



Health News Bias and its impact in Public Health

Pita-Costa, J., Stopar, L., Fuart, F., Grobelnik, M., Mladenić, D., Košmerlj, A., Belayeva, E., Leban, G., Rei, L., Fischhaber, S., & Wallace, J. G. (2019). Health News Bias and its impact in Public Health. In *Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD), Ljubljana 2019*

[Link to publication record in Ulster University Research Portal](#)

Published in:

Slovenian KDD Conference on Data Mining and Data Warehouses (SiKDD), Ljubljana 2019

Publication Status:

Published (in print/issue): 14/11/2019

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

Health News Bias and its impact in Public Health

J. Pita Costa, F. Fuart,
L. Stopar, M. Grobelnik, D. Mladenić
Quintelligence,
Jozef Stefan Institute, Slovenia Slovenia

A. Košmerlj, E.
Belayeva, L. Rei
Jozef Stefan Institute,
Slovenia

G. Leban
Event Registry,
Quintelligence,
Slovenia

S. Fischaber
Analytics Engines,
UK

J. Wallace
Ulster University,
UK

ABSTRACT

The impact of health-related news in today's society is increasing as is the awareness of the globalization of the worlds' habits and threats, and the impact on the continuous pursuit of a better quality of life. The risk of news media bias and the consequences it might have in the population is of great concern for public health, as are the available resources to identify the bias and further explore the news stories. In this paper we discuss several aspects, angles and perspectives on news media bias in the health domain, with a particular focus on digital epidemiology. We also present decision support tools developed to support decision makers in these explorations in the context of the MIDAS project, leveraging Big Data analytics to support decision making in public health. The presented resources provide health professionals with a global perspective on the worldwide news coverage of monitored health topics (such as, e.g., infectious diseases, mental health or childhood obesity), together with a workflow of tools allowing them to explore potential bias. Moreover, we discuss the specific challenges of news bias in the health domain, analyzing some typical examples, and using the Event Registry technology to further explore them. The exploration potential of the latter, in the health domain, is enhanced with the integration of an automated classifier based on MeSH Headings that allows researchers to explore the news using a similar workflow to that of exploring biomedical research in PubMed.

CCS CONCEPTS

• Information Systems • Human-centred Computing • Life and Medical Science

KEYWORDS

Data mining, health news, news media bias, Big Data, Public Health, digital epidemiology

1 HEALTH NEWS MEDIA BIAS

News media bias is a ubiquitous phenomenon that has generated various research studies in different fields. Practically any media outlet can be biased, but the public should be aware of it and news media bias should be minimized thereby offering more objectivity to the news reporting. Health related news, in particular, have a high impact on the population that tends to be more sensitive to their content. It is fairly well known that the media plays an influential role in public responses to health issues [6]. Although, the bias in health-related news can be considered in the same light as the overall news bias, with similar effect in most cases, it also

has very specific aspects deriving from the domain it is based on and the kinds of stakeholders it relates to. Often, the complexity of the information (due to the continuous innovation in medicine) along with the lack of detail can lead to misinterpretations and unconscious bias both at the media outlet and its audience. This is a common problem in the communication of science [13]. Examples of these are frequent in the context of precision medicine, where some difficult concepts and methods from genetics and life sciences play a key role while being a sensitive topic within the common public opinion.



Figure 1: MIDAS news dashboard screenshot showing a temporal intensity visualization module for the query *Ebola virus disease* to analyze and compare the media coverage on the disease outbreak during 2014 and 2019, considering only news sources located in the USA.

Another angle on news media bias is the amount of news on certain disease-related aspects, that are more abundant in media sources that are located far from where they are occurring, providing us with an idea that the occurrence is local. An example of this in the public health scenario, is the large amount of news published by news sources located in the USA about the Ebola outbreak, and the small number of cases detected in this country. The frequency of news items can be sometimes confused with its potential impact in the local citizens, by the less informed audiences. The transfer of that unclear message to a diversity of social media channels is then inevitable, as well as the subsequent accelerated proliferation of the misinformation and unconscious news bias.

The chart in Figure 1 shows two perspectives on health news bias while representing the news on Ebola virus disease media coverage limited to news sources in the USA. On one hand the peak in 2014 is not representative of the low number of cases identified in the USA. On the other hand, the weight of the disease in 2015 and now in 2019 is not representative to the high relevance of this topic to the global public health today. In July 17, 2019, the World Health Organization (WHO) was once again announcing an Ebola Outbreak in Congo with public health emergency of international

concern [14]. Though, the news coverage this time is much more local than it was back in 2015 as the reader can see in Figure 2.

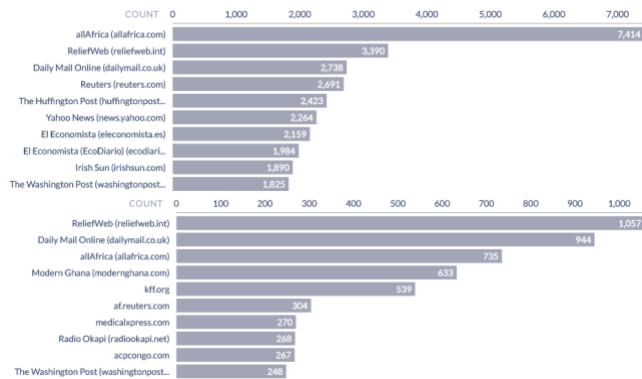


Figure 2: The coverage of the Ebola outbreak in 2014 (above) and in 2019 (below), showcasing the very different top 10 news sources covering the similar event.

A well-known generator of health news media bias was the case of the Google Flu Trends, a good example of collective intelligence estimating the influenza activity for more than 25 countries. This system was based on the queries for influenza related keywords on the Google search engine [3]. The Influenza season of 2012/2013 showed the inaccuracy of this system that, until then was closely following the data collected by the Centers for Disease Control (CDC), as seen in Figure 3. Although being more a case of algorithm bias per se, it was the responsible for false conclusions that could have had a bigger impact without the classical mechanisms in place by National and International Institutions.

2 CHALLENGES OF HEALTH NEWS

Unlike most news topics, health related topics are often close to the interest and well-being of the general public, with a large impact in the social media [2]. An example of that is the ongoing worldwide public discussion about child vaccination. However, aside from the popular diseases (such as influenza, measles, etc.), the media coverage is often not complete or does not provide an accurate coverage (not all topics are *news worthy*). The rising popularity of a certain disease implies the increase of the references to it in the media, not in parallel with the status of the disease itself. This sometimes falls into what is known as *mainstream bias*, i.e., a tendency to report what everyone else is reporting, and to avoid stories that will fall out of the core of popular news.

On the other hand, the awareness of a certain disease or the general status of public health is not always well represented in the media. Most of the times this lack of representation reflects the incomplete awareness of the general public to the state of the health nationwide. It is also the case when that awareness is higher in some countries and smaller in others. This is often the case differentiating the so-called developed countries to the so-called 3rd world countries. An example of this is the coverage of the news about the Zika virus outbreak in 2015/16.

Another angle that is relevant to this discussion is the different concepts of news media bias and what is ‘news-worthy’ discussed in [5]. Although the acknowledged importance of a complete global coverage of the status of the health of the population, some aspects of health have higher priority than others, independently of their relevance in the Public Health context. These priorities are defined by the media houses and publishers according to the expectation on the impact that the news will have in their audiences. In a more extreme sense, sensationalism is the bias in favor of the exceptional over the ordinary, aiming to give the impression that rare events, such as a victim of the Ebola outbreak in the USA, are more frequent than common events, such as a child with Type 2 Diabetes originating from obesity and a sedentary lifestyle.

A more accurate analysis of the media coverage of an epidemiological phenomenon needs to be handled in the same way studies are. While an epidemiological study results may reflect the true effect of an exposure to the development of the outcome under investigation, it should always be considered that the findings may in fact be due to an alternative explanation [8].

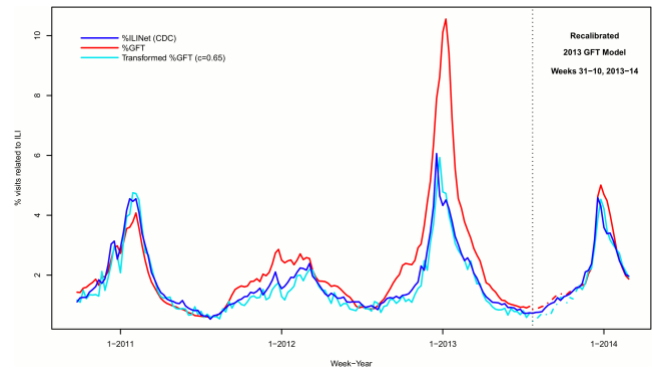


Figure 3: Incorrect estimations of the Google Flu Trends (in red) based on online queries, against the CDC (in dark blue) for the influenza season of 2012/13 [10].

3 THE TREND OF CHILDHOOD OBESITY IN WORLDWIDE NEWS

In the following section we focus on a specific Public Health priority, the well-known epidemiological case of childhood obesity that is of a major EU concern. To do that we will use Event Registry (ER), customized by the MIDAS Horizon 2020 project, funded under ‘Big Data supporting public health policies’ to develop a Big Data platform that facilitates the utilization of healthcare data, making that data amenable to enrichment with open and social data [1]. The Event Registry system collects and annotates in real-time news articles published by over 100,000 news publishers worldwide [4]. It provides the user with public health news articles in more than 10 languages as well as world events mentioned in these articles, permitting to explore what is currently being reported about in the media worldwide. ER can (a) identify and download news content from publicly available news sources, (b) analyze and semantically enrich the articles regardless of the language, (c)

combine the articles that report about the same event into a single event, (d) extract the relevant event information and (e) make all information searchable [7]. The worldwide health monitoring potential of this tool was discussed in [12] in the context of public health decision-making support, in particular as an epidemic / pandemic intelligence tool [15].

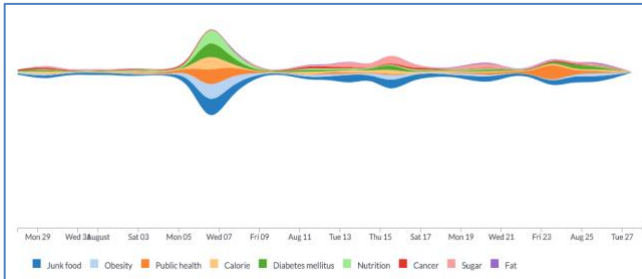


Figure 4: The concept trends associated with the query “childhood obesity” over the worldwide news in the past three years, highlighting related topics such as cancer, diabetes, nutrition, junk food or sugar.

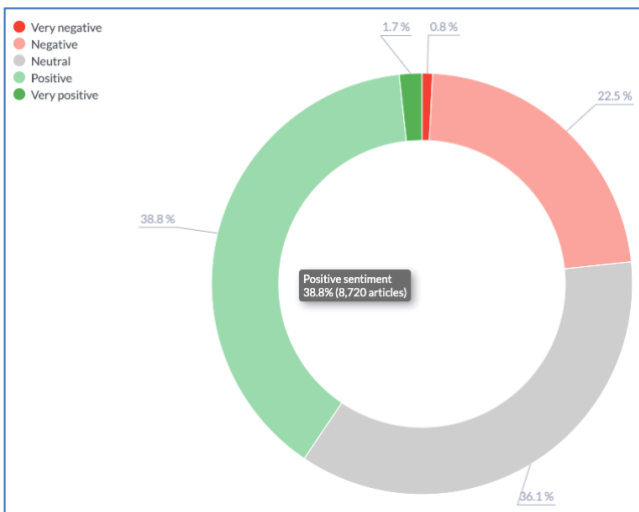


Figure 5: The sentiment analysis of worldwide news relating to “diabetes mellitus type 2”, with a 38.8% of positive sentiment identified highlighting the good results of the continuous efforts done to overcome this problem of modern society.

The user of ER can, query the dataset of worldwide news on “childhood obesity” to explore the trends and major concepts related with this query through the mentions in worldwide news (see Figure 4). Trending information is computed by comparing how frequently individual concepts/categories are mentioned in the articles. By default, trends are computed by comparing the total number of mentions of a concept in the last two days compared to the number of mentions in the two weeks before. The trend for each concept is computed as the Pearson residual. The returned concepts are the ones that have the highest residual [9]. The sentiment analysis in Figure 5 over the topic “diabetes mellitus type 2” serves

as a base of discussion to another angle on other aspect of news bias in the health domain: the sentiment expressed over news on a disease. The example shows that, although the immediate negative sentiment over such a topic, the positive sentiment can be identified in the good results from the continuous effort on fighting such a modern societal problem. In Figure 6 we show a screenshot of the pie chart of categories of worldwide news associated to the query “childhood obesity”. In that visualization module we observe that the media coverage of school meals nutrition in the context of public health was only 0.36% of all the news about childhood obesity during 2018. On the other hand, there was a coverage of 4.42% on child welfare in the context of the Society category, while the coverage on the business about sweeteners was 0.94%.

4 IMPACT IN PUBLIC HEALTH DECISION MAKING

Research in the field of automatic detection of news media bias shifts its attention to sentiment analysis and opinion mining in the news. Most sentiment and opinion mining analysis has been done on very subjective texts like product launch reports, movie reviews or blogs, where the opinion of the author is expressed freely in a very subjective and biased way. Recently, sentiment analysis of news articles, where an opinion of a journalist should not be present, is getting more attention. An example of such an approach is the news media bias analysis of finding over and under-stated facts of a particular news outlet [11].

In the example of Figure 1, the event of the Ebola outbreak is identified immediately after the news articles that report about it are collected. One can explore the evolution of the news publishers awareness of the epidemics in a timeline by looking at the related news articles represented in a world map, as they were identified or updated during a selected period of time. ER can find articles and events related to a particular entity, topic, date, location or category, as well as measure their impact on social media (Twitter).

For each world event, ER is able to provide extensive information. Its event clustering permits us to distinguish between subtopics and perspectives in the stories relating to that particular event. Beside the whole list of articles that describe the event, the user can also see the list of top concepts, the trending of the articles, and subcategories it falls into. If a user is not interested in all events, (s)he can easily limit the list of articles and events displayed based on specific interests, location, etc.

In particular, Governmental Institutions are interested on what is the public opinion over public health related legislation such as the Sugar Tax, where sentiment analysis can be an approach with great potential. This and related measures were applied in several EU countries to fight against childhood obesity and consequential diseases such as diabetes mellitus type 2. The role of social media is of great importance in this context but also contributes with a lot of noise. A query on “Sugar Tax” in ER allows us to identify not only the news articles about this measure but also the social media mentions of those news items, permitting the user to estimate the impact of the issue in the population.

A new version of ER is in development, integrating an automated annotator assigning MeSH Heading descriptors to snippets of free text provided by the user. This will allow the annotation of the news articles with those useful classes, designed to enhance the exploration of biomedical research in the well-established search engine PubMed. The latter is part of the workflow and know-how of most health professionals. The new ER instance, built in the context of the MIDAS project will allow those health professionals to explore the worldwide and local news using the MeSH Heading descriptors much as they use them in their searches in PubMed. The new system will also provide MeSH Heading-based visualization modules such as the one discussed above and in Figure 6, providing an efficient perspective of the news coverage over subtopics of the search query, allowing for a fast identification of potential news bias, designed for the health domain, to support decision making in public health.

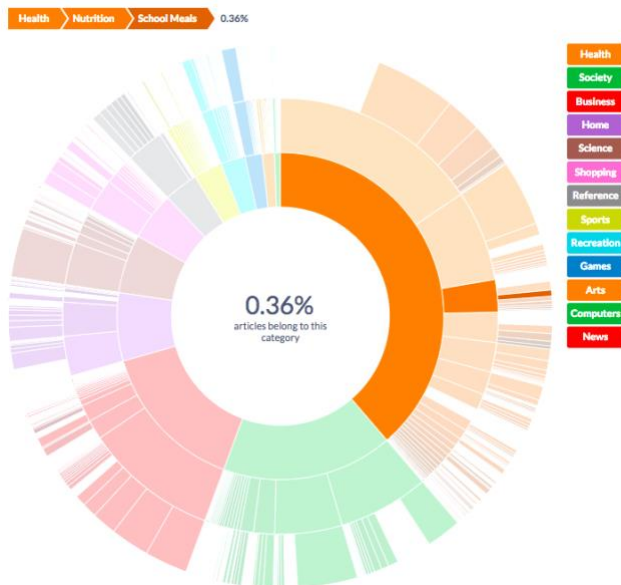


Figure 6: The categories associated to the query “childhood obesity” over the worldwide news during 2018. This shows that the media coverage of school meals nutrition in the context of public health was only 0.36% of all the news about this topic.

5. CONCLUSIONS AND FURTHER WORK

Media bias is a universal concern. Despite the fact that newspapers and reporters or journalists are supposed to provide the readers with impartial, objective, unbiased and reliable information, the reality is somehow different. Every news story has a potential to be biased. Every news story has a potential to be influenced by the attitudes, cultural background, political and economic views of the journalists and editors. In this paper we have discussed several angles of the news media bias in the context of health-related news with a particular focus on epidemiology. We have also presented some approaches and tools that permit data exploration and can help balancing the information in worldwide media coverage.

This includes some of the ER visualization modules which can help us to explore what is the news coverage of a certain health-related measure feeding the general population awareness. Moreover, the upcoming Event Registry instance built in the context of MIDAS will be offered with an automated MeSH classifier of text. This open source service will be used to classify the news with the MeSH headings and will enable queries using these. That will permit researchers to be closer to the information they are looking for, using a similar workflow as the one used in queries over PubMed. That will be useful to health professionals in particular that use PubMed daily in their biomedical research, and fully understand the usability of the MeSH descriptors. We will further analyze news bias when the first results of early adopters are available. Further work also includes the bias detection through advanced text mining techniques. This includes the analysis of the used metrics that can be itself a potential bias generator.

ACKNOWLEDGMENTS

We thank the support of the European Commission on the H2020 MIDAS project (GA nr. 727721).

REFERENCES

- [1] Brian Cleland et al. 2018. Insights into Antidepressant Prescribing Using Open Health Data. *Big Data Research* doi.org/10.1016/j.bdr.2018.02.002
- [2] D. Brossard et al (2013). Science, new media, and the public. *Science* 339.6115: 40-41.
- [3] J. Ginsberg et al (2009). "Detecting influenza epidemics using search engine query data." *Nature* 457.7232: 1012.
- [4] M. Grobelnik and G. Leban. Eventregistry's health panel. eventregistry.org. Accessed: 2019-08-22.
- [5] D. Kahneman et al (1982) *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- [6] J. Leask et al (2010). Media coverage of health issues and how to work more effectively with journalists: a qualitative study. *BMC public health* 10.1: 535.
- [7] G. Leban, B. Fortuna, J. Brank, M. Grobelnik (2014). Event Registry – Learning About World Events From News, WWW 2014, pp. 107-111.
- [8] C. H. Hennekens, J. E. Buring(1987). *Epidemiology in Medicine*, Lippincott Williams & Wilkins.
- [9] C. Manning et al. (2008). *Introduction to Information Retrieval*. Cambridge University Press 2008, pp. 269-273.
- [10] L. J. Martin, X. Biying and Y. Yasui (2014). Improving Google flu trends estimates for the United States through transformation. *PLoS one* 9.12: e109209.
- [11] S. Park, S. Kang (2009). "NewsCube: delivering multiple aspects of news to mitigate media bias." *Proceedings of the 27th international conference on Human factors in computing systems*.
- [12] J. Pita Costa et al. (2017). Text mining open datasets to support public health. *WITS 2017 Conference Proceedings*.
- [13] P. Snyder et al (2010). *Science and the media: Delgado's brave bulls and the ethics of scientific disclosure*. Academic Press.
- [14] WHO (2019). Ebola outbreak in the Democratic Republic of the Congo declared a Public Health Emergency of International Concern - News release 17 July 2019, <https://www.who.int/news-room/detail/17-07-2019-ebola-outbreak-in-the-democratic-republic-of-the-congo-declared-a-public-health-emergency-of-international-concern> Accessed: 2019-08-22.
- [15] WHO (2017). Epidemic intelligence definition. www.who.int/csr/alertresponse/epidemicintelligence/en/. Accessed: 2017-09-05