Contents lists available at ScienceDirect

ELSEVIER



Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

All are not created equal: Method descriptions in an epidemiology publication differ among media summaries – A case study comparison

Lilianne Samad, J.E. Reed

Research & Development, Bayer Crop Science, Chesterfield, MO, USA

ARTICLE INFO	A B S T R A C T		
Keywords: Science communication Public understanding of science Media (news, digital) Science journalism Mass media Public health	It is common to see mass media headlines about health-related topics in traditional and online news outlets, as well as on social media platforms. What a consumer might not realize is that often these headlines are a distillation of results reported in epidemiologic publications. Journalists make decisions about what information to include and exclude, hopefully without compromising the main conclusions. In this exercise, sixty-three media articles that summarized one peer-reviewed journal publication (Zhang et al., 2021) describing results from a cohort study on coffee and tea consumption and risk of stroke and dementia were compared to determine the consistency of details among them. The most heterogeneity was observed in whether articles compared results with other literature. There was some variation in inclusion of a measure of frequency within the study population, and in details describing measurement of exposure. However, most of the articles were consistent in either including or excluding other methodological details in the main text. The results of the present comparison have implications for readers, researchers, and journalists. Readers must know that media summaries of peer reviewed studies are just that – summaries. It is likely that some information from the original source is not represented by the article, and that additional information might be necessary to craft an informed opinion on a given topic.		

Introduction

It is common to see mass media headlines in traditional and online news outlets, as well as on social media platforms, about health-related topics. Headlines about what you should or should not eat, what will increase or decrease your risk of developing a particular disease, etc., are a part of life. What a consumer might not realize is that often these headlines are a distillation of results reported in epidemiology publications. Epidemiology is the study of the distribution of diseases in populations, and the factors that impact these patterns [1]. It is the job of epidemiologists to publish the results and a discussion of their studies that are designed to better understand health-related topics at the population level. It is then the job of some journalists to summarize the key findings for the general public from these peer-reviewed epidemiology reports. Journalists need to make decisions about what information to include, and what can be excluded from an article, hopefully without compromising the main conclusions the epidemiologists reported. How closely a media article's summary matches the results reported in a publication might have implications for the message conveyed to, and understood by, the reader. Which details of how a study was designed and conducted that are included in a media article can impact the extent to which a reader is able to suitably evaluate the study results and implications for the public. This concept has been explored in publications which highlight the prevalence [2] and potential impact [4] of missing details that would provide context to reported results. Some draw attention to lacking details which might lead to inaccurate representation of the level of certainty implicit in the conclusions [7]; for this reason, others encourage caution when sharing results with the general public without also explaining the extent to which they have or have not been replicated and validated [6].

Ensuring that shared information is factual is only part of the equation; equally important is that the information is portrayed consistently among sources. Readers cannot achieve a comprehensive understanding of a primary publication by reading a summary in a media article if details reported in the media article are incomplete. It stands to reason that if there is inconsistency in reporting among media articles, then this suggests that readers are heterogeneous in the extent to which their understanding of the primary study is equipped to be comprehensive. The extent to which representation of epidemiology studies is consistent across media articles warrants further examination,

https://doi.org/10.1016/j.gloepi.2025.100188

Received 10 December 2024; Received in revised form 3 February 2025; Accepted 7 February 2025 Available online 8 February 2025

^{*} Corresponding author. *E-mail address:* jennifer.reed@bayer.com (J.E. Reed).

^{2590-1133/© 2025} The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

and defines the scope of the present study. To further explore this concept, the present exercise focused on a topic familiar to many people: coffee or tea consumption and health. An epidemiology article that was published recently was selected on the possible association between coffee/tea consumption and human health outcomes [8]. In this exercise, sixty-three media articles that summarized one peer-reviewed journal article [8] were compared to determine the consistency of details among the media articles. The inclusion of details such as exposure, outcome, experimental groups, epidemiologic statistics, and limitations were of specific interest in this exercise. Also compared was the location of such details (i.e., in the title, and/or the main text), as information included in the title was presumed to be more noticeable to a reader than that which appeared only in the main text.

Methods

The epidemiology publication by Zhang et al. was an analysis of data from a cohort study that assessed a possible association between selfreported tea and/or coffee consumption and diagnosed stroke and/or dementia [8]). Exposure to each coffee and tea was assessed by questionnaire. Health outcome data for stroke and dementia were captured from hospital records. The main conclusions supported an association between coffee and tea consumption and risk of stroke and dementia. This article was selected to be the focus of the present exercise because the number of references within media articles was sufficient to inform an analysis, and because tea and coffee are common exposures across an array of audiences who might consume content from like sources. Altmetric Explorer [9] was used to search for media articles describing the Zhang et al. publication. The search was conducted on 15 February 2023, and returned 271 news stories from 192 outlets. News stories were excluded if they were exact duplicates, not accessible, not relevant, or not written in the English language. 63 stories were included in the analysis

This study focused on ten details describing the methods that are helpful to understanding an epidemiology study. These details are consistent with, or relevant to, some of the items listed in the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) reporting guidelines, the aim of which was to provide recommendations for reporting epidemiology research [5]. The elements assessed in the present exercise represent the subset of recommended components that are viewed by the authors to be most useful to a nontechnical audience.

- (1) How the exposure¹ was defined
- (2) How the exposure was measured
- (3) How the comparison group was defined
- (4) How the health outcome was defined
- (5) How the health outcome was measured
- (6) Whether the magnitude of the effect was specified
- (7) Whether a count (within the study population) was included to contextualize the effect estimate
- (8) Whether a count (outside the study population) was included to contextualize the public health impact
- (9) Whether limitations to claiming "cause-and-effect" relationship were acknowledged
- (10) Whether results were compared to those from other literature.

All 10 methodological details are reported in the [8] publication, either overtly in the publication text, or by way of in-text references. Whether or not an element was included in the media article was recorded for each of the ten elements and to what extent they were included. Definitions of the categories used to classify articles are listed in Appendix A. Whether the information was reported in the title or main text was also recorded. Data were extracted by JR. LS reviewed all extracted data. Discrepancies were adjudicated by joint discussion between JR and LS. Summary statistics were tabulated using Microsoft Excel.

Results

None of the media articles provided complete descriptions of all 10 methodological details. Fig. 1 shows the proportions of articles that include methodological details in either (or both) the title and the main text. More than 80 % of article titles included the health outcome definition, while fewer than 10 % included the definition of exposure. Fewer than 5 % of titles defined the comparison group. None of the other methodological details was included in the titles. Over 90 % of articles included information in the main text to describe definitions of the exposure and the health outcome, as well as the magnitude of the effect. A definition for the comparison group and a reference to limitations to causal inference were included in the main text of over 80 % of articles. The main text of around 70 % of articles described measurement of exposure and a measure of frequency within the study population. A comparison of results with existing literature was included in the main text of approximately 50 % of articles. Fewer than 10 % of articles described measurement of the health outcome or reported a measure of frequency outside the study population.

A summary of the counts and percentages of items observed in the articles based on their locations is reported in Table 1. Each of the methodological details observed anywhere in the articles (i.e., title or main text) always appeared in the main text; there were no details which appeared in the title, but not in the main text. Part of the exposure definition was included in 84 % of the article titles. A partial definition for the health outcome was reported in 11 % of the titles. Almost 2 % of article titles included a partial magnitude of effect. Fewer than 10 % of the articles included partial descriptions of any of the methodological details in the main text.

Discussion

In this study, 63 media articles reporting the same epidemiology study were compared to understand whether the methodological details selected for representation were consistent. The articles were compared based on the information included in different sections of the articles, specifically the title and main text. There were differences observed among the media articles; the most heterogeneity was observed in whether articles compared results with other literature, as 57 % of articles included this information. There was some variation in inclusion of a measure of frequency within the study population, and in details describing measurement of exposure. However, most of the articles were consistent in either including or excluding the other methodological details in the main text.

The health outcome definition was the only detail which was consistently represented in the article titles (84 %). All other details were consistently unrepresented in the majority of article titles; fewer than 10 % reported the exposure definition, and fewer than 5 % reported that of the comparison group. None of the other methodological details were reported in the titles. This has potential implications for readers who tend to read only the title of an article, and suggests that the information conveyed to this style of reader might typically relate to the health outcome only.

The fact that there is some observed variation among articles seems logical, given that they are a mechanism by which complex topics are summarized and repackaged for the general population. In order to keep readers engaged and to optimize use of limited space journalists must be deliberate in selecting and summarizing the pieces of information they choose to communicate. It follows that one journalist might prioritize such details differently from another due to divergence in what might

¹ In epidemiology, the term "exposure" refers to anything that might be associated with a particular health outcome [3].



Fig. 1. Proportions of articles (n = 63) that include complete (i.e., not partial) methodological details in either the title or the main text.

interest a target audience.

The results from this exercise are consistent with observations from other studies in that some information was not included in all articles. For example, most articles did not provide background incidence rate for the health outcome of interest, when including this might help the reader to understand the magnitude of a reported percent increase in its incidence. This is similar to observations made by Woloshin and Schwartz, who evaluated newspaper articles and television and radio transcripts that reported findings from presentations at scientific meetings. One of their key conclusions was that most sources reported results as relative estimates of change without providing a background rate for context and did not explain limitations of certain study designs. They also observed that few media pieces mention that additional research would be necessary to be more certain about the results [7].

The observed differences among the articles with regards to the elements that were selected for inclusion is consistent with observations by Hammes et al., who described the generally inconsistent reporting of epidemiological data on COVID-19 [2]. In the Hammes et al. report researchers assessed whether newspaper articles reporting COVID-19 data appropriately adjusted for population size and for the pandemic start date, whether logarithmic curves were presented and explained, and whether accuracy of case and death counts were discussed. It was found that only 25 % of newspaper articles adjusted for population size when reporting case or death counts. It was also reported that few sources provided a logarithmic curve, and that less than 30 % of those that did fail to explain its meaning [2], information which, had it been provided, could have helped readers to understand changes in the rate of disease spread over time as opposed to just the increase in the number of cases. This type of omission is suggested to have potential implications for public perception of risk and subsequent decision making [4].

Almost half of the media articles compared the results reported by Zhang et al. to those from other studies. Mentioning comparable results from other literature increases the credibility of the article making it easier for readers to trust the information being reported. Similarly, it is important for media articles to accurately represent the level of certainty with which one can infer a cause-and-effect relationship between exposure and outcome based on the reported results; around 80 % of the articles in this study did include some reference to this. Whitehouse explained that media articles on preliminary epidemiological studies are often insufficient to draw overarching conclusions on the particular research question without replication and validation [6]. Whitehouse urged researchers to consider the utility for the general public when considering when and how to share results from a given study, and stressed the importance of mentioning the limitations of these scientific studies and their reproducibility for the benefit of the readers.

The results of the present comparison have implications for readers, researchers, and journalists. Readers must know that media summaries of peer reviewed studies are just that – summaries. It is likely that some information from the original source is not represented by the article, and that additional information might be necessary to form a fully informed opinion on a given topic. Assessing information from multiple sources is one approach to achieve this. In some contexts researchers can be clear and explicit about which results from a study are most important for a lay reader to know. Understanding that not all relevant details will likely be represented in media articles might help to inform the way in which a study's results are communicated by the authors when

Table 1

Proportions of articles ($n = 63$) th	at included methodological	details by location

Methodological detail & Location	Included	Partially included	Not included
How the exposure was defined	n (%)	n (%)	n (%)
Anywhere (i.e., Title or Main text)	61	2 (3.2)	0 (0.0)
	(96.8)		
Title	5 (7.9)	53 (84.1)	5 (7.9)
Main text	61	2 (3.2)	0 (0.0)
	(96.8)		
How the exposure was measured			10 (00 0)
Anywhere (i.e., Title or Main text)	44 (60.8)	N/A	19 (30.2)
Title	(0.9.8)	N/A	63
1110	0 (010)		(100.0)
Main text	44	N/A	19 (30.2)
	(69.8)		
How the comparison group was			
defined			
Anywhere (i.e., Title or Main text)	53	N/A	10 (15.9)
Title	(84.1)	NI / A	62 (09 4)
Main text	1 (1.0) 53	N/A N/A	10 (15.9)
Main text	(84.1)	14/11	10 (13.5)
How the health outcome was defined	(0 111)		
Anywhere (i.e., Title, or Main text)	62	1 (1.6)	0 (0.0)
	(98.4)		
Title	53	7 (11.1)	3 (4.8)
	(84.1)		
Main text	62	1 (1.6)	0 (0.0)
How the boolth sutcome was measured	(98.4)		
Anywhere (i.e., Title or Main text)	5 (7 9)	N/A	58 (02 1)
Title	0(0.0)	N/A	63
		,	(100.0)
Main text	5 (7.9)	N/A	58 (92.1)
Specifies the magnitude of the effect			
Anywhere (i.e., Title or Main text)	60	2 (3.2)	1 (1.6)
	(95.2)		
Title Main text	0 (0.0)	1 (1.6)	62 (98.4)
Main text	60 (05-2)	2 (3.2)	1 (1.6)
Whether a count (within the study	(93.2)		
contextualize the effect estimate			
Anywhere (i.e., Title or Main text)	45	2 (3.2)	16 (25.4)
	(71.4)		
Title	0 (0.0)	0 (0.0)	63
			(100.0)
Main text	45	2 (3.2)	16 (25.4)
Whether a count (outside the study	(/1.4)		
population) was included to			
contextualize public health impact			
Anywhere (i.e., Title or Main text)	3 (4.8)	6 (9.5)	54 (85.7)
Title	0 (0.0)	0 (0.0)	63
			(100.0)
Main text	3 (4.8)	6 (9.5)	54 (85.7)
Whether limitations to claiming "cause-and-effect" relationship were			
acknowledged	52	N/A	11 (17 5)
Anywhere (i.e., The or Main text)	52 (82 5)	N/A	11 (17.5)
Title	0 (0.0)	N/A	63
	0 (0.0)		(100.0)
Main text	52	N/A	11 (17.5)
	(82.5)		
Whether results are compared with			
those from other literature	06	NT / A	07 (40.0)
Anywhere (i.e., Title or Main text)	36	N/A	27 (42.9)
Title	0 (0 0)	N/A	63
	5 (0.0)		(100.0)
Main text	36	N/A	27 (42.9)
	(57.1)		

amplifying conclusions beyond the publication, itself. Journalists can focus on representing the elements which are most relevant to interpreting and using study results in practice, keeping in mind that this might differ among papers of differing subject matter.

Some of the articles assessed in the present study comprised elements and/or sections that appear to be duplicates of those in other articles. Presumably any one reader might have a greater likelihood of reading duplicated text if it appears in multiple articles compared to the likelihood of reading text that appears in only one story. This, of course, would be determined in part by the relative readership of each source. Further research is needed to understand the extent to which article elements are duplicated, and to assess the inclusion of the methodological details explored in this study in the duplicated articles compared to that in those which are completely unique. Future research should also assess the consistency among total articles in circulation, including complete duplicates which were not included in the present data set. Additionally, future research should assess other properties of the media articles, including word count, and whether the authors of the primary publication were interviewed by the journalist, in addition to characteristics of the media outlets, including whether the content is available exclusively online, whether it is a general or specialized media outlet, its country of origin, and size of its audience. Additional research might focus on heterogeneity among media articles in their inclusion and representation of additional elements not included here, including whether a biologically plausible explanation is known between an exposure and an outcome. Future research might aim to determine whether heterogeneity in reporting practices differs among articles about other types of health outcomes or exposures. For some articles there might be a difference in the information provided and that which might be necessary to make the results actionable for the reader. The scope of the present study was limited to consistency of reporting; future research should determine the extent to which the results reported in media articles are accurate, and should address whether differences in reporting practices or deviations from the original publication influence the degree to which the reader's understanding matches the results reported in the peer-reviewed study. Future work might include conducting a survey among readers to assess the extent to which the primary publication is understood given varying levels of detail provided among media articles. This is particularly relevant to articles which convey a call to action or intervention to alter the course of a presumed developing health outcome. It is further recommended that the journalism and epidemiology communities collaborate to establish a checklist, template, or guide that is mutually agreeable according to the unique demands and priorities of each discipline.

Funding details

Lilianne Samad was a paid contingent worker at Bayer Crop Science during the time this manuscript was written. Jennifer E. Reed is a paid employee at Bayer Crop Science.

CRediT authorship contribution statement

Lilianne Samad: Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **J.E. Reed:** Writing – review & editing, Supervision, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Lilianne Samad was a paid contingent worker at Bayer Crop Science during the time this manuscript was written. Jennifer E. Reed is a paid employee at Bayer Crop Science.

Acknowledgements

The authors thank Dr. Kevin C. Glenn and Dr. John L. Vicini for peer

review of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2025.100188.

References

- [1] Gordis L. Epidemiology (fourth edition). Saunders; 2009.
- [2] Hammes LS, Rossi AP, Pedrotti LG, Pitrez PM, Mutlaq MP, Rosa RG. Is the press properly presenting the epidemiological data on COVID-19? An analysis of newspapers from 25 countries. J Public Health Policy 2021;42(3):359–72. https:// doi.org/10.1057/s41271-021-00298-7.
- [3] Lee TA, PharmD PhD, Pickard AS. Developing a protocol for observational comparative effectiveness research: a User's guide. (Chapter4 exposure definition

and measurement). Agency for Healthcare Research and Quality; 2013. https://www.ncbi.nlm.nih.gov/books/NBK126191/.

- [4] Romano A, Sotis C, Dominioni G, Guidi S. The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. Health Econ 2020;29(11): 1482–94. https://doi.org/10.1002/hec.4143.
- [5] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61 (4):344–9. https://doi.org/10.1016/j.jclinepi.2007.11.008.
- [6] Whitehouse A. Epidemiology and the media. The Conversation. https://theconversation.com/epidemiology-and-the-media-15972; 2013, July 10.
- [7] Woloshin S, Schwartz LM. Media reporting on research presented at scientific meetings: more caution needed. Med J Aust 2006;184(11):576–80. https://doi.org/ 10.5694/j.1326-5377.2006.tb00384.x.
- [8] Zhang Y, Yang H, Li W, Wang Y. Consumption of coffee and tea and risk of developing stroke, dementia, and poststroke dementia: a cohort study in the UK biobank. PLoS Med 2021, November 16;18(11):e1003830. https://doi.org/ 10.1371/journal.pmed.1003830.
- [9] 2022 Altmetric database in JSON format. Unpublished raw data. 2024.