



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

ORIGINAL ARTICLES

Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: validating a maximum entropy machine-learning model

Steven J. Hoffman^{a,*}, Victoria Justicz^b

^aGlobal Strategy Lab, Faculty of Law, University of Ottawa, 57 Louis Pasteur St, Ottawa, Ontario, Canada K1N 6N5

^bHarvard Global Health Institute, Harvard University, 42 Church St, Cambridge, Massachusetts 02138, USA

Accepted 8 December 2015; Published online 7 March 2016

Abstract

Objective: To develop and validate a method for automatically quantifying the scientific quality and sensationalism of individual news records.

Study design: After retrieving 163,433 news records mentioning the Severe Acute Respiratory Syndrome (SARS) and H1N1 pandemics, a maximum entropy model for inductive machine learning was used to identify relationships among 500 randomly sampled news records that correlated with systematic human assessments of their scientific quality and sensationalism. These relationships were then computationally applied to automatically classify 10,000 additional randomly sampled news records. The model was validated by randomly sampling 200 records and comparing human assessments of them to the computer assessments.

Results: The computer model correctly assessed the relevance of 86% of news records, the quality of 65% of records, and the sensationalism of 73% of records, as compared to human assessments. Overall, the scientific quality of SARS and H1N1 news media coverage had potentially important shortcomings, but coverage was not too sensationalizing. Coverage slightly improved between the two pandemics.

Conclusion: Automated methods can evaluate news records faster, cheaper, and possibly better than humans. The specific procedure implemented in this study can at the very least identify subsets of news records that are far more likely to have particular scientific and discursive qualities. © 2016 Elsevier Inc. All rights reserved.

Keywords: Health communication; Mass media; News; Pandemics; Machine learning; Validation studies

1. Introduction

The news media is one of the most powerful societal influences and most important sources of publicly available health information. It can significantly influence people's health-related behaviors [1], clinical practices [2], and policymaking processes [3]. Yet, current news coverage of health issues is not optimal. Prior studies have identified instances of health information being distorted or misreported in the news, presumably resulting in gaps among what researchers know about health issues, how journalists convey this information, and, ultimately, the reports on which

health professionals, policymakers, and the public act [4–7].

Accurate health news coverage is particularly important in the context of crises like pandemics, when events are rapidly unfolding, when facts are constantly changing, and when more credible sources may be unavailable or inaccessible [8]. But news coverage is probably no better during crises and may actually be worse. A recent systematic review, which integrated findings from 13 content-analytic studies, concluded that the news media may have dramatized the A/H1N1 influenza (H1N1) pandemic of 2009–2010 through excessive coverage and overemphasis on the threat posed by the virus rather than available self-protection measures [9]. Initial genomic studies of the H1N1 virus were reported sensationally and in isolation without being put in the context of the larger body of research to which they contributed. Worst-case scenarios for the H1N1 pandemic were sometimes laid out theatrically without caveating possible risks with any sense of

Conflicts of interest: We declare there are no competing interests.

Funding: S.J.H. is financially supported by the Canadian Institutes of Health Research, Research Council of Norway, and Trudeau Foundation.

* Corresponding author. Tel.: +1-613-562-5800x8811; fax: +1-613-562-5124.

E-mail address: steven.hoffman@uottawa.ca (S.J. Hoffman).

What is new?

Key findings

- A new automated method for quantitatively evaluating the relevance, scientific quality, and sensationalism of individual news records was developed and successfully modeled, applied, and validated on a huge corpus of news records mentioning two pandemics.

What this adds to what was known?

- Even rudimentary machine-learning models can accurately classify text documents for complex attributes like scientific quality and sensationalism.

What is the implication and what should change now?

- Automated text analysis and machine-learning modeling represent exciting frontiers in health research and news media analysis.
- With further developments, these approaches should be able to help detect performance gaps, identify problems, develop solutions, evaluate interventions, and hold organizations accountable.

the likelihood (or unlikelihood) in which they may or may not be realized. When high quality, specific information was available, the journalistic imperative of balanced coverage too often resulted in trustworthy evidence from credible scientists reported alongside ill-informed opinions from the most popular celebrities and conflicted lobbyists

[10]. Similar concerns were raised following the severe acute respiratory syndrome (SARS) outbreak in 2003 [11,12]. Likewise, the 2014 Ebola outbreak was consistently front-page news around the world for weeks—drawing unprecedented public interest (see Fig. 1)—despite only a single Ebola death outside of West Africa [13].

To researchers, this “research-to-reporting gap” and the broader “research-to-action gap” that it perpetuates is frustrating. But to those people who rely on the media as a primary source of health information—the health professionals who provide treatment, the policymakers who direct government action, and the public who make personal health decisions every day—this gap is potentially harmful. It means people may be routinely left to act on suboptimal information and unnecessary fear, and therefore cannot make informed decisions about how to respond to pandemics.

At the very least, suboptimal media coverage of pandemics reduces capacity to quickly access, assess, adapt, and apply emerging information as it is generated, disseminate public health guidance, and coordinate responses of health system stakeholders. More broadly, suboptimal coverage can diminish public discourse on policy issues, trust in science, and accountability for decisions, thereby affecting good governance, oversight and broader principles of civic engagement and democratic responsibility [3,14].

This study developed a systematic and comprehensive method for automatically quantifying the scientific quality and sensationalism of news media coverage which was then validated on a corpus of news records published during the SARS and H1N1 pandemic alert periods. Scientific quality is about accurate reporting that reflects truth and avoids bias [15]. Sensationalism is a discourse strategy of presenting news as more extraordinary, interesting, or relevant than is objectively warranted [16]. Analysis of vast quantities of qualitative data is aided by advances in automatic and

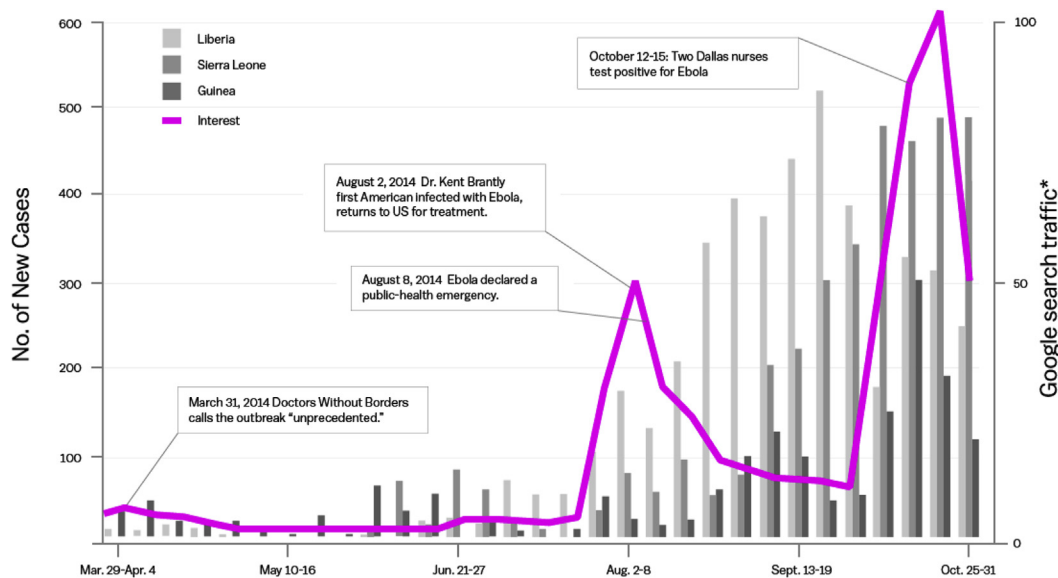


Fig. 1. Google searches for “Ebola” versus New Ebola Cases. Figure reproduced with permission from Vox Media Inc.

computer-assisted methods of extracting, organizing, and consuming knowledge from unstructured text [17,18]. These machine-learning methods allow classifications of text documents according to user-chosen categories by applying human classifications of a small subset of documents to the rest of the documents [19].

The ultimate goal of this study is to advance methods for “technoregulation” of the news media, which represents the deliberate use of technology to regulate an industry that is mostly impervious to traditional law-based regulatory mechanisms, due to, in this case, constitutional freedoms of speech and the press [20]. For example, since in most countries it is impossible to pass a law mandating high-quality news coverage, interested parties could use automated methods to continuously publish evaluations of news media organizations with the goal of incentivizing top-notch work and shaming suboptimal sources. Hopefully the development of methods for assessing news media coverage can also facilitate evaluations of interventions to improve it.

2. Methods

The scientific quality and sensationalism of news media coverage mentioning pandemics were evaluated by optimizing the retrieval of relevant news records, developing tools for quantitatively measuring these qualities on a random sample of news records, using a maximum entropy model for automatic unstructured text classification, and validating the classification model by measuring its accuracy. See Appendix 1 for a detailed explanation of these steps which are only summarized in the following paragraphs.

2.1. Retrieving news records

News records were retrieved from the LexisNexis database using a search protocol that was developed in consultation with a social science librarian and continually optimized over 3 stages of pilot tests to maximize sensitivity (i.e., true positives) and specificity (i.e., true negatives). LexisNexis provides access to over 15,000 sources, including over 3,000 newspapers, 2,000 magazines, and many newswires, blogs, and television broadcasting transcripts from around the world [21]. The following search was implemented to retrieve English-language records published on SARS from March 15, 2003 to May 18, 2004: “SARS” or “severe acute respiratory syndrome” or “coronavirus” or “SARS-CoV”. The following search retrieved records on H1N1 from April 23, 2009 to September 10, 2010: “H1N1” or “A(H1N1)” or “S-OIV” or “swine origin influenza” or ([“flu” or “influenza”] and [“pig” or “swine” or “hog”]).

2.2. Measuring scientific quality and sensationalism

The scientific quality of individual news records was quantitatively measured using an adapted version of the

Index of Scientific Quality outlined in [22]. This index was used because it was the only empirically validated tool for measuring scientific quality that was found after extensive literature searches, it was devised with input from 38 research methodologists and additional journalism scholars, and it was specifically developed for evaluating health news records. The index facilitates systematic scoring of news records by integrating human ratings on five-point Likert-type scales measured along seven dimensions: (1) applicability; (2) opinions versus facts; (3) validity; (4) magnitude; (5) precision; (6) consistency; and (7) consequences. A score of “1” or “2” indicates the news record contains “critical or extensive shortcomings,” a score of “3” indicates “potentially important but not critical shortcomings,” and a score of “5” indicates “minimal shortcomings” [22]. Other approaches to measure scientific quality tend to rely on proxies, such as author affiliation [23], sources of information [24] and referencing practices [25].

For this study, the Index of Scientific Quality was slightly updated to improve clarity based on pretesting with three research assistants (RAs) and consultation with a professional copy editor. This included dropping the “magnitude” dimension and merging “consistency” and “consequences” into a single rating due to overlap and highly correlated responses in pretesting. Illustrative examples of statements that would militate toward different scores for the news records were added to boost interrater reliability; the unit of analysis remained the news record as a whole, which is the unit of analysis for which the Index of Scientific Quality was originally developed [22].

The sensationalism of news records was measured using a new tool developed from a pragma-linguistic framework of five “sensationalist illocutions”—exposing, speculating, generalizing, warning, and extolling—that [16] identified as indicative of sensationalist reporting through surveys and focus groups. This framework was used because it facilitated direct measurement of sensationalism by conceptually identifying its facets and dividing it into discrete components. Other approaches were either: too specifically tailored for evaluating news about particular events like suicides [26,27] and anthrax attacks [28]; depended on elements not found in text-based databases like background music [29] and camera positions [30]; incorporated consideration of the topic cover [31]; used simple dictionary methods like counting intensifying adjectives [32]; or relied on proxy indicators like newspaper page number [27], article length [26], and off-record attribution [28]. One previous study examined the 1918 influenza pandemic, but its approach assessed news media coverage broadly rather than measured the sensationalism of individual news records [33].

Questions, examples, and corresponding five-point Likert-type scales were crafted to assess the five components of sensationalism identified in [16]. A score of “1” indicates the news record was “not at all sensationalizing,” a “2” indicates there was “not too much sensationalizing,” a “3” indicates the record was “somewhat sensationalizing,” a “4”

corresponds to “fairly sensationalizing,” and a “5” means it was “very sensationalizing.” This means that a “5” is the worst score possible, unlike with scientific quality—a virtue, rather than a vice—where a “5” is the best score possible. Only minor word changes were made after pretesting. See [Appendices 2 and 3](#) for the pretested and final tools used to measure the scientific quality and sensationalism of individual news records.

Three RAs independently assessed a simple random sample of 500 retrieved LexisNexis news records, first for relevance based on whether they were actually focused on the SARS or H1N1 pandemics, and then, if so, to score them using the tools developed for measuring scientific quality and sensationalism. This sample size was chosen based on previous work that suggests the advantages of more human coding begins to experience diminishing returns at this point [14]. Disagreements on relevance were resolved by consensus. Three-rater Fleiss’ kappa, intraclass correlation coefficients (ICCs), and Krippendorff’s alphas were calculated to assess interrater reliability.

2.3. Maximum entropy modeling for classifying relevance, scientific quality, and sensationalism

These 500 news records served as a “training set” for development of a maximum entropy model that probabilistically classifies text documents [34–36]. Maximum entropy modeling is equivalent to multinomial logistic regression—both using maximum likelihood estimation—albeit the two methods are derived differently. Specifically, logistic regression maximizes the log-likelihood of model parameters knowing the exponential form of posterior probability functions, which is equivalent to the dual problem of maximum entropy modeling’s unconstrained optimization [37–40].

This modeling involved a computationally intensive inductive machine-learning procedure that (1) processed the training set to remove punctuation, capitalization, non-English words, white spaces, symbols, and non-ASCII letters; (2) converted it into a document-term matrix for quantitative analysis; (3) identified relationships distinguishing the 500 news records by how the RAs assessed relevance; (4) combined these relationships as constraints into a multinomial logistic regression that best predicts records’ relevance; (5) applied this regression to 10,000* randomly selected news records mentioning pandemics for determining relevance by the least biased maximum likelihood estimate on the available information; (6)

* The maximum entropy model was applied to 10,000 news records instead of all 163,433 records given the model’s exponentially increasing demands on computing resources as the corpus of records expands. A run of this model on all 163,433 records using default computing resources available through the Harvard-MIT Data Center’s Interactive Computing Cluster did not finish within 7 days. Applying the model to 10,000 news records took approximately 12 hours.

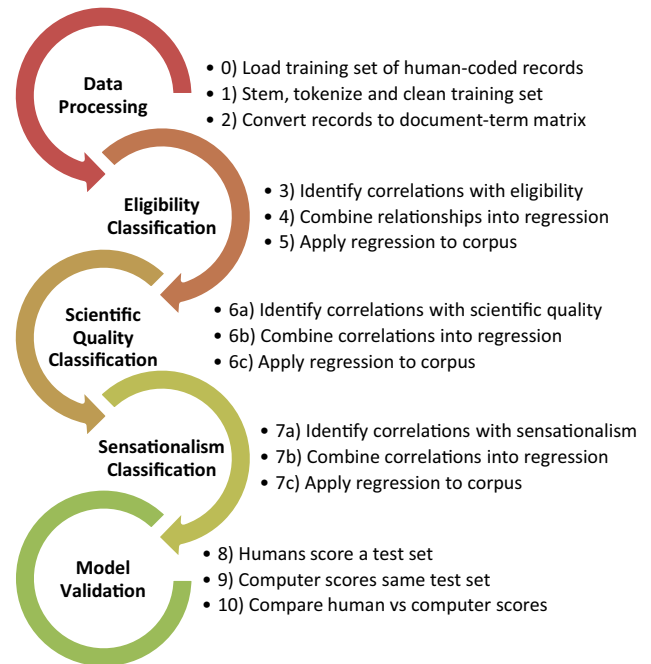


Fig. 2. Flowchart of maximum entropy modeling and validation.

repeated steps 3–5 using RA scores for a second analysis to evaluate each relevant record for scientific quality; and (7) repeated steps 3–5 using RA scores for a second analysis to evaluate sensationalism. See [Fig. 2](#) for a flowchart of this modeling.

More simply, a statistical model was trained to predict whether news records mentioning pandemic-related terms were actually about pandemics (i.e., first application) and whether they deserved a “1,” “2,” “3,” “4,” or “5” score for scientific quality and for sensationalism (i.e., second application). The first application is important to boost specificity after it was sacrificed in the optimized search for greater sensitivity.

Maximum entropy modeling was chosen from among the many machine-learning approaches that can be used for text analysis because it does not assume independence of terms; in future applications, this would allow the use of bigrams and phrases in modeling without the possibility of overlapping or double counting words that often appear together such as “World Health Organization” [34].

Data processing, statistical analyses, and text classification were conducted using the MaxEnt package (v1.3.3.1) for R statistical software (v2.15.1). See [Appendix 4](#) for R code implementing these procedures.

2.4. Validating the model

To assess the model’s internal validity, a test set of 200 news records was randomly drawn from the corpus of retrieved news records (excluding those in the training

		Training set (500 records)			MaxEnt model (10,000 records)				
		Human mean	Fleiss' kappa	ICC statistic	Total mean	SARS mean	H1N1 mean	2-sample t-test 95%CI	t-test's p-value
	relevance (application 1)	195/500	1.00	1.00	3625/10k	1101/5192	2524/4808	[0.2949, 0.3309]	<0.0001*
Scientific quality	applicability	4.66	0.82	0.99	4.83	4.81	4.84	[0.0064, 0.0633]	0.0165*
	opinion vs fact	3.32	0.66	0.97	3.31	3.15	3.38	[0.1478, 0.3125]	<0.0001*
	validity	2.62	0.58	0.95	2.54	2.48	2.57	[0.0382, 0.1267]	0.0003*
	precision	3.28	0.63	0.98	3.44	3.62	3.36	[-0.3126, -0.1982]	<0.0001*
	context	3.25	0.64	0.98	3.40	3.55	3.34	[-0.2683, -0.1548]	<0.0001*
	overall	3.17	0.84	0.99	3.32	3.22	3.32	[0.0982, 0.2083]	<0.0001*
Sensationalism	exposing	1.66	0.79	0.96	1.30	1.46	1.23	[-0.2903, -0.1694]	<0.0001*
	speculating	1.73	0.66	0.95	1.70	1.83	1.64	[-0.2556, -0.1279]	<0.0001*
	generalizing	1.54	0.79	0.95	1.33	1.40	1.29	[-0.1636, -0.0584]	<0.0001*
	warning	1.73	0.68	0.96	1.88	2.13	1.78	[-0.4199, -0.2758]	<0.0001*
	extolling	1.20	0.83	0.94	1.09	1.15	1.06	[-0.1211, -0.0590]	<0.0001*
	overall	1.81	0.68	0.97	1.73	1.94	1.63	[-0.3707, -0.2549]	<0.0001*
	Overall	2.50	0.74	0.98	2.49	2.56	2.45	[-0.1318, -0.0765]	<0.0001*

Fig. 3. Summary of the training and maximum entropy modeling exercises. ICC, intraclass correlation coefficient; Kripp's alpha, Krippendorff's alpha for ordinal data; 10k, 10,000; SARS, severe acute respiratory syndrome (2003–2004); H1N1, influenza A subtype H1N1 (2009–2010); 95% CI, 95% confidence interval for estimated difference between SARS and H1N1 means.

set) and classified independently by two RAs for relevance, scientific quality, and sensationalism. The mean RA score for each relevant record in the test set was assumed to be “correct” and used as a benchmark against which the model's second-application classifications were judged. Two-rater Cohen's kappa, ICCs, Krippendorff's alphas, and two-way paired *t*-tests were calculated to assess the model's reliability. Accuracy was calculated based on the percentage of news records that the model classified the same as the two RAs.

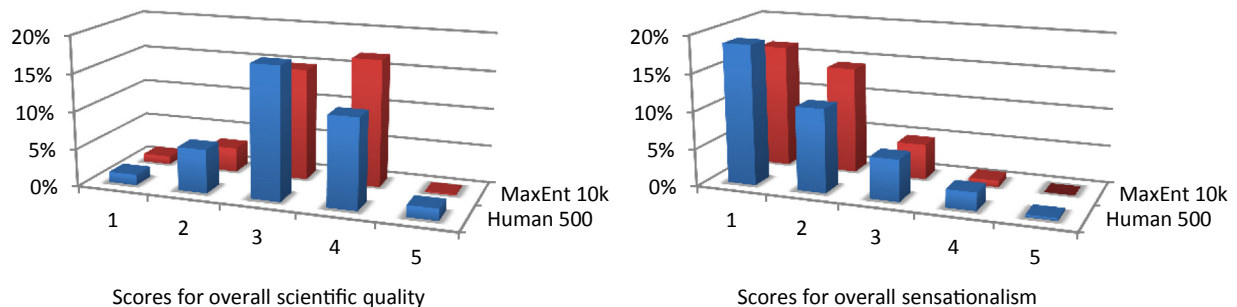
3. Results

The optimized search protocols conducted from 17 to 19 October 2013 on LexisNexis identified 89,846 news records mentioning SARS and 73,587 records mentioning H1N1, for a total of 163,433 records. RAs deemed 195 of the

500 training set records to be relevant. This means the LexisNexis searches yielded an estimated 63,739 news records that were actually focused on SARS or H1N1 for an estimated specificity of 39.0%.

The 195 relevant records in the training set had a mean overall scientific quality score of 3.17 and a mean overall sensationalism score of 1.81 with an overall Fleiss' kappa of 0.74, ICC of 0.98, and Krippendorff's alpha of 0.82, indicating substantial interrater reliability among RAs [41,42]. Scores for specific dimensions of scientific quality ranged from 2.62 for validity (“not assessed or very misleading”) to 4.66 for applicability (“minimal ambiguity”). Scores for sensationalism ranged from 1.20 for extolling to 1.73 for speculating (both indicating “minimal” presence of these illocutions).

Maximum entropy modeling of the 10,000 randomly selected news records provided revised aggregated estimates of scientific quality (total mean of overall



Note: The histograms exclude those cases that were deemed irrelevant to either the SARS or H1N1 pandemics in the first screening.

Fig. 4. Histograms of scores from the training set (500 records) and maximum entropy model (10,000 records). The histograms exclude those cases that were deemed irrelevant to either the SARS or H1N1 pandemics in the first screening.

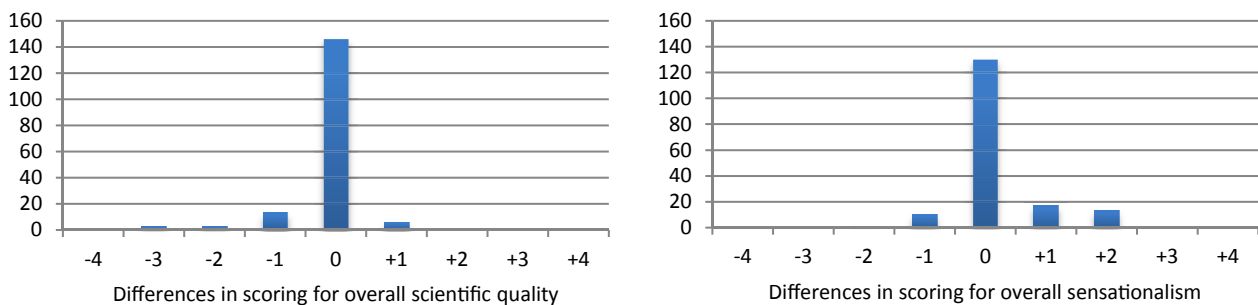
		Validation test set (200 records)					
		Human mean	Computer mean	Cohen's kappa	ICC statistic	Accuracy ±0	Accuracy ±1
Scientific quality	relevance (application 1)	65/200	74/200	0.68	0.81	86%	NA
	applicability	2.89	4.73	0.56	0.75	61%	70%
	opinion vs fact	3.54	3.20	0.72	0.84	74%	82%
	validity	1.40	2.62	0.48	0.69	60%	73%
	precision	2.40	3.54	0.60	0.77	65%	74%
	context	2.75	3.15	0.66	0.80	68%	77%
	overall	2.55	3.16	0.66	0.81	65%	78%
Sensationalism	exposing	1.88	1.45	0.54	0.70	72%	79%
	speculating	2.11	1.84	0.43	0.60	67%	76%
	generalizing	1.40	1.43	0.45	0.62	74%	81%
	warning	2.37	1.66	0.53	0.70	68%	76%
	extolling	1.09	1.09	0.58	0.74	81%	84%
	overall	2.06	1.66	0.47	0.64	73%	82%
	Overall	2.20	2.46	0.60	0.76	69%	78%

Fig. 5. Summary of the validation exercises. All Cohen's kappa and ICC statistics in the validation exercise had *P*-values well below 0.00001 indicating the high values of these statistics should not be due to chance. The scientific quality and sensationalism accuracy percentages carried forward the inaccuracies of the relevance screening which means they represent overall accuracy percentages measuring error in both applications. ICC, intraclass correlation coefficient; Kripp's alpha, Krippendorff's alpha for ordinal data; NA, not applicable.

score = 3.32, ranging from 2.54 for validity to 4.83 for applicability) and sensationalism (total mean of overall score = 1.73, ranging from 1.09 for extolling to 1.88 for warning). This means the average news record had “potentially important but not critical shortcomings” in scientific quality with “not too much sensationalizing.”

These records can also now be stratified at the individual record level for subgroup comparisons such as between records published about different pandemics; in this example, news coverage of the H1N1 pandemic was found to be statistically significantly better than the earlier SARS outbreak (two-sample *t*-test 95% confidence interval for overall scientific quality score = [0.0982, 0.2083], *P* < 0.0001; for overall sensationalism score = [−0.3707, −0.2549], *P* < 0.0001). See Figures 3 and 4 for a summary of these results.

The model performed well in the validation exercise. In the first application, the model determined relevance with 86% accuracy—which means the model and RAs were in agreement 86% of the time on which of the 200 news records in the test set were about pandemics, and disagreed on 14% of the 200 records. In the second application, overall scientific quality was scored accurately 65% of the time (or 78% if allowing ±1 deviations on the five-point scale). The model's overall sensationalism scoring was 73% accurate (or 82% if allowing ±1 deviations). These statistics indicate substantial agreement between the human and computer scoring for scientific quality and sensationalism [42]. Notwithstanding errors, population-wide estimates from the model should be relatively unbiased given the histogram of misclassifications appear to be equally biased upward and downward. See Figures 5 and 6 for a summary of these results.



Note: Counts exclude the 29 cases (14%) where the human scorers and MaxEnt model disagreed on the record's relevance to either the SARS or H1N1 pandemics in the first screening.

Fig. 6. Histograms of differences in computer scores to human scores in the validation test set (200 records). Counts exclude the 29 cases (14%) where the human scorers and maximum entropy model disagreed on the record's relevance to either the SARS or H1N1 pandemics in the first screening.

4. Discussion

4.1. Principal findings

A new method for quantitatively evaluating the relevance, scientific quality, and sensationalism of individual news records was developed and successfully modeled, applied, and validated on a huge corpus of news records mentioning two pandemics. Analyses confirmed that news media coverage of pandemics is not perfect, especially its scientific quality if not also its sensationalism. Slight improvements were observed between the SARS and H1N1 pandemics. Possible explanations for this improvement include the media learning from experience with the first pandemic and/or better crisis communications from public health authorities throughout the second pandemic. It could also be a reaction to the 2005 revision of the International Health Regulations [43–45].

4.2. Strengths and limitations

This study has several strengths compared to previous work in news media analysis on which it builds. First, it drew from over 15,000 sources of news records, making results more generalizable across a broader range of contexts. Second, it used pilot tests to optimize searches for maximal sensitivity. Third, it drew on existing tools—an empirically

validated index and rich pragma-linguistic framework—when developing new metrics for quantitatively measuring news records' scientific quality and sensationalism. Fourth, the study assessed 10,000 randomly sampled news records from a massive corpus of 163,433 records (instead of just a small sample feasible for human scoring) using recent advances in machine-learning methods and computing power. This means that population-wide estimates incorporate more information from more sources and that detailed subgroup analyses are theoretically possible given most-likely scores are available at the individual record level. Fifth, the study incorporated a relevance screening into the modeling procedures to boost specificity. Sixth, there were multiple RAs scoring the training and test set records to reduce human errors and biases. Seventh, the study validated the approach and measured its reliability and accuracy. Overall, the study showed that automated methods can quantify characteristics of news records faster (i.e., within seconds), cheaper (i.e., fewer human resources), and possibly better than humans (i.e., avoiding silly mistakes and rater drift). The specific procedure implemented in this study can at the very least identify subsets of news records that are far more likely to have particular scientific and discursive qualities.

This study also has several limitations. First, it has all the usual pitfalls of automated text classification, including its

Rank	Organization	Country	Number of records	Overall scientific quality	Overall sensationalism
1.	The Scotsman	UK	55	3.64	1.89
2.	Midland Independent Newspapers	UK	65	3.52	1.58
3.	AllAfrica	Multiple	112	3.44	1.75
4.	The Liverpool Daily Post & Echo	UK	51	3.41	1.43
5.	New Straits Times Press	Malaysia	112	3.37	1.63
6.	The New York Times	USA	100	3.35	1.71
7.	The Globe and Mail	Canada	110	3.34	1.98
8.	The Times	UK	116	3.32	1.91
9.	Guardian	UK	59	3.31	2.17
10.	Express Newspapers	UK	91	3.30	1.64
11.	Singapore Press	Singapore	58	3.29	1.76
12.	The Australian	Australia	390	3.29	1.63
13.	BBC	UK	418	3.29	1.30
14.	The Washington Post	USA	87	3.28	1.89
15.	South China Morning Post	China	208	3.27	1.95
16.	National Post	Canada	111	3.26	1.80
17.	The Sun/The News of the World	UK	108	3.24	1.69
18.	Toronto Star	Canada	155	3.23	1.98
19.	The Irish Times	Ireland	55	3.22	1.87
20.	The Daily Mirror	UK	102	3.09	1.70

Fig. 7. Ranking news media organizations by the scientific quality of their pandemic coverage, including all sources for which there were more than 50 relevant news records.

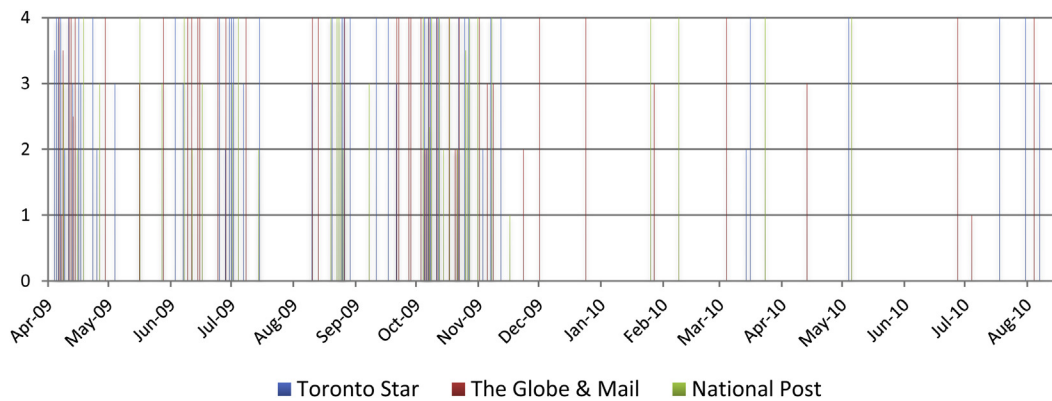


Fig. 8. Average daily scores of news records' overall scientific quality during the H1N1 pandemic period among the three Canadian newspapers with more than 50 relevant records.

many assumptions and simplifications [12]. Second, its maximum entropy model leaves information on the table by relying on a multinomial regression that treats the scientific quality and sensationalism scores as nominal rather than ordinal data. Third, it required a substantial initial investment of human resources to score the training and test sets. Fourth, the final model was not perfectly accurate with classification errors compounded across two applications. Fifth, the performance of this particular machine-learning approach was not compared to others.

4.3. Future research directions

The suboptimal news media coverage of pandemics that was found in this study affirms the need to further research this problem and identify prospects for amelioration. Also needed are advances in the imperfect methods and metrics for making these assessments.

Specifically, the ability to automatically score individual news records for their scientific quality and sensationalism should be applied to track changes, make comparisons, identify outliers, find correlations, and evaluate interventions. This could include, for example, constructing day-by-day time series of these characteristics that could be stratified to compare countries, rank news media organizations, or even judge individual journalists (see Figs. 7 and 8). Publishing rankings comparing news media organizations or journalists could encourage them to compete on quality and enhance their reporting practices. These data could also help find factors broadly associated with better news coverage (e.g., record length, readership, political affiliation) or predictive of rapid changes in scientific quality and sensationalism (e.g., new event, major announcement, public scolding of news media). Record-level data can also be used for rigorous impact evaluations such as quasi-experimental interrupted time-series analyses of interventions aimed at improving news media coverage. The feasibility of real-time analysis of news media coverage on emerging pandemics should also be explored.

Another opportunity for future research is to improve automated methods of quantitatively measuring the scientific quality and sensationalism of news records as well as other characteristics of other qualitative texts. All automated text classification methods use necessarily wrong models of text designed to help draw inferences from data; this means that diverse methods should be explored for assessing news records, and their merit evaluated according to how well they perform specific tasks, especially because more realistic or sophisticated models do not always offer better performance [19]. Methodological advances that improve models' accuracy and applicability to out-of-sample records would allow researchers to make more helpful inferences with fewer resources.

5. Conclusion

News media coverage of emerging pandemics is not as good as it should be. Developing new methods for automatically quantifying characteristics of news media coverage is an important step toward improving it. These methods represent an exciting frontier in public health research and news media analysis, because they can help detect performance gaps, identify problems, develop solutions, evaluate interventions, and hold news organizations accountable for their health reporting.

Acknowledgments

The authors thank Mathura Mahendren, Navneet Natt, Vivian Tam and Julia Woo for undertaking the human coding of news records, and Robert Kindman, Gary King, Alex Storer and Steven Worthington for statistical advice.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2015.12.010>.

References

- [1] Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. *Lancet* 2010;376:1261–71.
- [2] Grilli R, Ramsay C, Minozzi S. Mass media interventions: effects on health services utilisation. *Cochrane Database Syst Rev* 2002; CD000389.
- [3] Otten AL. The influence of the mass media on health policy. *Health Aff (Millwood)* 1992;11:111–8.
- [4] Bubela TM, Caulfield TA. Do the print media “hype” genetic research? A comparison of newspaper stories and peer-reviewed research papers. *Can Med Assoc J* 2004;170:1399–407.
- [5] Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 2000;342:1645–50.
- [6] Ransohoff DF, Ransohoff RM. Sensationalism in the media: when scientists and journalists may be complicit collaborators. *Eff Clin Pract* 2001;4(4):185–8.
- [7] Sumner P, Vivian-Griffiths S, Boivin J, Williams A, Venetis CA, Davies A, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* 2014;349:g7015.
- [8] Reissman DB, Watson PJ, Klomp RW, Tanielian TL, Prior SD. Pandemic influenza preparedness: adaptive responses to an evolving challenge. *J Homel Secur Emerg Manag* 2006;3(2). Article 13. Available at <http://www.bepress.com/jhsem/vol3/iss2/13>. Accessed March 26, 2016.
- [9] Klemm C, Das E, Hartmann T. Swine flu and hype: a systematic review of media dramatization of the H1N1 influenza pandemic. *J Risk Res* 2016;19:1–20.
- [10] Laing A. The H1N1 crisis: roles played by government communicators, the public and the media. *J Prof Commun* 2011;1:123–49.
- [11] Naylor D, Basur S, Bergeron MG, Brunham RC, Butler-Jones D, Dafoe G, et al. Learning from SARS: renewal of public health in Canada: a report of the National Advisory Committee on SARS and Public Health. Ottawa: Health Canada; 2003. Available at <http://www.phac-aspc.gc.ca/publicat/sars-sras/pdf/sars-e.pdf> Accessed March 26, 2016.
- [12] World Health Organization. World Health Report 2007: A Safer Future: Global Public Health Security in the 21st Century. Geneva: World Health Organization; 2007.
- [13] World Health Organization. Ebola Situation Report—17 June 2015. Geneva: World Health Organization; 2015. Available at <http://apps.who.int/ebola/current-situation/ebola-situation-report-17-june-2015>. Accessed March 26, 2016.
- [14] Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington DC: National Academy Press; 2001.
- [15] Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44:1271–8.
- [16] Molek-Kozakowska K. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse Commun* 2013;7:173–97.
- [17] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457:1012–4.
- [18] Hopkins D, King G. A method of automated nonparametric content analysis for social science. *Am J Pol Sci* 2010;54:229–47.
- [19] Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit Anal* 2013;21:267–97.
- [20] Brownsword R. Code, control, and choice: why East is East and West is West. *Leg Stud* 2005;25:1–21.
- [21] LexisNexis. LexisNexis Academic 2014. Available at <http://www.lexisnexis.com/en-us/products/lexisnexis-academic.page>. Accessed March 26, 2016.
- [22] Oxman AD, Guyatt GH, Cook DJ, Jaeschke R, Heddle N, Keller J. An index of scientific quality for health reports in the lay press. *J Clin Epidemiol* 1993;46:987–1001.
- [23] Soot LC, Moneta GL, Edwards JM. Vascular surgery and the Internet: a poor source of patient-oriented information. *J Vasc Surg* 1999;30:84–91.
- [24] Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999;53:105–11.
- [25] Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 2002;287:2691–700.
- [26] Niederkrotenthaler T, Voracek M, Herberth A, Till B, Strauss M, Etzersdorfer E, et al. Role of media reports in completed and prevented suicide: Werther v. Papageno effects. *Br J Psychiatry* 2010;197:234–43.
- [27] Pirkis JE, Burgess PM, Francis C, Blood RW, Jolley DJ. The relationship between media reporting of suicide and actual suicide in Australia. *Soc Sci Med* 2006;62(11):2874–86.
- [28] Swain KA. Outrage factors and explanations in news coverage of the anthrax attacks. *J Mass Commun Q* 2007;84:335–52.
- [29] Grabe ME, Zhou S, Barnett B. Explicating sensationalism in television news: content and the bells and whistles of form. *J Broadcast Electron Media* 2001;45:635–55.
- [30] Vettehen PH, Nuijten K, Peeters A. Explaining effects of sensationalism on liking of television news stories the role of emotional arousal. *Hum Commun Res* 2008;35:319–38.
- [31] Tannenbaum PH, Lynch MD. Sensationalism: the concept and its measurement. *J Mass Commun Q* 1960;37:381–92.
- [32] Burgers C, de Graaf A. Language intensity as a sensationalistic news feature: the influence of style on sensationalism perceptions and effects. *Communications* 2013;38:167–88.
- [33] Spratt M. Science, journalism, and the construction of news: how print media framed the 1918 influenza pandemic. *Am J* 2001;18(3):61–79.
- [34] Anjaria M, Guddeti RMR. A novel sentiment analysis of social networks using supervised learning. *Soc Netw Anal Min* 2014;4:1–15.
- [35] Berger AL, Pietra VJD, Pietra SAD. A maximum entropy approach to natural language processing. *Comput Linguist* 1996;22:39–71.
- [36] Jaynes ET. Information theory and statistical mechanics. *Phys Rev Ser II* 1957;106:620–30.
- [37] Malouf R. A comparison of algorithms for maximum entropy parameter estimation. COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002) 2002. Available at <http://aclweb.org/anthology/W/W02/W02-2018.pdf>. Accessed March 26, 2016.
- [38] Mount J. The equivalence of logistic regression and maximum entropy models 2011. Available at <http://www.win-vector.com/dfiles/LogisticRegressionMaxEnt.pdf>. Accessed March 26, 2016.
- [39] Nigam K, Lafferty J, McCallum A. Using maximum entropy for text classification. Paper presented at: IJCAI-99 Workshop on Machine Learning for Information Filtering. 1999; Stockholm, Sweden.
- [40] Qian M. The equivalence of logistic regression and maximum entropy modeling 2013. Available at <http://web.engr.illinois.edu/~mqian2/upload/research/notes/The%20Equivalence%20of%20Logistic%20Regression%20and%20Maximum%20Entropy%20Modeling.pdf>. Accessed March 26, 2016.
- [41] Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–82.
- [42] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [43] Hoffman SJ. The evolution, etiology and eventualities of the global health security regime. *Health Policy Plan* 2010;25:510–22.
- [44] World Health Organization. International Health Regulations (1969). Geneva: World Health Organization; 1969. Available at <http://whqlibdoc.who.int/publications/1983/9241580070.pdf>. Accessed March 26, 2016.
- [45] World Health Organization. International Health Regulations (2005). Geneva: World Health Organization; 2006. Available at www.who.int/csr/ihr/en. Accessed March 26, 2016.