Misinformation from Web-based News Media? Computational Analysis of Metabolic Disease Burden for Chinese

Angela Chang(✉)

University of Macau, Macau SAR, China
wychang@um.edu.mo

Abstract. Raising the question of what and how metabolic syndrome is covered through the renowned online press, hence, this study attempted to determine the ways in which the newspapers selects chronic diseases and how the syndrome informs the public of causation and adverse outcome. Web-based news platforms from four leading newspapers in mainland China were sampled for machine computational analysis. Two prominent metabolic diseases, its causation and causal inference were identified in a total of 16,005 articles in the past 10 years. Descriptive statistics is reported while a word cloud displays the entire news content as a graphical representation of word frequency with recurring expressions. A dendograms was summarized to provide a richer context to interpret association between metabolic syndrome and causal inference to verify if misinformation was spread through the renowned online press. Results indicate that news stories tend to provide equivocal descriptions of metabolic disease burden while implying explicitly that alcohol, tobacco, or genes might be the main cause of adverse outcomes for metabolic syndrome. This study emphasized the importance of taxonomy of coding, causal assessments of news stories to inform debates on addressing causal inferences, message design, and unique frames in the media content. It concludes with limitation of datamining and directions for future research.

Keywords: Diabetes · Alcohol overconsumption · Causal inference · Misinformation type

1 Introduction

An important topic in population health studies has been metabolic disease since the major cause of death and the risk factors for metabolic syndrome affect disability-adjusted life span, especially for women in Mainland China [1]. The most common form of metabolic disease is type 2 diabetes, which accounts for 95% of diabetes cases in adults aged 18–99 years old. In China, 120 million adults have been diagnosed with diabetes as of 2017 [2]. The rate of metabolic disease continues to escalate, which calls into question the media's role in covering and promoting knowledge about metabolic syndrome for treatment and prevention (e.g., [3]). Hence, more news coverage about metabolic syndrome is assumed to raise awareness of the disease, while promoting knowledge of causation and causal inference for metabolic disease to the general public.

M. van Duijn et al. (Eds.): MISDOOM 2020, LNCS 12259, pp. 52–62, 2020.
https://doi.org/10.1007/978-3-030-61841-4_4

1.1 Misinformation Analysis

The problem of population health could be worsened by health-related misinformation from news media. The concept of misinformation refers broadly to honest mistakes, bias, unknown inaccuracies, uncertainties, and ignorance as distinct from disinformation [4]. Prior studies surveyed misinformation in which it was implied that something or someone might be the cause of an adverse outcome but less emphasis in these studies was placed on the function of explicitly or implicitly stating the likely cause of an outcome [5]. Hence, this study proposes to examine what is being published in web-based news media by studying the way a likely cause of an outcome is explicitly or implicitly presented. The aim of the present study was to provide reliable and up-to-date information on the prevalence and associated factors of metabolic syndrome. It allows the researchers to examine how different news outlets discuss the diseases, its cause, and the causal inferences that typically accompany it. The computational analysis improves empirical study for rendering news selection in a more efficient and better-understood process. In addition, it established a preliminary understanding of the dominated news values and the relevant risks for potentially leading news audiences to incorrect assumptions about how metabolic disease develops.

Computational analysis of misinformation has been popular in computational communication studies. Examples of text analysis applications include the examination of news stories from CNN and BBC by computer-assisted software to study the section prominence of the SARS crisis [6]. A more recent study examines how social media became a channel to terminate rumours in China by influencing the perceived credibility of health information, in contrast to the popular impression that social media degenerates into a rumour mill [7]. A similar approach adopted the concept of infodemiology for studying the science of information distribution and determinants of information in an electronic medium, with the goal of informing public health and policy [8].

Recently developed computer-aided analysis for identifying health issues has become more sophisticated in ascertaining metadata collection, media effects, particularly by integrating data engineering and machine learning [9–11]. Earlier studies argue that several news values dominate not just in the popular press laymen can understand but also in quality newspapers [12]. Based on all above mentioned concerns by employing computer-aided text processing and automatic content analysis to examine metabolic syndrome and its burden for potential impact on perception, three research questions (RQ) were formulated:

The first RQ focused on how much coverage was devoted to metabolic syndrome in mainland China press for the past 10 years? To be specific, what is the trend of the metabolic diseases with causes and casual inference covered in Chinese newspapers?
The second RQ explored how the metabolic diseases covered by considering the extent to which news selections may be misinforming?
The third RQ examined how the causal inference of metabolic syndromes and causal effect were employed to inform the public for outcome? Did the coverage provide misinformation to interpret association between metabolic syndrome and causal inference?

2 Methodology

In line with the research objectives, the platform of DiVoMiner is deployed and proves ideal at answering research questions concerning changes in the quantity of coverage over time and across newspapers. The platform is equipped with a powerful keyword retrieval function in both English and Chinese from which researchers can easily develop and identify text units containing one keyword or a combination of phrases. With optional filtering of articles, the tailored task powered by DiVoMiner is coming with a web browser extension through which news articles can be categorized. DiVoMiner is operated by the Boyi Data technology company[1] and has been considered as a powerful machine computational tool in big data collection and analysis in real time. Moreover, the embedded software in DiVoMiner provides automated content analysis while validation and reliability check are also functional. Thus, machine computational method contributes valuable clues for examining news selection in a more efficient, effective, and transparent process.

2.1 Sampling

Four Chinese newspapers that publish an electronic version on their websites with free access from January 1, 2010 to December 31, 2019 were selected for this study: *People's Daily News* (PD) (人民日報), *Beijing Evening News* (BE) (北京晚报), *Southern Metropolis Daily News* (SMD) (南方都市報), and *Guangzhou Daily News* (GD) (廣州日報). A communist party-owned outlet, PD has provided direct information from the central government since 1948 and boasts the second largest circulation of any Chinese newspaper, with 2.52 million copies sold daily [13, 14]. A state-run tabloid outlet, BE from Beijing, was founded in 1958 and has the largest-circulation of any newspaper in the capital [14]. In comparison, SMD has been a market-driven newspaper since 1995 and circulates in the Pearl River Delta with 1.40 million readers; SMD has gained extensive attention for daring to challenge institutional restrictions [15]. The official newspaper of the Guangzhou municipal party, GD was established in 1952, with the highest circulation in the metropolitan area with approximately 1.85 million copies sold daily [14].

2.2 Computational Taxonomy

DiVoMiner assisted in the implementation of the coding taxonomy to measure the concept of misinformation. The process involves pilot coding, subsequent modification of the coding scheme, and double coding adapted from earlier studies [9, 10, 16]. Special emphasis is placed on the assumptions of all emerging epidemic of metabolic syndrome that underlies all causation, casual inferences, and the news languages used in formulating these assumptions. An exploratory test was run several times by computer-aided automatic scanning and human manual assessment to ensure accuracy and relevance of the test data.

One of the main tasks is to capture both explicit and implicit assertions of casual linkages. To be specific, an explicit link is given when the assertion is that X causes, triggers, leads to, or generates Y, or that Y originates from or is attributed to X. The explicit

[1] BoYiData Homepage, https://www.boyidata.cn, last accessed 2020/01/01.

association between two elements is formulated using concepts or terms that signify causation, causal inferences, or causal language and effects. Additionally, implicit causal inferences do not use such specific terms but still suggest or insinuate causation. Thus, analogical reasoning on linguistic regularities for developing a more comprehensive and reliable coding which was behind a dictionary-approach to measure the concept of misinformation was considered. Moreover, corresponding to previous studies in conceptualizing mediated associations was also taken into consideration. Eventually, the code taxonomy started with the 3 types of metabolic diseases, 14 independent factors of causation contribute to metabolic diseases, and 11 types of causal inference [9, 17]. Codes are typically data-driven and consistency is less of an issue.

Our 6 keywords and 14 phrases resulted from several pilot tests that created a logical phrase the machine can assist in analyzing news data. This process involved searching for diseases names, multiple terms, phrases, concepts, and alternative terms for more precise and reliable results [9, 18]. It is worth to note that Chinese words and phrases into the meaning group have a unique morphological system than English. It is required to select different settings on semi-prefixes or semi-suffixes for computational analysis [9, 10, 17, 18]. For instance, the term of diseases frequently employed in describing metabolic syndrome were: 代謝病 (metabolic disease), 糖尿病 (diabetes), 肥胖 (obesity), 慢性腎臟病 (chronic kidney disease), 肝衰竭 (liver failure), 痛風 (gout); multiple terms of metabolic syndrome in Chinese appeared in the news stories included the following examples: 代謝性疾病, 代謝病, 代謝疾病, 新陳代謝失調症, 代謝缺陷, 代謝失調, 代谢障碍, 代謝紊亂, 代謝異常, 代謝旺盛; to some extent, these Chinese terms were often synonymous with four English phrases of metabolic disorder, metabolic defect, metabolic abnormality, or metabolism.

3 Results

The tailored platform creates researcher-defined codes entered manually and norm files are generated based on frequency analysis of keywords, phrases and content categories. The development over time of metabolic disease coverage shows a clear trend of increasing attention from journalistic decisions. Overall speaking, the highest surge of metabolic related coverage in China is observed in 2019 and 2018, but it shows a big decrease of coverage in 2014 and 2010. *People's Daily* has the least coverage on the metabolic disease with causes and casual inference which illustrated a particular discourse strategy from their editorial strategy. In comparison, *Guangzhou Daily News* starts its increased report of metabolic disease with causes and casual inference in 2017. While *Southern Metropolis Daily News* shows a sharp decline in the number of coverage in 2014, *Beijing Evening News* also has a big decrease in 2016 but no coverage data found in 2010. Figure 1 shows an overview of the development of metabolic diseases coverage in four newspapers in Mainland China for the past 10 years.

A total of 16,005 articles that covered metabolic diseases in conjunction with causation and causal inference were identified, corresponding to an average of 1,601 stories ever year., with an annual low of 906 to a high of 2,244 stories. *Southern Metropolis Daily News* and *Guangzhou Daily News* mirrored this overall development, while *Beijing Evening News* and *People's Daily* provided comparatively less coverage, below the

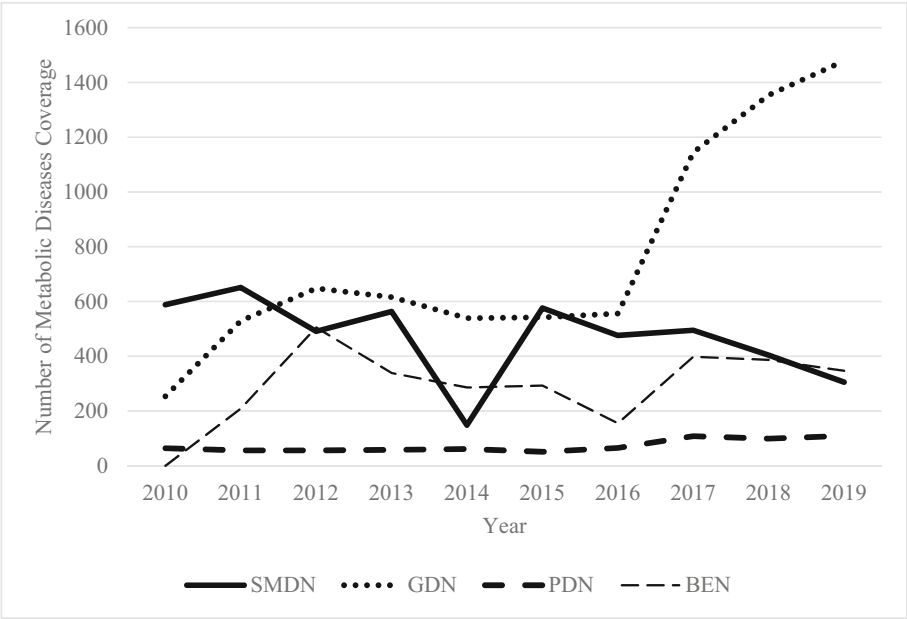

Fig. 1. Development of metabolic diseases coverage in four newspapers in Mainland China from 2010–2019. Note: SMDM stands for Southern Metropolis Daily News; GDN stands for Guangzhou Daily News, PDN stands for People's Daily News, and BEN stands for Beijing Evening News.

average of 400 stories per year. The development of metabolic diseases coverage in Chinese newspapers shows a clear trend of increasing attention over time ($x^2 = 1895.59$, df $= 40$, $p < .001$). Turning to individual newspapers, one of the nationwide newspapers, *People's Daily*, had the least coverage, while *Guangzhou Daily News* had the highest coverage of metabolic diseases and burden. Table 1 displays the distribution of metabolic diseases related articles in news media in China from 2010-2019.

For casual inference of metabolic diseases, burdens are presented in four newspapers with statistically significant difference ($x^2 = 149.08$, df $= 20$, $p < .001$). Overall, the consequence of loss of wealth outnumbered the other attributions in all newspapers. Employer cost was emphasized in the *Southern Metropolis Daily,* besides the main consequence of individual's wealth loss. As links of consequences at the environmental level, the frame was prominent in displaying societal costs explicitly in news media. Table 2 displays the casual inference of metabolic diseases covered in four newspapers in mainland China spanning 10 years, from 2010 to 2019.

A word cloud is composed of single words that denote a graphical representation of word frequency [10]. The attention of news coverage was given to metabolic syndrome constitutes 158 words from all press stories. Two of the most prominent terms highlighted are "diabetes" and "obesity and overweight", followed by "alcohol", "government", and "mental stress & burden"; next to the above-mentioned terms, the three most frequently used terms are: "smoking", "economic loss & wealth reduction", and "genes". Figure 2

Table 1. Number of metabolic diseases related articles in four Chinese web-based news media from 2010–2019

	SMDN 4697 (%)	GDN 7662 (%)	PDN 727 (%)	BEN 2919 (%)	All 16005 (%)
2010	588 (12.5)	254 (3.3)	64 (8.8)	0 (0.0)	906 (5.7)
2011	651 (13.9)	529 (6.9)	56 (7.7)	209 (7.2)	1445 (9.0)
2012	491 (10.5)	648 (8.5)	56 (7.7)	505 (17.3)	1700 (10.6)
2013	563 (12.0)	616 (8.0)	58 (8.0)	339 (11.6)	1576 (9.8)
2014	148 (3.2)	539 (7.0)	61 (8.4)	286 (9.8)	1034 (6.5)
2015	576 (12.3)	542 (7.1)	51 (7.0)	293 (10.0)	1462 (9.1)
2016	476 (10.1)	556 (7.3)	65 (8.9)	155 (5.3)	1252 (7.8)
2017	495 (10.5)	1145 (14.9)	108 (14.9)	398 (13.6)	2146 (13.4)
2018	404 (8.6)	1354 (17.7)	99 (13.6)	387 (13.3)	2244 (14.0)
2019	305 (6.5)	1479 (19.3)	109 (15.0)	347 (11.9)	2240 (14.0)

Note. SMDM for Southern Metropolis Daily News; GD stands for Guangzhou Daily News, PD stands for People's Daily News, BE for Beijing Evening News

Table 2. Casual inference of metabolic diseases related articles in four Chinese web-based news media from 2010–2019

	SMDN 1625 (%)	GDN 1994 (%)	PDN 303 (%)	BEN 753 (%)	All 4675 (%)
Loss of wealth	684 (42.1)	899 (45.1)	103 (34.0)	313 (41.6)	1999 (42.8)
Loss quality life	233 (14.3)	459 (23.0)	71 (23.4)	221 (29.3)	984 (21.0)
Employer cost	302 (18.6)	273 (13.7)	53 (17.5)	91 (12.1)	719 (15.4)
Societal cost	209 (12.9)	209 (10.5)	37 (12.2)	71 (9.4)	526 (11.3)
Unemployment	181 (11.1)	115 (5.8)	32 (10.6)	41 (5.4)	369 (7.9)
Premature death	16 (1.0)	39 (2.0)	7 (2.3)	16 (2.1)	78 (1.7)

Note. SMDM stands for Southern Metropolis Daily News; GD stands for Guangzhou Daily News, PD stands for People's Daily News, BE stands for Beijing Evening News.

displays a word cloud visualization of metabolic disease with the most frequently used causation and causal inferences.

An deductive textual analysis revealed that the coverage of metabolic diseases included diabetes mellitus (50.4%, $n = 8{,}906$), followed by overweight/obesity (43.3%, $n = 7{,}662$), and chronic kidney diseases (4.2%, $n = 738$). Frequency ranking and cluster analysis were further conducted for graphical representation of the hierarchical tree in order to verify if misinformation was spread through the renowned online press. The causations and casual inferences of metabolic diseases were grouped into three clusters

Fig. 2. Lexical groupings of overall metabolic disease covered in four leading Chinese web-based newspapers, 2010–2019

by integrating news languages used. Specifically, one cluster integrated six variables performing better than the others; the news expression included six themes of news information: governmental management, mental status, alcohol consumption, genetic concern, smoking tobacco, and economic loss. To illustrate, a dendrogram shows the hierarchical relationship between objects employed. It is most commonly created as an output from hierarchical clustering analysis to allocate causes and causal inferences by showing items are similar.

The result of the hierarchical relationship is illustrated in Fig. 3 dendograms in which there are five clusters. One cluster (A) combines three causation and two causal inferences (family condition, drugs except alcohol and tobacco, economic loss, social & economic system, and premature death); a second cluster (B) combines two causation and one causal inference (lack of exercise, wrong diet, and loss of quality life); a third cluster (C) groups one causation and three causal inferences (societal cost, unemployment, environmental pollution, and cost from patients' employer); a fourth cluster (D) combines two causes and one causal inferences (smoking, wealth reduction, and genes); and a fifth cluster (E) shows three causes are very similar (alcohol, mental stress, and government). In the dendrogram, the width of the dendrogram indicates the order in which the clusters were joined. A more informative dendrogram was also created to show where the widths reflect the distance between the clusters B and C, or D and E is close, as is shown in Fig. 3. In the findings, the dendrogram displays that a big difference is between cluster A versus that of cluster E. Overall, a dendograms provides a richer context of misinformation in interpreting associations between causation and causal inference of metabolic syndrome.

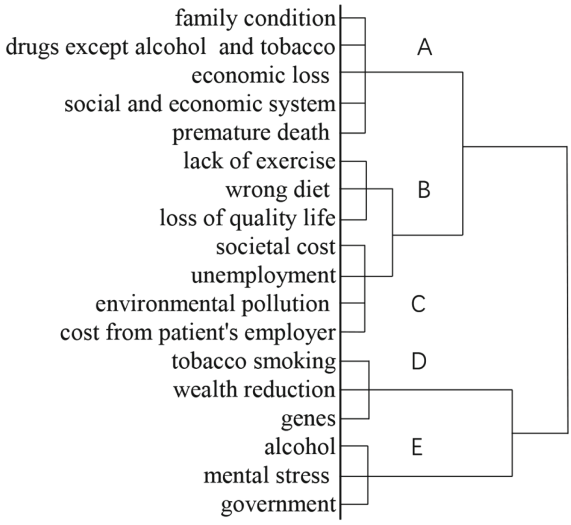

Fig. 3. Dendograms of clustering causation and causal inference of metabolic syndrome from Chinese web-based news media

4 Discussion

News decisions were influential in awareness or knowledge of the public might be informed and misinformed. The trend of coverage for the past ten years indicates that efforts to enhance awareness of metabolic diseases were mainly implemented by the web-based SMD periodical, followed by GD. In comparison, the story of metabolic disease seems to not be newsworthy enough for PD and BE, although both newspapers provide direct information on relevant policies from central government in Beijing. The trends observed over the past ten years could be part of a larger cycle from individual's press and may be persist into the future coverage of metabolic diseases. Even our forecasts appear to be increasing coverage of metabolic diseases from regional newspapers, it should be cautious about forecasts that are more than ten years of time periods. Because trends can be volatile, researchers should usually only forecast 2 periods into the future pattern for Chinese readers.

A word cloud of the entire content allows researchers to identify recurring expressions. The most frequently used phrases for causal inference expression emphasized an individual's economic loss and wealth reduction, while the recurring causation phrases focused on four diverse expressions regarding an individual's condition (i.e., alcohol consumption, mental stress and burden, genetic disposition, and tobacco smoking). For additional insight into this finding, that "alcohol" is the most popular term highlighted reflects a concerted effort by journalists as well as China's government to address alcohol overconsumption as a health issue. We wonder whether alcohol is really such a large problem in China, or whether this is a preferred discourse strategy relative to discussion of diabetes.

The causal inference of type 2 diabetes can increase risk of heart disease and stroke and cause major health complications, particularly in the smallest blood vessels in the

body, which nourish the kidneys, nerves, and eyes [2]. In comparison, the equivocal information provided by the general news media tends to imply that alcohol, tobacco or genes might be the main cause of metabolic disease. As a result of news selection and salience, some facts are highlighted (e.g., diabetes is linked to adverse outcomes like kidney failure, heart disease, and stroke), while some other facts are overlooked (e.g., diabetes contributes to cardiovascular disease and other health complications).

In this study, Chinese news's mentions of specific keywords – such as childhood obesity in connection with adult aging, incidence of fatty liver, hypertension, hyperlipidaemia, and type 2 diabetes – are limited to one paragraph and not necessarily linked to what follows or precedes them. The selection of ignorance in the news content also lacked details of gender issues in covering causal inference of metabolic syndrome. For example, hypertension was the most prevalent component of metabolic syndrome in Chinese males, while the most prevalent component of metabolic syndrome for Chinese females was central obesity [17, 19]. Although the scientific-based and behavioral-focused study has indicated that the age, urolithiasis, hyperuricemia, coronary artery disease, thiazide drugs intake, family history of diabetes, and hypertension were all significantly associated with an increased risk of metabolic diseases in China [19]. Unfortunately, the casual links of metabolic diseases in the scientific-based information were not reflected sufficiently in overall web-based news media. The published news pieces provides fragmented casual information, rather covering a comprehensive explanation of likely causes of an outcome of metabolic disease burden. This type of misinformation existed, particularly in state-owned news media which demonstrated its journalistic decisions on health news selection. In cluster analysis, another misinformation was evidenced in which news pieces frequently lack explicit explanation of causal inferences; instead, the news stories provided equivocal and uncertainty descriptions of metabolic syndrome for potentially leading news readers to generate incorrectly assumptions about how metabolic diseases were developed and treated.

4.1 Limitations

Several limitations are noteworthy in computational process and news data analysis. First, the result does not show absolute frequency, but relative frequency due to confined data crawling. It can be used for longitudinal comparisons of obtained frequencies to previous saved files. Secondly, there is no clear-cut criteria for the coding taxonomy developed to distinguish misinformation from proper journalism. Yet, the current coding schemes are still insufficient to recognize and filter content about metabolic syndrome circulated online. Last but not least, a word cloud is typically used to depict keyword metadata to visualize text but the analysis of word cloud lack of context in visual representation. Hence, one way to enrich the analysis of coverage bias would be to compare the news coverage with the actual metabolic disease prevalence in those locales to see whether coverage reflects the severity of the problem on the ground. Similarly, the casual inference could be compared to the ranking of causes as well as potential adverse outcomes of metabolic syndrome by epidemiologists.

5 Conclusion

The present work is about the news talking about the syndrome of metabolic diseases, but not about the syndrome itself. This paper aims at investigating the web-based news selection of metabolic syndrome, particularly on diabetes type 2 as a possible example of misinformation. Therefore, this present work explored a data set of Chinese news concerning metabolic disease information in investigating the information in order to verify if misinformation was spread through the renowned online press in China. One of the contributions was to provide reliable and up-to-date information on the prevalence and associated factors of metabolic syndrome.

The method of analysis is based on data mining by examining how machines and algorithmic systems are increasingly utilized to make complex judgements regarding unstructured data. In today's digital landscape, content such as texts, images, and recorded sounds are increasingly subjected to automatic or even semi-automatic processes of classification. When put to action, automatic content analysis methodology is improving validity and reliability in separating biased/unbiased forms of communication and is used to secure the value, authenticity, origin, and ownership of content. Although the limitations and analysis are still quite broad, the study can be a good example of a source of misinformation and possible impacts on society. Just as this study put taxonomy of news value by applying computational collection and analysis, the result explores the extent to which researchers would revisit list of news values by given the challenges faced by the emergence of web-based media today.

Acknowledgements. The views and opinions expressed in this article are from the individual author and not from sponsor organization. This study was funded by the University of Macau grant MYRG2018-0062-FSS & Higher Education Bureau of Macao grant (HSS-UMAC-2020-09). The author would like to thank Professor Peter J. Schulz, Professor Angus W. H. Cheong and anonymous reviewers for their valuable comments.

Conflicts of Interest. The authors declare no conflict of interest. The founding sponsor had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Forouzanfar, M.H., et al.: Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the global burden of disease study 2015. The Lancet **338**(10053), 1659–1724 (2016)
2. CDS Homepage. https://www.idf.org/our-network/regions-members/western-pacific/diabetes-in-wp.html. Accessed 25 May 2019
3. Dan, V., Raupp, J.: A systematic review of frames in news reporting of health risks: characteristics, construct consistency vs. name diversity, and the relationship of frames to framing functions. Health Risk Soc. **20**(5-6), 203–226 (2018)
4. Søe, S.O.: Algorithmic detection of misinformation and disinformation: Gricean perspectives. J. Documentation **74**(2), 309–332 (2018)

5. Rich, P.R., Zaragoza, M.S.: The continued influence of implied and explicitly stated misinformation in news reports. J. Exp. Psychol. Learn. Mem. Cogn. **42**(1), 62–74 (2016)
6. Yan, T., Stewart, C.M.: Framing the SARS crisis: a computer-assisted text analysis of CNN and BBC online news reports of SARS. Asian J. Commun. **15**(3), 289–301 (2005)
7. Zhao, L., Yin, J., Song, Y.: An exploration of rumor combating behavior on social media in the context of social crises. Comput. Hum. Behav. **58**, 25–36 (2016)
8. Eysenbach, G.: Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J. Med. Internet Res. **11**(1), e11 (2009)
9. Chang, A.: Digitalized news on non-communicable diseases coverage-what are the unhealthy features of media content induced for Chinese? In: Staab, S., Koltsova, O., Ignatov, D.I. (eds.) Social Informatics: Lecture Note Series in Computer Science, vol. 11186, pp. 29–39. Springer Nature, Switzerland (2018)
10. Chang, A., Hu, J.Y., Liu, Y.C., Liu, M.: Data mining approach to Chinese food analysis for diet-related cardiometabolic diseases. In: 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), IEEE Xplore digital library. https://doi.org/10.1109/icdew.2019.00-29
11. Dutta, M.J.: Content effects: personal and public health. In: Rössler, P., Hoffner, C.A., Zoonen, L.V. (eds.) The International Encyclopedia of Media Effects. John Wiley & Sons, New Jersey (2017)
12. Harcup, T., O'neill, D.: What is news? news values revisited (again). Journalism Stud. **18**(12), 1470–1488 (2017)
13. Hu, S., Zhang, Z.: The discursive construction of party identification of Chinese party-owned media by means of social media—Setting "xiake_island" as an example. Commun. Soc. **48**, 57–91 (2019)
14. Velker, M.: Top 10 daily newspapers in China. http://www.china.org.cn. Accessed 05 May 2019
15. Zhang, W.: Online UGC, public power conflicting issues and media publicity: a study of the routinely use of online UGC in China's market-driven newspaper. Commun. Soc. **48**, 131–162 (2019)
16. Kaefer, F., Roper, J., Sinha, P.: A software-assisted qualitative content analysis of news articles: example and reflections. In: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, vol. 16, no. (2), pp. 1–55 (2015)
17. Li, R., et al.: Prevalence of metabolic syndrome in mainland China: a meta-analysis of published studies. BMC Public Health **16**, 296–305 (2016). https://doi.org/10.1186/s12889-016-2870-y
18. Boumans, J.W., Damian, T.: Taking stock of the toolkit: an overview of relevant automated content analysis approaches and techniques for digital journalism scholars. Digit. Journalism **4**(1), 8–23 (2016)
19. Lan, Y., et al.: Prevalence of metabolic syndrome in China: an up-dated cross-sectional study. PLoS ONE **13**(4), e0196012 (2018)