcomes.^{6,9} Response information focuses on specific actions by health organizations and the public that can mitigate threats.¹⁰

Despite CDC's increasing use of direct communication channels, such as social media,¹¹ the organization still relies on mainstream mass media to disseminate its messages.⁵ The CERC framework provides communicators with guidelines for cooperation with journalists and recognizes the importance of meeting the needs of the media.⁶ Studies have shown that, despite an increase in alternative sources, mainstream outlets remain the primary source of news information for most audiences.¹² When looking for health information, people often look to the news rather than medical sources.¹³ Consequently, news media still have a significant effect on the public's agenda.^{14,15} However, during crises, the news media may not provide audiences with sufficient exposure to the organization's messages.¹⁶ Moreover, the information that will reach audiences is likely to be processed, edited, and changed by journalists.^{17,18} Thus, audiences may not be exposed to an adequate quantity and quality of information that will allow the mitigation of risks.

The construction of news involves a process of selection, where some issues, events, and sources get more media attention than others, as they better fit journalists' needs.¹⁹⁻²¹ By emphasizing different aspects of outbreaks, the media construct epidemics in various ways that may have different effects on public perceptions and behaviors.^{22,23} Two main approaches, general and specific, were taken by researchers who analyzed media portrayals of crises and epidemics. The general approach employs common, broad, empirically supported social constructions, such as the classic episodic and thematic frames analysis²⁴ in the context of infectious diseases.²⁵ Scholars working in the specific approach, on the other hand, develop theoretically driven categories that represent the unique characteristics of crises. For example, Shih and colleagues suggested a typology of consequence, uncertainty, action, reassurance, conflict, and new evidence.²⁶ However, as both approaches rely on predetermined categories to analyze new data, scholars have argued that researchers may be prone to consciously or unconsciously find what they were looking for in the data.²⁷

Therefore, in some studies researchers extracted coverage categories from a specific corpus dedicated to 1 disease,

such as the construction of SARS in specific Asian countries or the construction of HIV.^{28,29} This approach may result in insightful findings, but generalizability to other diseases is questionable. To address these issues in this work, we have built a corpus consisting of the coverage of more than 1 disease and further reduced researcher biases by harnessing computational tools for the identification of consistent coverage categories. To test whether the categories are distinct in terms of their use of crisis information, manual content analysis was conducted using a codebook built around the CERC framework.

This study examined the content of news articles on 2 levels: topics and clusters of topics (themes). A topic is a subject that occurs with some degree of frequency throughout and across a corpus.³⁰ A common approach for extracting thematic substance from complex data is the use of co-occurrences. This method identifies and clusters words that tend to appear in the same documents, while distancing them from words that tend not to appear in the same documents.^{31,32} The underlying assumption of co-occurrence is that words that appear together often in the same documents also share thematic meaning.³³ Identifying co-occurrences in large corpora is resource intensive, and human coders are not well suited for identifying textual patterns that manifest across texts, contexts, and time periods. Therefore, researchers have recently employed automated analysis tools to identify patterns of co-occurrences in complex data sets.^{34,35} To begin the analysis of media coverage, we asked:

RQ1: What topics were prominent in American newspapers' coverage of epidemics?

Topics in this study are extracted from co-occurrences in articles, and news articles tend to focus on 1 disease at a time. As a result, some topics are prone to be disease-specific. For example, the topic of the Olympic Games was prevalent during the Zika virus epidemic of 2015-16, but the topic is not expected to be relevant to the coverage of Ebola or H1N1. Similarly, the topic of Ebola patients hospitalized in the United States is mostly irrelevant to other diseases (albeit future epidemics may be discussed using comparisons to previous cases). Therefore, a second level of clustering was performed to identify topics that share similar words. These clusters of topics that share similar vocabularies were labeled "themes." We explored these questions:

RQ2a: What themes were prominent in American newspapers' coverage of the epidemics?

RQ2b: What were the similarities and differences in the use of themes between different epidemics?

In order to examine whether the themes extracted with our method varied in terms of the prevalence of crisis information, a manual content analysis was conducted on representative articles from each theme. According to the CERC guidelines, communications during an epidemic should focus on health and social risks and on steps that can be taken by citizens and health organizations to mitigate them.^{4,36,37} The next research question asked:

RQ3a: Did the prevalence of crisis information recommended in the CERC framework (medical/health disruptions, social/economic disruptions, individual response, and organizational response) differ by themes?

In order to be able to reject the alternative explanation that differences were not the result of themes but of coverage of different diseases, we asked:

RQ3b: Did the prevalence of crisis information recommended in the CERC framework (medical/health disruptions, social/economic disruptions, individual response, and organizational response) differ across diseases?

Finally, crisis communications are time-dependent and time-sensitive.³⁸ The CERC framework emphasizes that crises are dynamic and that crisis communications should adapt according to crisis phases and developments in knowledge regarding risks and behavioral recommendations.⁶ Although a detailed analysis of specific changes over time and with respect to real-world events during each disease is beyond the scope of this study, the fourth research question asked:

RQ4: How did the use of themes change over the timeline of each disease?

METHOD

Automated Content Analysis

A total of 5,006 news articles were collected from Lexis Nexis (the *New York Times*, the *Washington Post*, and *USA Today*) and Factivia (the *Wall Street Journal*) using the search terms “swine flu,” “H1N1,” “Ebola,” and “Zika.” Outlets were selected based on their centrality and position as leading prestigious newspapers that can serve as a proxy for the American news environment.³⁹

For each disease, data were collected from the first mention of the disease in the sources. The H1N1 corpus consisted of 1,798 articles, spanning between January 9, 2009, and September 8, 2010, a month after WHO declared the epidemic was over. The Ebola corpus consisted of 2,113 articles between March 20, 2014, and July 16, 2015, a month after WHO declared Liberia to be Ebola-free. The Zika corpus consisted of 1,095 articles, between December 29, 2015, and December 7, 2016, a month after WHO

declared that Zika was no longer a global threat. After preprocessing (removing punctuation, numbers, and stop-words and using a sparsity level of 0.99 to remove uncommon words*), the final vocabulary consisted of 4,538 unique words (tokens).

Structural topic modeling (STM) was used to estimate a model in R.⁴⁰ † Topic modeling is a semi-automated, unsupervised method that uses a Bayesian generative approach for text analysis, in which the model “mimics” the writing process of a given corpus of documents. The approach assumes that topics are specified before the data were generated and that authors have used combinations of topics in each article.⁴¹ In other words, the model uses an observed set of documents to infer the latent topic structure that could generate it. “Topics” are sets of frequency distributions of words, based on co-occurrences. In topic models, every word in the corpus has a probability to appear in each topic, and every document is a mixture of all topics.³⁴ Topic modeling is a “bag-of-words” approach, meaning that narrative, location in the text, and syntax are not taken into consideration.⁴² Topic modeling was found to be useful and efficient for analyzing corpora in the social sciences, including news articles.⁴³ Structural topic modeling allows documents to vary in the distribution of words and topics based on covariances.⁴⁰ The current study uses the covariances of date of publication and news outlet. In order to determine the number of topics, models with different numbers of topics were compared in terms of coherency, exclusivity, residuals, and held-out likelihood for different numbers of topic (using the “searchk” function in the structural topic modeling package in R). Based on these criteria, a model of 35 topics was selected, compromising between coherence and exclusivity,[‡] while providing a relatively high held-out likelihood.⁴⁰ Labels were given to the topics by the researcher based on a hermeneutic qualitative procedure, after reading top words in each topic (highest probability to be included in the topic), top FREX words (top exclusive words for each topic, see Roberts et al.⁴⁰), and the 30 articles that were most representative of each topic.

Next, in order to cluster topics based on shared vocabulary, a network of topics was created, treating individual

*Sparsity is used to reduce the number of words in the final vocabulary used for the model by removing words that appear infrequently. Sparsity refers to the threshold of relative document frequency for a term, calculated as $N \times (1 - \text{Sparsity level})$, where N is the number of documents. Sparse=0.99 will remove only terms that are sparser than 0.99. In the current study, all terms that appear in more than $5006 \times (1 - 0.99)$ documents will be retained; others will be omitted. In the current case, words that appear in more than 50 documents will be retained in the corpus. Near the other extreme of sparsity=0.1, only terms that appear in nearly every document will be retained.

†<https://www.r-project.org/about.html>.

‡The semantic coherence criterion was developed by Mimno and colleagues⁴⁴ and is closely related to pointwise mutual information. It is maximized when the most probable words in a given topic frequently co-occur together. Exclusivity is measured through Roberts et al’s (2014) FREX criterion,⁴⁰ a weighted harmonic mean of the word’s rank in terms of exclusivity and frequency. Exclusivity of a model will be higher when top frequent words are less shared between topics. There is a trade-off between coherence and exclusivity. Models with more topics tend to have higher exclusivity but also lower semantic coherence. For more information on the held-out likelihood criterion, see Wallach et al,⁴⁵ and for residual analysis, see Taddy.⁴⁶

topics as nodes. First, a matrix was created, representing the probability distribution of each word conditional on each topic (in topic modeling, every word has a probability to appear in every topic, but the probability changes from topic to topic). Then, a Pearson correlation matrix was calculated to evaluate the edge weight or “distance” between each pair of topics. One topic was omitted from analysis as it consisted of linguistic artifacts that could not be interpreted as part of the epidemics phenomenon (see “irrelevant” topic in Table 2). Next, a weighted (based on Pearson correlation), undirected network was built based on the correlation matrix. Since in topic modeling each word appears in each topic, the resulting network was fully connected, as all topics shared the same vocabularies. To reduce the number of edges and retain only substantial connections, a backbone filtering method was used.⁴⁷ A sensitivity analysis was conducted using different community-detection techniques, comparing the methods of leading eigenvectors⁴⁸ and Louvain⁴⁹ with varied levels of filtering. Clusters were virtually the same, and the final model used a Louvain approach with a backbone Alpha of .15 (the maximum point of reduction, after which complete topics were omitted from the model). The results indicated that there were 3 communities in the network, with a modularity score of .56. The method automatically indicates which topics belong to which community.

Manual Content Analysis

To understand the content of each theme, a sub-corpus for coding was created, consisting of the articles that are most representative for each theme in each disease. For each article, the model provided Theta values, representing the per-document probabilities of each topic (Theta values for all topics always sum up to 1 for each article). The calculation of representativeness was done by summing up the Theta values of all the topics included in that cluster. For example, an article with a Theta value of 99.9 for cluster A is an article in which 99.9% of its words belong to topics from cluster A. Specifically, the sub-corpus included 126 articles, with 42 articles per disease (14 articles from each cluster). To improve comparability, articles shorter than 100 words or longer than 2,000 words were omitted and replaced by the next most representative ones. A codebook was created based on the model of effects described in the CERC framework (see Measures section).

Three human coders were randomly assigned 84 articles each. Coders were first trained through several coding sessions, done on a separate set of articles (representative of the clusters in each disease, but not the top, most representative ones). The codebook was refined and improved during training sessions. After reaching a satisfactory reliability on all key variables (Krippendorff's Alpha > .80), the coders used the refined codebook to analyze the target corpus of

126 articles. Each article in the corpus was coded by 2 coders.

Measures

The codebook was built around audiences' information needs described in the CERC guidelines: medical/health disruptions, social/economic disruptions, individual response, and organizational response.⁶ Since the aim of this study is to identify broad categories of coverage and not the prevalence of specific information (eg, health effects in general rather than specific symptoms), the analysis focuses on the centrality of information components, as described below.

Medical/health disruptions information captured the existence or absence of information about health effects, tolls, vulnerability, symptoms, connection to previous outbreaks, technical information, and use of narratives. After coding for the existence or absence of these components in an article, the coder was asked for the centrality of the information in the article: “You’ve just answered questions about facts and risk information (prognosis, symptoms, biology, history of viruses, risk information, description of cases and transmissions, death tolls, susceptibility, health consequences). How central would you say the topic of facts and risk information is to the article?” Possible answers were:

1. Not central at all—there are no mentions of such information in the article.
2. Slightly central—there are some mentions, but they are not central to the main topic.
3. Central—the information is important to the main topic of the article, but there are other kinds of information that are also important.
4. Very central—the information is the main topic of the article.

Individual response information was measured using the following item: “According to the article, what can individuals (citizens, not doctors, etc.) do to avoid infection or ameliorate risks? (check all that apply)” An example of an answer is “get vaccinated.” After coding for these components, the coder was asked: “You’ve just answered questions about decision making and efficacy information (information that directs readers to take an appropriate action to reduce or eliminate risks. These DO NOT INCLUDE actions by official organizations, CDC, government etc.). How central would you say the topic of decision making and efficacy information was to the article?” The same possible answers were used here as for medical/health disruptions.

Organizational response information was measured using the following item: “Which steps taken by countries, organizations, and institutions to reduce risks and avoid infections are mentioned in the article? (check all that apply)”

An example of an answer is “improving safety in laboratories.” After coding the components, the coder was asked: “You’ve just answered questions about resource allocation information (information that explains to readers what was done by official governmental or nongovernmental organizations to reduce or eliminate risks. These DO NOT INCLUDE actions by unofficial individuals, citizens, etc.). How central would you say the topic of resource allocation information was to the article?” The same possible answers were used here as for the other dimensions.

Social/economic disruptions information was measured using the following item: “Does the article mention the impact of the disease on the following? (check all that apply)” An example for an answer is “sports events or athletes.” After coding the component, the coder was asked: “You’ve just answered questions about normalcy and well-being information (information about the non-health consequences of the disease, such as impact on cultural events, politicians, economies, etc.). How central would you say the topic of normalcy and well-being information was to the article?” The same possible answers were used here as for the other dimensions.

RESULTS

The corpus contained 5,006 articles from 4 newspapers and about 3 diseases. The amount of coverage by each outlet for each disease can be seen in Table 1.

RQ1 examined the topics that were prominent in the American newspapers’ coverage of the epidemics. The model produced 35 topics. Table 2 presents the prevalence of each topic in the corpus, its label (assigned by the researcher), the most common words in the topic, and the top FREX words.

As expected, some topics were specific to 1 disease, such as the “Ebola toll” topic, describing the growing number of infections and fatalities in Africa; the “H1N1 vaccine” topic about the efforts to create a specific vaccine against the H1N1 epidemic; and the “microcephaly” topic, which focused on the effects of Zika virus on fetuses, resulting in abnormally small heads. The model estimated that a substantial percentage, about 13.2%, of the language in Ebola articles came from the “Ebola toll” topic. However, the topic was not used at all in articles about Zika and H1N1. Similar patterns of dominance in 1 disease and absence in others could be seen for the “H1N1 vaccine”

(10.9% of H1N1 coverage) and “microcephaly” (16.6% of Zika coverage) topics. In contrast to these (and other) specific topics, some topics were broader and more general and were used for all 3 diseases. For example, the topic about scientific research dominated 1.8% of Ebola coverage, 2.0% of H1N1, and 2.4% of Zika. The topic labeled “disaster response,” focusing on government efforts to form response teams and to improve preparedness to disaster situations, was used in the coverage of Ebola (5.3%), H1N1 (3.1%), and Zika (3.2%).

While topics could be general or specific to a disease, themes attempted to capture broader thematic substance used across epidemics. For that purpose, RQ2a examined what themes were prominent in the coverage. As explained above, themes were operationalized as clusters of topics, found through a Louvain community-detection method. The 3 themes can be seen in Figure 1, using the labels from Table 2. For convenience, each theme was assigned a letter, as detailed in Table 3.

RQ2b examined the differences between diseases in the prevalence of themes. Prevalence was calculated using the Theta object derived from the topic model. Theta is the per-document probability of a topic, and it “has the posterior probability of a topic given a document that this function uses” (<https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf>). The diseases were relatively similar on the existence of theme C, its relative percentage ranging from 44.1% (Ebola) to 45.2% (Zika). More substantial differences were found for the other 2 themes. Theme A was most prominent in H1N1 (42.2%) and Zika (36.9%) coverage, while it was relatively lower in Ebola (21.7%). Theme B was most prominent in Ebola (34.2%) and relatively low in Zika (17.7%) and H1N1 (12.8%) coverage. The prevalence of each theme in each disease can be seen in Table 4.

Next, RQ3a examined differences in prominence of crisis information between themes. The following analysis is structured around the categories derived from the CERC framework.

Medical/health disruption—ANOVA tests found significant differences between themes in terms of the centrality of information about medical/health disruption ($p < .01$). In terms of the evaluation of centrality on a 1 to 4 scale, it was most central to theme A (average centrality for articles from this theme was $M = 3.27$, $SD = .65$), where 78.5% of the articles included information about health effects, 69% information about subgroups’ vulnerability, 57.1% information

Table 1. Number of News Articles per Outlet, by Disease

	<i>Wall Street Journal</i>	<i>Washington Post</i>	<i>New York Times</i>	<i>USA Today</i>	Total
H1N1	399	583	553	263	1,798
Ebola	482	574	810	247	2,113
Zika	159	340	456	140	1,095
Total	1,040	1,497	1,819	650	5,006

Table 2. Topics' Prevalence and Top (lower-cased) Words, from Most to Least Prevalent

#	%	Label	Top Words	Top FREX Words
1	5.7	Ebola Toll	ebola, said, outbreak	leone, sierra, guinea
2	4.5	H1N1 Outbreak	said, flu, cases	confirmed, cases, deaths
3	4.1	H1N1 Vaccine	vaccine, flu, said	vaccine, doses, shots
4	4.1	Disaster Response	health, public, will	resources, organizations, systems
5	3.7	Microcephaly	zika, virus, said	microcephaly, babies, pregnancy
6	3.5	Flu Narratives	said, family, day	mother, son, daughter
7	3.3	Tourism	said, year, company	sales, revenue, profit
8	3.3	Medical Treatment	flu, people, children	tamiflu, asthma, adults
9	3.3	Markets & Stocks	percent, economic, economy	investors, markets, stock
10	3.2	Florida	zika, said, virus	mosquito, mosquitoes, florida
11	3.2	Global Co-op	said, united, states	ukraine, korea, syria
12	3.2	Thomas Duncan	ebola, said, hospital	duncan, dallas, texas
13	3.1	Midterm Elections	republican, democrats, republicans	republican, voters, election
14	3.0	Dr. Kent Brantly	said, ebola, patients	brantly, unit, emory
15	2.7	Ebola Drugs	drug, said, vaccine	drug, trials, fauci
16	2.7	Legal Issues	said, government, law	court, law, legal
17	2.6	H1N1	flu, virus, pandemic	pigs, pandemic, pork
18	2.6	Ebola Team	obama, president, white	obama, clinton, white
19	2.5	Schools	school, schools, students	students, schools, school
20	2.5	Screening Passengers	travel, will, said	travelers, airport, travel
21	2.4	Funding	congress, bill, house	funding, congress, bill
22	2.4	Liberians	said, ebola, people	monrovia, village, liberian
23	2.3	Sport Events	team, game, said	game, coach, players
24	2	Research	research, said, can	technology, cells, science
25	2	Media	news, media, times	twitter, media, internet
26	2	Quarantines	new, york, city	mayor, cuomo, york
27	1.9	Rio Athletes	olympics, said, games	gold, olympics, athletes
28	1.9	Blood Donation	blood, virus, test	blood, test, tests
29	1.7	Mexico	mexico, mexican, city	mexican, mexico, mecixos
30	1.7	Hospitals & Doctors	patients, doctors, medical	doctors, patients, hospitals
31	1.6	Mosquitoes	mosquitoes, water, species	species, animals, bats
32	1.4	Hygiene	hands, hand, masks	masks, hands, hand
33	1.2	Rio Games	rio, brazil, games	rousseff, brazils, brazilian
34	1.2	Biosafety	says, usa, cdc	says, usa, lab
-	5.9	Irrelevant	like, people, can	things, book, seems

about tolls, and 28.5% information about symptoms. Medical/health disruption information was central, although to a lower degree, to theme B ($M=2.27$, $SD=.89$), where 64.2% of the articles included health effects information, 30.9% symptoms information, 35.7% tolls information, and only 16.6% vulnerability information (the widest gap from theme

A on health/medical disruptions). For theme C, medical/health disruptions were not central at all ($M=1.04$, $SD=.18$); none of the articles included any information about health effects, symptoms, tolls, or vulnerability.

Individual response—ANOVA tests found significant differences between themes in terms of the centrality of

Figure 1. A topics network, with clustering based on the community detection method of Louvain and backbone reduction with $\alpha = .15$. The size of the circle represents the relative size of the topic. The labels were applied by the researcher.

individual response information. It was evaluated as most central in articles from theme A ($M = 1.77$, $SD = .97$), followed by theme B ($M = 1.17$, $SD = .50$), and not central at all to theme C ($M = 1.00$, $SD = 0.0$). Half of the articles in theme A consisted of efficacy information, while only 14.2% of those in theme B included such information. No articles in theme C included any efficacy information.

Organizational response—ANOVA tests found significant differences between themes in terms of the centrality of information about organizational response. It was evaluated as most central to articles in theme B ($M = 3.50$, $SD = .81$), followed by theme A ($M = 2.09$, $SD = .98$). It was least central to theme C ($M = 1.04$, $SD = .24$). Most theme B articles (97.6%) included organizational response information, compared to 71.4% in theme A and 2.3% in theme C.

Social/economic disruption—ANOVA tests found significant differences between themes in terms of the centrality of information about social/economic disruption. It was evaluated as most central to theme C ($M = 3.16$, $SD = .64$), followed by theme A ($M = 1.5$, $SD = .71$), and least central

to theme B ($M = 1.09$, $SD = .31$). All of the articles in theme C included social/economic disruption, compared to 40.4% in theme A and 9.5% in theme B.

Next, each theme was labeled by the researcher. The process of labeling is hermeneutic and reflects the interpretation of the researcher. Theme A emphasized medical/health disruption, and to a lesser degree individual response and organizational response. The theme consisted of topics such as microcephaly, flu outbreak in Mexico, and H1N1 biology, but also topics about biosafety and vaccines. As this theme is primarily about the scientific attempt to understand and cope with the medical conditions through research and development of treatments, it was labeled the “scientific” theme.

Theme B emphasized organizational response and, to a lesser degree, medical/health disruptions. It consisted of topics about diseases abroad (mostly Ebola), narratives about specific people (doctors and patients) who could bring the disease into the United States, and attempts to respond to a potential American disaster by screening passengers at airports and by promoting attention to

Table 3. Topics in Each Theme

Theme	Topics
A	Biosafety, research, Ebola drugs, mosquitoes, H1N1 vaccine, Florida, medical treatment, H1N1, blood donation, microcephaly, H1N1 outbreak, NY & NJ quarantines, Mexico, media
B	Hospitals & doctors, Thomas E. Duncan, disaster response, Dr. Kent Brantly, screening passengers, Ebola toll, hygiene, Liberians
C	Schools, flu narratives, sport events, legal issues, withdrawing athletes, Rio games, midterm elections, tourism, Ebola team, funding, international cooperation, markets and stocks

Table 4. The Relative Prevalence of Themes by Disease

	<i>H1N1 (%)</i>	<i>Ebola (%)</i>	<i>Zika (%)</i>	<i>Grand Mean (%)</i>
Theme A	42.2	21.7	36.9	32.4
Theme B	12.8	34.2	17.7	22.9
Theme C	44.7	44.1	45.2	44.5

personal hygiene. As this theme is mostly about a disease abroad and the attempts to prevent it from spreading in the United States, it was labeled the “pandemic” theme to characterize its emphasis on global diseases.

Theme C was devoted almost completely to discussions of the impact of diseases on the social/economic disruption. It included no health or scientific information. Its topics included impacts on tourism, stock markets, sports events, legal issues, schools, and elections and topics about international cooperation, funding, and personal stories. As this theme is primarily about the impact on cultural, economic, and political institutions, it was labeled the “social” theme. Table 5 summarizes the differences between themes in terms of centrality of CERC variables.

RQ3b examined whether there were differences in the prominence of crisis information recommended in the CERC guidelines between the coverage of different diseases. As compared to the differences found between themes, ANOVA tests found no significant differences on any of the variables between the 3 diseases ($p > .05$).

RQ4 examined the broad changes in the use of themes over the timeline of each disease. Several patterns can be seen in Figure 2, showing changes in prevalence of themes over time by disease. First, the pandemic coverage seems to be higher at the beginning of new epidemics, and it declines with time. This is especially discernible in Ebola, where the first months of coverage dealt almost exclusively with the threat of an outbreak taking place abroad. Similar patterns can be seen for H1N1. Zika, however, is more complicated, as new cases in Puerto Rico and Florida during the summer of 2016, as well as the return of visitors from Brazil following the Olympics, raised new concerns about potential outbreaks in the United States.

Generally, the use of the social theme increases over time, slowing down toward the end of the epidemic. The social theme is almost absent at the very beginning of coverage of all 3 diseases (especially Zika and Ebola) and seems to rise over time. The use of the scientific theme is relatively high at the beginning of all 3 epidemics. For Ebola and Zika, it drops over time, before becoming more prevalent again toward the end of the outbreak. For H1N1, the use of the scientific theme

is at its peak during October 2009, around the time the H1N1 vaccine was released to the public (see Figure 2).

DISCUSSION

Media coverage of epidemics is essential for the dissemination of CDC’s messages during crises. In order to ensure adequate quantity and quality of media coverage, the CERC framework urges health communicators to meet the needs and agenda of the media.⁶ However, data on the behavior of newspapers during outbreaks are limited and tend to focus on specific diseases or to use predefined categories.

This study examined topical and thematic coverage patterns used by American newspapers during 3 recent epidemics. It is expected that newspapers will treat each disease somewhat differently, as they varied in terms of spread, lethality, severity, and other aspects. However, due to journalistic routines and the commonalities that diseases do share, it was expected that broader themes could be extracted from the combined corpus. Two clustering methods were used: Topic modeling allowed us to examine the specificities and idiosyncrasies of each disease, while the network analysis revealed broader themes based on the common linguistic grounds that connect topics and diseases. For example, while the social implications of Zika differ from those of H1N1 and Ebola, they all have social implications that generally use similar language, as can be demonstrated by the existence of the social theme.

The analysis identified 34 specific and general topics and clustered them into 3 themes. The content of themes was analyzed by human coders to evaluate the centrality of crisis information emphasized in the CERC framework. The analysis showed differences between themes in the use of crisis information. The social theme focused exclusively on the impact on social/economic disruption. The scientific theme focused strongly on medical and health risks and moderately on organizational efforts to contain the disease. The pandemic theme was dominated by a discussion of organizational response. The use of themes throughout the timeline of diseases was dynamic and changed between diseases.

Table 5. Centrality of CERC Components by Theme

	<i>Grand Mean</i>	<i>Scientific Theme</i>	<i>Pandemic Theme</i>	<i>Social Theme</i>	<i>p Value (F)</i>
Risk and facts	2.198	3.274	2.274	1.048	<.001
Decision making	1.317	1.773	1.178	1.00	<.001
Resource allocation	2.214	2.095	3.500	1.048	<.001
Normalcy and well-being	1.920	1.095	1.500	3.167	<.001

Note: Centrality was measured on a 4-point Likert scale from 1 (not central at all to the article) to 4 (very central to the article).

Figure 2. Theme prevalence over time for each disease, with local-regression (Loess) smoothing of span = 0.3. Prevalence is calculated as the accumulated Theta values for all the topics in each theme in all the articles in any point in time. Since the irrelevant topic was omitted, for each article the sum of themes A, B, and C is always equal to 1.

An examination of the prevalent information also reveals what was absent from coverage. Instead of providing readers with a comprehensive depiction of a disease, articles dominated by different themes emphasized specific infor-

mation types and omitted others. For example, articles emphasizing social and economic disruptions tended to lack a discussion of health implications. Articles that focused on medical and health implications, on the other hand, tended not to discuss social implications. Both articles that focused on health and those that focused on social implications tended not to discuss what can be done to ameliorate risks. These discussions were most prevalent in articles about organizational response, which in turn tended not to include information about health or social consequences. Most alarming is the fact that individual response information, a crucial component that was stressed as central in most theories of health behavior and decision making and was found to be crucial for the formation of self-efficacy and the promotion of healthy behaviors,⁵⁰⁻⁵² was scarce across all 3 themes.

Health communicators working with journalists during epidemics could use these findings to develop a better understanding of the media's needs and practices. The CERC framework recommends that communicators stick to specific topics when speaking with journalists. The results suggest that these topics should emphasize what individuals might do to reduce threats and risks. Also, since the CERC guidelines recommend providing journalists with background material on risk issues, we encourage communicating to journalists the importance of that information and the need to include it even in articles that focus on other aspects of the disease, such as its social implications.

We also suggest that health communicators use comprehensive and multi-layered messages in the organizations' press releases, emphasizing health and social risks as well as information about individual and organizational response. While this may be an idealistic portrayal of journalism, the addition of a paragraph or 2 about the threats of Zika and ways to mitigate it to an article about, for example, a golfer withdrawing from the Olympic games might go a long way toward improving crisis communications.

Finally, because health communicators' ability to shape news coverage is limited, we encourage health organizations to include all CERC components in their own public communications. For example, when creating social media messages, health communicators should attempt to fill the gaps discerned in this analysis and provide audiences with information they may not receive from the media, especially practical advice that can inform decision making.

This study has several limitations. First, a systematic content analysis is an essential step toward understanding the media's needs during epidemics, as well as gaining insight into the type of information the public receives from it during a crisis. However, content differences are not effects.⁵³ Future studies should test whether exposure to articles with different themes have different influences on perceptions (such as uncertainty and efficacy), attitudes (including trust), and behaviors (compliance with and support of CDC). Second, future studies may look into other channels, such as CDC's social media. Third, topic

models may introduce uncertainty and error, as different topic models and clustering techniques may yield somewhat different results. To address that issue, sensitivity analyses were conducted comparing the results of the study to alternative models using different numbers of topics and different clustering techniques. Results remained generally the same across models. Fourth, content was evaluated using extreme cases, where articles were very strongly dominated by a specific theme. As a result, many articles from the social theme articles discussed diseases only tangentially. However, these articles represent a common phenomenon in the corpus.

Despite these limitations, our results shed light on the media's needs and agenda during previous epidemic outbreaks. It identified 3 distinct themes used by the media when covering epidemics using a methodology applying novel network analysis approaches to the analysis of topic structures for theme identification⁵⁴ and pointed to the discrepancies between CERC recommendations and the information that dominates each of these themes. Understanding journalistic routines during outbreaks could allow health communicators to better shape their own messages, as well as their relationships with journalists.

ACKNOWLEDGMENTS

The author wishes to thank Joseph N. Cappella, Kathleen Hall Jamieson, Robert C. Hornik, and Dror Walter for their insightful contributions to this project.

REFERENCES

- Seeger MW, Reynolds B. Crisis communication and the public health: integrated approaches and new imperatives. In: Seeger MW, Sellnow TL, Ulmer RR, eds. *Crisis Communication and the Public Health*. Cresskill, NJ: Hampton Press; 2008.
- Coombs TW. Conceptualizing crisis communication. In: Heath RL, O'Hair HD, eds. *Handbook of Risk and Crisis Communication*. Abingdon, UK: Routledge; 2009:99-118.
- Sellnow TL, Seeger MW. *Theorizing Crisis Communication*. West Sussex, UK: John Wiley & Sons; 2013.
- Veil S, Reynolds B, Sellnow TL, Seeger MW. CERC as a theoretical framework for research and practice. *Health Promot Pract* 2008;9(4 Suppl):26S-34S.
- Reynolds B, Seeger MW. Crisis and emergency risk communication as an integrative model. *J Health Commun* 2005; 10(1):43-55.
- Reynolds B, Seeger M. *Crisis and Emergency Risk Communication*. Atlanta: Centers for Disease Control and Prevention; 2014. https://emergency.cdc.gov/cerc/resources/pdf/cerc_2014_edition.pdf. Accessed April 10, 2018.
- Coombs TW. Revising situational crisis communication theory: the influence of social media on crisis communication theory and practice. In: Austin LL, Jin Y, eds. *Social Media and Crisis Communication*. New York: Routledge; 2017:21-37.
- Weick KE. Enacted sensemaking in crisis situations. *Journal of Management Studies* 1988;25(4):305-317.
- Reynolds B. Crisis and Emergency Risk Communication: Pandemic Influenza. Atlanta: Centers for Disease Control and Prevention; 2007. <https://emergency.cdc.gov/cerc/resources/pdf/cerc-pandemicflu-oct07.pdf>. Accessed April 11, 2018.
- Lundgren RE, McMakin AH. *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks*. Hoboken, NJ: John Wiley & Sons; 2013.
- Glowacki EM, Lazard AJ, Wilcox GB, Mackert M, Bernhardt JM. Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: an analysis of a Zika live Twitter chat. *Am J Infect Control* 2016;44(12):1709-1711.
- Hindman M. *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press; 2008.
- Lewis N, Martinez LS, Freres DR, et al. Seeking cancer-related information from media and family/friends increases fruit and vegetable consumption among cancer patients. *Health Commun* 2012;27(4):380-388.
- Djerf-Pierre M, Shehata A. Still an agenda setter: traditional news media and public opinion during the transition from low to high choice media environments. *J Commun* 2017; 67(5):733-757.
- Shehata A, Strömbäck J. Not (yet) a new era of minimal effects: a study of agenda setting at the aggregate and individual levels. *International Journal of Press/Politics* 2013; 18(2):234-255.
- Hornik RC. Exposure: theory and evidence about all the ways it matters. *Social Marketing Quarterly* 2002;8(3):31-37.
- Covello VT. Best practices in public health risk and crisis communication. *J Health Commun* 2003;8(Suppl 1):5-8; discussion 148-151.
- Veil SR, Ojeda F. Establishing media partnerships in crisis response. *Communication Studies* 2010;61(4):412-429.
- Althaus SL, Tewksbury D. Agenda setting and the "new" news patterns of issue importance among readers of the paper and online versions of the *New York Times*. *Communication Research* 2002;29(2):180-207.
- Galtung J, Ruge MH. The structure of foreign news: the presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *J Peace Res* 1965;2(1):64-90.
- Gans HJ. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek and Time*. New York: Vintage Books; 1979.
- Fischer HW III. *Response to Disaster: Fact Versus Fiction and Its Perpetuation*. 2d ed. Lanham, MD: University Press of America; 1998.
- Singer E, Endreny PM. *Reporting on Risk: How the Mass Media Portray Accidents, Disease, Disasters, and Other Hazards*. New York: Russell Sage Foundation; 2003.
- Iyengar S. *Is Anyone Responsible? How Television Frames Political Issues*. Chicago: University of Chicago Press; 1994.
- Lee ST. Predictors of H1N1 influenza pandemic news coverage: explicating the relationships between framing and news release selection. *Int J Strateg Commun* 2014;8(4):294-310.
- Shih T-J, Wijaya R, Brossard D. Media coverage of public health epidemics: linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics. *Mass Commun Soc* 2008;11(2):141-160.
- Tankard JWJ. The empirical approach to the study of media framing. In: Reese SD, Gandy OH Jr, Grant AE, eds.

- Framing Public Life: Perspectives on Media and Our Understanding of the Social World*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001:95-106.
28. Powers JH, Xiao X. *The Social Construction of SARS: Studies of a Health Communication Crisis*. Amsterdam: John Benjamins Publishing; 2008.
 29. Herek GM, Capitanio JP, Widaman KF. Stigma, social risk, and health policy: public attitudes toward HIV surveillance policies and the social construction of illness. *Health Psychol* 2003;22(5):533-540.
 30. Jockers M. *Macroanalysis: Digital Methods and Literary History*. Champaign, IL: University of Illinois Press; 2013.
 31. David CC, Atun JM, Fille E, Monterola C. Finding frames: comparing two methods of frame analysis. *Commun Methods Meas* 2011;5(4):329-351.
 32. Matthes J, Kohring M. The content analysis of media frames: toward improving reliability and validity. *J Commun* 2008;58(2):258-279.
 33. Sinclair J. *Trust the Text: Language, Corpus and Discourse*. Abingdon, UK: Routledge; 2004.
 34. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit Anal* 2013;21(3):267-297.
 35. Jockers M. *Text Analysis with R for Students of Literature*. Cham, Switzerland: Springer; 2014.
 36. Quinn SC, Thomas T, McAllister C. Lessons from the 2001 anthrax attack: a conceptual model for crisis and emergency risk communication. In: Seeger MW, Sellnow TL, Ulmer RR, eds. *Crisis Communication and the Public Health*. Cresskill, NJ: Hampton Press, Inc.; 2008:23-42.
 37. Reynolds B. Zika Crisis and Emergency Risk Communication (CERC) discussion: what the public needs when risks are uncertain [PowerPoint]. Centers for Disease Control and Prevention; 2016. <https://emergency.cdc.gov/cerc/ppt/zika-cerc-discussion-uncertainty.pdf>. Accessed April 11, 2018.
 38. Seeger MW, Sellnow TL, Ulmer RR. *Communication and Organizational Crisis*. Westport, CT: Greenwood Publishing Group; 2003.
 39. Jamieson KH, Campbell KK. *The Interplay of Influence: News, Advertising, Politics, and the Mass Media*. 5th ed. Belmont, CA: Wadsworth Publishing; 2000.
 40. Roberts ME, Stewart BM, Tingley D, et al. Structural topic models for open-ended survey responses. *Am J Polit Sci* 2014;58(4):1064-1082.
 41. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3:993-1022.
 42. Wallach HM. Topic modeling: beyond bag-of-words. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York: ACM; 2006:977-984.
 43. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. government arts funding. *Poetics* 2013;41(6):570-606.
 44. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2011:262-272. <http://dl.acm.org/citation.cfm?id=2145432.2145462>. Accessed April 11, 2018.
 45. Wallach HM, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. New York: ACM; 2009:1105-1112.
 46. Taddy MA. On estimation and selection for topic models. In: *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*. 2012. <http://proceedings.mlr.press/v22/taddy12/taddy12.pdf>. Accessed April 11, 2018.
 47. Serrano MA, Boguñá M, Vespignani A. Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci U S A* 2009;106(16):6483-6488.
 48. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E* 2006;74(3):036104.
 49. Meo PD, Ferrara E, Fiumara G, Provetti A. Generalized Louvain method for community detection in large networks. In: *2011 11th International Conference on Intelligent Systems Design and Applications*; 2011:88-93.
 50. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl PDJ, Beckmann DJ, eds. *Action Control*. Berlin, Heidelberg: Springer; 1985:11-39. http://link.springer.com/chapter/10.1007/978-3-642-69746-3_2. Accessed April 11, 2018.
 51. Rosenstock IM. Historical origins of the health belief model. *Health Educ Behav* 1974;2(4):328-335.
 52. Bandura A. Evaluation of AIDS prevention and education programs: perceived self-efficacy in the exercise of control over AIDS infection. *Eval Program Plann* 1990;13(1):9-17.
 53. Jamieson KH, Cappella JN. Bridging the disciplinary divide. *PS Polit Sci Polit* 1996;29(1):13-17.
 54. Walter D. *It's the Structure, Stupid: Exploring Thematic Diversity in News Coverage and Social Media Activity of Political Candidates Using Unsupervised Machine Learning* [doctoral dissertation]. Philadelphia: Annenberg School for Communication, University of Pennsylvania; 2018.

*Manuscript received December 29, 2017;
revision returned February 12, 2018;
accepted for publication February 20, 2018.*

Address correspondence to:
Yotam Ophir, PhD
University of Pennsylvania
Annenberg School for Communication
3620 Walnut Street
Philadelphia, PA 19104

Email: yophir@asc.upenn.edu