

Quality of news media reports about the effects and costs of health interventions: Systematic review protocol

Oxman M

Working paper, April 26, 2018

Colophon

Title **Quality of news media reports about the effects and costs of health interventions: Systematic review protocol**

Authors Oxman, Matt ^{1,2,3}

1. University of Oxford, United Kingdom 2. Faculty of Health Sciences, Oslo Metropolitan University, Norway 3. Centre for Informed Health Choices, Norwegian Institute of Public Health, Norway

Corresponding author

Matt Oxman

matt@mattoxman.com

Telephone: +47 98636788

Oslo Metropolitan University, Faculty of Health Sciences

Section for Research & Development

Pilestredet Campus

P.O. Box 4 St. Olavs plass

N-0130 Oslo, Norway

Keywords News media, Systematic review, Unreliable information, Key Concepts

Citation Oxman, M., *Quality of news media reports about the effects and costs of health interventions: Systematic review protocol. IHC Working Paper*. Oslo: Informed Health Choices. ISBN: 978-82-8082-971-9

Article category

- About Informed Health Choices
- Key concepts and glossary
- Learning resources
- Systematic reviews
- Development and evaluation of learning resources
- Claim evaluation tools
- Editorials and commentaries
- Grant applications
- Other

Date April 26, 2018

Abstract

Background: Unreliable information about health care, combined with the inability to assess the reliability of such information, can lead to uninformed decisions and, ultimately, waste and unnecessary suffering. Journalism is a particularly important source of health information. There have been several scientific studies of the quality (reliability) of news media reports about the effects of health interventions. There does not appear to be a systematic review of these studies. The planned review is intended to inform: the production and consumption of health news; as well as further research, including the development of interventions to help people assess health care information in the news, so they can make well-informed decisions.

Objectives: 1) To assess criteria used to measure the quality of news media reports about the effects and costs of health interventions; 2) to assess the quality of such reports; and 3) to explore factors that might explain variation in quality.

Eligibility: We will include scientific studies of print, broadcast and online news reports (the population). At least one explicit criterion must have been used by the researchers to measure quality (the condition). It must be possible and sensible to frame the results for at least one criterion as the proportion of reports that satisfied it (e.g. the proportion of reports in which effect estimates are absolute, not only relative). There will be no limits on language, geography or time period of the reports (the context). To be included, studies must specify: the sampling frame; the selection criteria; and the selection technique.

Methods: Two reviewers will extract data, assess risk of bias, and make judgements about Informed Health Choices (IHC) Key Concepts captured by criteria used in scientific studies. If necessary, a third reviewer will help reach consensus on any judgement. Meta-analyses will be conducted to estimate the prevalence of news reports that satisfy respective criteria. To explore variation, subgroup analyses will be conducted for: medium (broadcast vs. other); time period (decade); financial model (commercial vs. non-commercial); journalist specialisation (health or science vs. other); and country income level (low-income vs. other).

Background

The problem

Unreliable health information

This protocol is for a systematic review that is ultimately intended to help address a major, global problem. The problem has three parts. The first is unreliable health information.

In effect, there is endless information about the effects of health interventions, a health intervention being any action intended to improve or maintain the health of individuals or groups. This includes information about: “modern”, “academic”, “conventional” or “western” medicine; “complementary”, “alternative”, “traditional” or “natural” medicine; screening; surgery and devices; diet, exercise and lifestyle; and systems and policies. Some of this information, if not most of it, is unreliable in one way or another. The information can be directly misleading, such as the explicit claim that an intervention causes an outcome when it is only associated with the outcome. Or it can be misleading by omission, such as relative effect estimates without absolute estimates, particularly when the baseline risk is small (Woloshin, et al., 2008).

Researchers have found unreliable lay information about the effects of health interventions in:

- Patient materials (Coulter, et al., 1999); product labels (United States Government Accountability Office, 2010),
- Various types of websites (Culver, et al., 1997; Wolfe, et al., 2002; Glenton, et al., 2005; Spencer, et al., 2016)
- Advertisements (Sansgiry, et al., 1999; Frosch, et al., 2007; Frosch, et al., 2011; Faerber & Kreling, 2013; Groven & Braitwaite, 2016)
- The news (Appendix 1)

Information in the news media is of particular interest, for reasons described later in the background section.

Finally, research referenced in the following subsection shows people are unable to assess information about the effects of health interventions. Logically, if people are unable to assess such information, they are more likely to spread that which is unreliable. Conversely, people spreading unreliable information, as documented in the studies referenced in this subsection, suggests they are unable to assess such information.

Inability to assess health information

The second part of the problem is that people are unable to assess the reliability of information about the effects of health interventions. As part of the Informed Health Choices (IHC) project (www.informedhealthchoices.org), the ability to assess such information and make informed decisions about health interventions has been broken down into the ability to apply specific concepts called the IHC Key Concepts. For example, people need to be able to apply the concept that association is not the same as causation. The original list included 32 concepts (Austvoll-Dahlgren, et al., 2015). The newest iteration includes 36 (Box 1) (Chalmers, et al., 2018). The ability to apply IHC Key Concepts is part of health literacy and has been measured both directly and indirectly.

There have been large surveys of health literacy, in Europe by the European Health Literacy Project Consortium (HLS-EU Consortium, 2013) and in the United States by the National Center for Education Statistics (Kutner, et al., 2006). According to the European survey, “Health literacy is linked to literacy and entails people’s knowledge, motivation and competences to access, understand, appraise, and apply health information in order to make judgments and take decisions in everyday life concerning healthcare, disease prevention and health promotion to maintain or improve quality of life during the life course” (HLS-EU Consortium, 2013, p. 7). For the American survey, the term was defined as: “The degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” (Kutner, et al., 2006, p. iii).

The ability to apply IHC Key Concepts (i.e. assess the reliability of information about the effects of health interventions and make well-informed decisions) is logically included in both definitions. However, neither survey appears to have objectively measured this. The European survey measured the subjective ease of 47 tasks. Tasks related to applying IHC Key Concepts, including “judge if the information about illness in the media is reliable”, were ranked amongst the most difficult (HLS-EU Consortium, 2013).

Box 1: Short titles for the Informed Health Choices (IHC) Key Concepts

Recognising an unreliable basis for a claim

- 1.1. Treatments can harm.
- 1.2. Anecdotes are unreliable evidence.
- 1.3. Association is not the same as causation.
- 1.4. Common practice is not always evidence-based.
- 1.5. Newer is not necessarily better.
- 1.6. Expert opinion is not always right.
- 1.7. Beware of conflicting interests.
- 1.8. More is not necessarily better.
- 1.9. Earlier is not necessarily better.
- 1.10. Hope may lead to unrealistic expectations.
- 1.11. Explanations about how treatments work can be wrong.
- 1.12. Dramatic treatment effects are rare.

Understanding whether comparisons are fair and reliable

- 2.1. Comparisons are needed to identify treatment effects.
- 2.2. Comparison groups should be similar.
- 2.3. Peoples' outcomes should be analysed in their original groups.
- 2.4. Comparison groups should be treated equally.
- 2.5. People should not know which treatment they get.
- 2.6. Peoples' outcomes should be assessed similarly.
- 2.7. All should be followed up.
- 2.8. Consider all the relevant fair comparisons.
- 2.9. Reviews of fair comparisons should be systematic.
- 2.10. Peer review and publication does not guarantee reliable information.
- 2.11. All fair comparisons and outcomes should be reported.
- 2.12. Subgroup analyses may be misleading.
- 2.13. Relative measures of effects can be misleading.
- 2.14. Average measures of effects can be misleading.
- 2.15. Fair comparisons with few people or outcome events can be misleading.
- 2.16. Confidence intervals should be reported.
- 2.17. Do not confuse 'statistical significance' with 'importance'.
- 2.18. Do not confuse 'no evidence of a difference' with 'evidence of no difference'.

Making informed choices

- 3.1. Do the outcomes measured matter to you?
- 3.2. Are you very different from the people studied?
- 3.3. Are the treatments practical in your setting?
- 3.4. Do treatment comparisons reflect your circumstances?
- 3.5. How certain is the evidence?
- 3.6. Do the advantages outweigh the disadvantages?

The American survey suggested a minority of American adults (12%) had “proficient” health literacy (Kutner, et al., 2006, p. v). However, only a sample of the health literacy questions used in that survey are publicly available and none of the available questions measure the ability to assess information about causality (www.nces.ed.gov/NAAL/sample.asp).

Another surrogate measure of people’s ability to apply IHC Key Concepts is their expectations about effects. In a systematic review of patients’ expectations, the majority of participants overestimated benefits and underestimated harms (Hoffmann & Del Mar, 2015). A systematic review of clinicians’ expectations showed clinicians do the same (Hoffmann & Del Mar, 2017).

Other substitutes for ability are attitudes and beliefs. According to a study in the United Kingdom, patients and general practitioners appear to place a disproportionate amount of trust in the reputation of the organisation conducting a clinical trial, the qualifications of the researchers, and peer-review, versus the methods, which are what in fact determine the reliability of the results (The Academy of Medical Sciences, 2016). Moreover, about a third of the respondents (37%) placed a high level of trust in data from medical trials, while about two thirds (65%) placed a high level of trust in the experiences of friends and family, i.e. anecdotal evidence.

There have indeed been studies directly measuring the ability to apply Key Concepts. In Norway, where I am based, Oxman et al. tested a random sample of 626 Norwegian adults (Oxman, et al., 2017). About one in five (19%) showed they were able to distinguish between an association and causation. Other surveys suggest Norwegian students in post-secondary school (Pettersen, 2007) and secondary school (Pettersen, 2005) also struggle with assessing claims about the effects of treatments.

Another product of the Informed Health Choices project is the CLAIM Evaluation Tools database (www.informedhealthchoices.org/claim-evaluation-tools) (Chalmers, et al., 2018). The CLAIM Evaluation Tools are multiple-choice questions developed specifically to measure people’s ability to apply the IHC Key Concepts (Austvoll-Dahlgren, et al., 2016; Austvoll-Dahlgren, et al., 2017). Before the development of the database, in a systematic review of tools for measuring said ability, Austvoll-Dahlgren et al. identified 215 discrete instruments or procedures, capturing up to 15 of the concepts, whereof four captured 10 or more (Austvoll-Dahlgren, et al., 2016).

A validated set of 26 questions measuring the ability to apply 13 of the concepts (i.e. two questions per concept) were used to measure the effects of IHC learning resources, in two randomised trials (Nsangi, et al., 2017; Semakula, et al.,

2017). In the first trial, the control group was 4430 Ugandan primary school children, after loss to follow-up (Nsangi, et al., 2017). In the second trial, after loss to follow-up, the control group was 273 parents and guardians of children participating in the intervention or control group of the first trial (Semakula, et al., 2017). Less than half of the participants in the control group for either trial answered both questions correctly for any of the 13 included concepts.

Uninformed health decisions

The combination of unreliable information about the effects of health interventions and people's inability to assess such information can result in uninformed decisions. This is the third part of the problem. People acting on unreliable information, or failing to act on reliable advice, can in turn lead to waste and unnecessary suffering. This is why the problem is important.

In a systematic review, low health literacy was found to be associated with worse health outcomes (Berkman, et al., 2011). Granted, what was measured in the included studies, as with the aforementioned surveys of health literacy, may not include the objective ability to apply IHC Key Concepts.

People certainly are making poor decisions. Systematic reviews have found: worldwide overuse of medical services more likely to do harm than good (Brownlee, et al., 2017); underuse of effective services (Glasziou, et al., 2017); and overtesting and undertesting in primary care (O'Sullivan, et al., 2018). Meanwhile, fortunes are spent on "alternative" medicine and dietary supplements without reliable evidence that they are beneficial or safe (Optum, 2013; Starr, 2015). The inability to apply IHC Key Concepts is logically a factor in many of these decisions.

This review

Why do the review?

While I am aware of scientific studies of the quality of news media reports about the effects of health interventions (Appendix 1), I was unable to find a systematic review of such primary studies in Epistemonikos (www.epistemonikos.org), Trip Pro (www.tripdatabase.com), or Google Scholar (www.scholar.google.com), as laid out in Appendix 2. Neither am I aware of a systematic review of how quality has been measured in such studies. Zeraatkar et al. conducted a literature search before developing their own measurement tool, named the Quality Index for health-related Media Reports (QIMR) (2017). However, they only assessed validated tools, and they did not compare criteria to the IHC Key Concepts.

By assessing the methods used to assess quality in eligible studies, the planned review can inform the development and use of methods in further research. By providing an overview of the quality of news media reports about the effects of health interventions, the review can inform the generation of new research questions, for example the geographic focus of those questions.

In two ways, the review can help address the problem of unreliable information about the effects of health interventions, inability to assess that information, and uninformed decisions about the interventions. First, the results can be used to raise awareness about the problem. Second, by showing how news media reports about the effects of health interventions tend to be unreliable (i.e. the specific criteria that the reports tend not to satisfy), the results can inform the prioritisation of IHC Key Concepts when developing interventions to help journalists produce reliable reports and help consumers assess information in the news.

Why focus on the news media?

The news media play a special role in the dissemination of health information. In the United Kingdom, despite an increase in the use of social media, the majority of adults still find out about science most regularly through news media, according to a representative survey of 1,749 adults (Castell, et al., 2014). Granted, the proportion of people finding out about specifically health science through the news may be different.

Meanwhile, Chew and Eysenbach found that in 5395 tweets about the H1N1 outbreak in 2009, most links were to a news website (23%) (Chew & Eysenbach, 2010). Another 12% linked to news blogs, feeds or niche news, compared to 1.5% that linked to government and public health agencies.

While there are other important sources of lay information about the effects of health interventions, including them in this review would be “mixing apples and oranges”. In other words, the quality of information in those sources is likely to be significantly different. For example, patient materials are likely to include more information as well as more reliable information than news reports, given that the materials are prepared by health professionals who are trying to inform patients, not attract an audience. The methods used to assess information from different sources may also be different. Furthermore, including sources besides the news media could make the review unwieldy.

Rather, this study can be used to inform reviews focusing on other sources. As stated in the *Cochrane Handbook for Systematic Reviews of Interventions*, there

are pros and cons to both broad and narrow review questions, but with the advent of overviews of systematic reviews, “It may increasingly be considered desirable to plan a series of reviews with a relatively narrow scope, alongside an Overview to summarize their findings” (Higgins & Green, 2011).

Objectives

Primary

1. Assess criteria used to measure the quality of news media reports about the effects and costs of health interventions
2. Assess the quality of news media reports about the effects and costs of health interventions

Secondary

3. Explore factors that might explain variation in quality of news media reports about the effects and costs of health interventions

Eligibility

CoCoPop

Munn et al. note that questions about prevalence do not fit the population, intervention, comparator and outcome (PICO) framework, used for systematic reviews about efficacy (Munn, et al., 2015, pp. 148-149). There is no intervention or exposure, nor is there an outcome on which an intervention or exposure can have an effect. Therefore, instead of the PICO framework, I will use the mnemonic CoCoPop (condition, context, and population) to set inclusion criteria, as Munn et al. recommend. I have added criteria for study design, as per the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-P) statement (Moher, et al., 2015).

Population

In *News Around the World: Content, Practitioners and the Public*, Shoemaker and Cohen write: “[The journalist] typically constructs a method for fulfilling the daily job requirements. He or she rarely has an underlying theoretical understanding of what defining something or someone as newsworthy entails” (Shoemaker & Cohen, 2006, p. 7). In other words, even journalists will struggle to tell you what exactly “news” is. This includes me.

In *A history of news*, Stephens uses the definition: “New information about a subject of some public interest that is shared with some portion of the public” (Stephens, 2007, p. 4). In terms of the review, applying such a broad definition would be problematic in several ways. For example, advertisements, press releases and journal articles could all be considered news.

Instead of using a specific definition, I will be pragmatic. I will only consider studies of information that: 1) is in newspapers or magazines (print), radio, podcasts or television (broadcast) or news websites (online); and 2) is labelled by the authors as “news” or any common synonym. In terms of online, I will focus on dedicated news websites, i.e. not social media platforms, although news reports appear on such platforms.

I will only include studies of reports about the potential health effects (negative or positive changes in health outcomes) and the monetary costs of health interventions (any action intended to improve or maintain the health of an individual or group). The reports can all be about the same condition or intervention, or they can be a mix.

If an otherwise eligible study’s sample includes news reports about the effects of health interventions as well as other types of news reports, I will extract data

only about the former, if possible. Otherwise, I will leave out the study and list it in a table of excluded studies. In the same table, I will list any studies of the quality of other health information in news reports, besides information about the effects of health interventions.

Condition

By “quality”, I mean an attribute that is either conducive to informed decisions about health interventions (e.g. presents potential effects in absolute numbers) or misleading (e.g. only presents potential effects in relative numbers), or the sum of such attributes. The condition of my population—news media reports—is either having satisfied at least one explicit criterion for quality or not (e.g. either presenting potential effects in absolute numbers or not).

In potentially eligible primary studies of which I am aware (Appendix 1), researchers have used different sets of criteria to measure quality. Zeraatkar et al. conducted a literature search before developing their own tool to measure the quality of reporting on health research, named the Quality Index for health-related Media Reports (QIMR) (2017). They identified one other validated tool with the same purpose, named the Index of Scientific Quality (ISQ) (Oxman, et al., 1993). They also identified and assessed instruments for measuring the quality of patient information: Ensuring Quality Information for Patients (EQIP) (Moult, et al., 2004) and DISCERN (Charnock, et al., 1999). Zeraatkar et al. concluded that both EQIP and DISCERN have poor content relevance to health news reports.

It would be problematic to restrict the review to studies that use either the QIMR or ISQ to measure quality. This would exclude studies of reports where the health information is not based on research—e.g. information based on anecdotal evidence. Moreover, the QIMR was published as recently as 2017, and in most of the potentially eligible studies of which I am aware, neither the QIMR or the ISQ were used to measure quality. In several studies, researchers have used criteria based on those used by Moynihan et al. (Moynihan, et al., 2000)—e.g. Smith et al. (2005). In other studies, researchers have developed and used their own tools, e.g. Marcon et al. (2017). A criterion may be sensible even if the tool was not validated, and it may be sensible to group the criterion with others for meta-analysis.

Context

The population for the review is limited to news in the aforementioned mediums: newspapers and magazines; radio, podcasts and television; and news websites. I will not set any limits in terms of geography, language or time period.

Study design

For a study to be eligible, the results must include dichotomous or categorical data for at least one explicit criterion used to measure quality. It must be possible to dichotomise the data as either satisfying the criterion or not, for two reasons: to ensure the results can be easily interpreted, and to synthesise findings across studies.

Some nuance, such as the number of reports that “partially” satisfy a criterion, may be lost in this synthesis. The importance of such detail is questionable; if a quality is important enough to be measured, the criterion used to measure that quality should be clearly satisfied, assuming it is relevant to the specific report.

Studies that only provide a global quality score, such as a star rating, will be excluded and listed in the table of excluded studies. Global scores are difficult to interpret, since they do not tell you which parts of a news report are problematic. Furthermore, they are difficult to compare across studies using different scoring systems.

To be included, studies must specify: the sampling frame (i.e. where the news reports were sampled from); the selection criteria for the reports; and the selection technique. This excludes case studies of single reports. I will list any case studies that would otherwise be eligible in the table of excluded studies.

Methods

Search strategy

To make sure my search is systematic and transparent, I will base my approach on the guideline developed by Kable et al. (2012). As noted in the description of my search for an existing review (Appendix 2), while my question is interdisciplinary, I consider it unlikely that a systematic review designed to answer it would have been published in an academic journal focused on journalism. However, I consider it more likely that an eligible primary study has been published in such a journal. Therefore, in addition to PubMed (www.ncbi.nlm.nih.gov/pubmed), I will search Google Scholar (www.scholar.google.com) and Scopus (www.scopus.com). I will search for grey literature in Open Grey (www.opengrey.org) and Grey Literature Report (www.greylit.org), as well as theses in ProQuest Dissertations & Theses (Global Full text plus UK and Ireland abstracts) (www.proquest.com). I will not place limits on language or time period in any of my searches.

In the search for an existing review, I only included terms for the population (news media reports) when searching Epistemonikos and Trip Pro, since I could filter for systematic reviews, which reduced the results to a manageable number. In the upcoming search for primary studies, given there is not a reliable filter I can use for study design, I will also include terms for the condition (quality) or terms related to the analysis, to improve the specificity of the search strategy.

I will identify additional terms for the PubMed search, including Medical Subject Headings (MeSH) terms, by looking up the studies in PubMed or, if the study is not in PubMed, in another database, and extracting relevant terms. The aim is to ensure my search captures all studies of which I am aware, as well as similar studies. I will also add terms for specific categories of media (e.g. print). I will conduct test searches to inform pragmatic decisions about editing or dropping terms, with the aim of making the search and screening as efficient as possible, while minimising the risk of missing eligible studies. I will use the PubMed search as the basis for searches in other databases.

Besides searching databases, I will check the reference lists of eligible studies. In Scopus, I will screen studies citing those that are eligible, as well as studies citing the development of the ISQ or the QIMR. I will take advantage of my professional network, conducting targeted dissemination of the protocol via email and broad dissemination via Twitter, to crowdsource eligible studies. Finally, I will contact the authors of eligible studies and the developers of the ISQ and QIMR, to ask about unpublished studies.

Study language is not an eligibility criterion, but to be pragmatic, I will only search English language databases. To screen any studies without an English abstract for eligibility, I will use Google Translate (www.translate.google.com). If necessary, I will seek help from someone fluent in the language e.g. an acquaintance or one of the authors. To extract data from a non-English paper, I will again seek help. In the excluded studies table, I will list any apparently eligible studies from which I am unable to extract data because of a language barrier or because of inadequate reporting and inability to obtain missing information from the authors.

A second reviewer and I will independently judge eligibility by reviewing each abstract and, if necessary, (sections of) the full text. Appendix 3 is the form we will use to confirm eligibility. Any disagreements will be discussed before moving on to data extraction and assessment of risk of bias. A third reviewer will be brought in to arbitrate if necessary.

Data extraction

A second reviewer and I will independently extract data about the population, study design, criteria used to assess the quality of the reports, and the results. We will again discuss any disagreements, bringing in a third reviewer to arbitrate if necessary. Appendix 4 is the form we will use to extract data about population and design. Appendix 5 is the separate spreadsheet in which we will enter the individual criteria and results for each criterion.

I have determined what data to extract based on what is reported in the potentially eligible studies of which I am aware (Appendix 1), various tools (Vandenbroucke, et al., 2007; The Joanna Briggs Institute, 2014; Moher, et al., 2015; Munn, et al., 2015), my expertise, and common sense. Appendix 6 contains dummy tables for all the data to be extracted, as well as the risk of bias assessments and analyses.

Population

A second reviewer and I will extract data about the following variables, in terms of the samples of news reports: medium of publication; geographic area of publication; time period of publication; category of intervention reported on; and whether the outlet is commercial or not. In terms of medium, radio and podcasts will be one category, in addition to the categories: newspapers; magazines; television; and news websites. The nine predetermined categories of intervention are: “modern” medicine (aka. “academic”, “conventional” or “Western” medicine); “alternative” medicine (aka. “complementary”, “traditional” or “natural” medicine); screening; surgery; devices; diet; exercise; lifestyle; and systems and policies.

Study design

For each study, we will record: the instrument used to measure quality; the stated objectives; the sampling frame; the selection criteria and technique; and any reported subgroup analyses.

Criteria, response options and results

Finally, we will record each eligible criterion used to assess the quality of reports, as well as the response options, in a spreadsheet separate from the form used to collect data about the population and study design (Appendix 5). We will enter the results for each criterion into the same spreadsheet. Where data are available for subgroups of news reports, we will record the overall results and the results for each subgroup.

Assessing risk of bias

In a systematic review of tools used to assess the quality of prevalence studies and other observational research, Shamliyan et al. identified 46 scales and 51 checklist, none of which discriminated between reporting versus methodological quality, or external versus internal validity (2010). Through citation searches, I identified several tools developed after the review, of which I have chosen to use the latest, developed by Munn et al. (2014), as a starting point, based on face validity and simplicity. While Munn et al. do not refer to the review by Shamliyan and colleagues, they compare their tool to one developed by Hoy et al. (2012) who do consider said review. The critical appraisal tool developed by Munn et al. includes ten items, with the response options “Yes”, “No”, “Unclear” and “Not applicable”:

1. Was the sample representative of the target population?
2. Were study participants recruited in an appropriate way?
3. Was the sample size adequate?
4. Were the study subjects and the setting described in detail?
5. Was the data analysis conducted with sufficient coverage of the identified sample?
6. Were objective, standard criteria used for the measurement of the condition?
7. Was the condition measured reliably?
8. Was there appropriate statistical analysis?
9. Are all important confounding factors/subgroups/differences identified and accounted for?
10. Were subpopulations identified using objective criteria?

Of the ten items, three assess risk of bias and are relevant and applicable to this review: items 2, 6 and 7. Item 2 assesses risk of selection bias, while items 6 and 7 measure detection bias (aka. Information bias). Focusing on these two types of bias is consistent with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (Vandenbroucke, et al., 2007, p. 814) and methods used by the Cochrane Methodology Review Group to systematically review the prevalence of a condition in a population of information, as opposed to people (Welch, et al., 2010, p. 7).

A second reviewer and I will independently assess risk of bias in each study, criterion by criterion, given that some criteria used to measure the quality of news reports require more subjective judgement than others. These assessments will be used to describe the strengths and weaknesses of the studies. If deemed necessary, they will also be used to conduct a sensitivity analysis. It is unclear how

a greater risk of bias would impact the results of studies to be included in this review, so I have no hypothesis about the results of such an analysis.

In terms of selection bias, if the news reports were randomly or sequentially sampled from the target population, we will consider the risk of bias as low. If the reports were purposively sampled, we will consider the risk of bias as high. If any other sampling technique was used, we will consider the risk of bias as moderate or high.

In terms of detection bias, if a criterion requires minimal judgement (e.g. a criterion that risks be presented in absolute numbers), we will consider the risk of bias as low. If the criterion requires substantial judgement (e.g. whether there is sufficient information about potential harms), two researchers assessed the reports, and they were blinded to the journalist and publication, we will consider the risk of bias as moderate. If the criterion requires substantial judgement, and only one researcher assessed the reports or researchers assessing the reports were not blinded to the journalist and publication, we will consider the risk of bias as high.

If the risk of either selection or detection bias is high, we will consider the overall risk of bias as high. If the risk of either is moderate, we will consider the risk as moderate. Appendix 6 includes a dummy table for the risk of bias assessments. Appendix 7 is the form that we will use to assess risk of bias. It includes a space for recording any conflicts of interests. Any such conflicts will be discussed separately when interpreting the results of the review.

Analyses

Criteria used to measure quality

How quality has been measured will be assessed by comparing criteria to the IHC Key Concepts (Box 1). Each eligible criterion from each study will be entered into a spreadsheet, diagonally across from the full concept list (Appendix 8). A second reviewer and I will judge which concepts are captured by each criterion, representing the judgements with marks in the spreadsheet. We will discuss any agreements and bring in a third researcher to arbitrate, if necessary.

The IHC Key Concepts differ from other lists of criteria for making causal inferences in two ways (Chalmers, et al., 2018, pp. 30-31). First, the list is different in that it is both: a) developed using a systematic, transparent and iterative process; and b) intended to help children, the general public and health practitioners make informed decisions, not just health researchers. Second, unlike other tools such as checklists and tip sheets, it is not an intervention in and of itself, limited to a particular population or context, but a comprehensive framework that can be used for mapping skills, developing interventions and evaluating interventions.

Using the IHC Key Concept list to assess the criteria used to measure quality of the news reports has several advantages. First, it provides a framework for grouping criteria from different studies, so results can more easily be synthesised. Second, it will shed light on relevant qualities that have not typically been measured or measured at all. Conversely, it may reveal concepts captured by

the criteria missing from the IHC Key Concept list. Finally, using the list can inform the interpretation of results from this review, including what the results suggest about what concepts are most important when developing interventions for improving reporting on the effects of interventions, or helping people assess information in such reports.

A second reviewer and I will judge what criteria used in different studies can be grouped. Once more, we will discuss any disagreements and bring in a third reviewer to arbitrate if necessary.

Quality of news reports

As far as the data permits, we will meta-analyse the proportions of samples that either satisfied or failed to satisfy respective criteria, i.e. the prevalence of reports satisfying or failing to satisfy a given criterion. See dummy tables in Appendix 6. We will dichotomise categorical data where possible and sensible. Furthermore, we may reframe data that already is dichotomous as satisfying or not satisfying a given criterion.

Not satisfying a criterion will include anything other than completely satisfying it. In other words, if news reports “partially” satisfy a criterion or it is unclear whether the reports satisfy it, we will consider the reports as not having satisfied the criterion. If a criterion has been deemed inapplicable to reports, we will exclude those reports from the meta-analysis for that criterion. In addition to potentially giving us larger samples for meta-analysis, the dichotomisation will make interpreting the results more straightforward.

Across studies, I anticipate reports will generally fail to satisfy most criteria, despite differences in populations and methods. Therefore, I expect mean proportions with 95% confidence intervals will be meaningful. I will prepare forest plots and visually analyse the extent to which there are meaningful differences between point estimates, and I will conduct a Chi-square test to see if the heterogeneity is larger than one would expect by chance. For a given criterion, if the visual interpretation and Chi-square test suggest it is appropriate, I will conduct a meta-analysis using a random effects model. I will also conduct six subgroup analyses, when data are available.

Variation in quality

Again, I do not expect a lot of variation in the quality of reports across studies. To the extent that there is variation, I do not expect a lot of data for exploring it. That said, there are several variables that could logically explain variation, evidence supporting some of those hypotheses, and reasons to test them.

Table 1 is an overview of: the variables for which we will conduct subgroup analyses; the subgroups; the hypothesised differences; and the rationales for the hypotheses and for conducting the analyses. The variables are: the medium in which a report was published; the time period in which a report was published; whether the publishing outlet was commercial; whether the journalist was specialised in health or science; and the income level of the country in which the outlet is situated.

On the one hand, I will be cautious about extrapolating findings if there is no direct evidence for a subpopulation (e.g. broadcast reports) and there is reason to be uncertain about the applicability of the evidence from other subpopulations (e.g. print and online reports). On the other hand, subgroup analyses are frequently misleading, as shown by Sun et al. (2014).

To avoid spurious results, we will use explicit criteria for assessing the credibility of any difference, based on Sun and colleagues' criteria:

- The difference is of practical importance.
- The difference is bigger than we would expect by chance.
- The difference is in the hypothesised direction.
- The difference is unlikely to be explained by other known variables.
- The difference is consistent across studies with different methods.
- The difference is consistent across studies with different outcome measures.
- The difference is consistent across studies with different levels of risk of bias.

For each explanatory variable, given sufficient data, we will conduct univariate regression analyses with the variable (e.g. medium) as the independent variable, and with the proportion of subpopulations (e.g. reports from respective decades) that satisfy the criterion as the dependent variable. We will also conduct multiple regression analyses with all of the explanatory variables. We will note any within-study subgroup analyses where the researchers have used appropriate statistical methods or provided the necessary data for us to do an appropriate statistical analysis ourselves. We will consider those analyses when making judgements about the credibility of subgroup differences.

I have decided against conducting a subgroup analysis for the type of report, i.e. whether the report is a feature or a breaking news report. Features are typically longer in words and prepared on longer deadlines. They might logically be higher quality than breaking news reports. However, an analysis by type of report would likely be confounded by medium, financial model, and specialisation. Furthermore, how researchers have defined feature versus breaking news, and how they have measured length of report or deadline, may be so different across studies that combining the data would not make sense.

Table 1: Subgroup analyses

Variable	Grouping	Hypothesis	Rationale and importance
Medium	Broadcast vs. other (print and online)	Broadcast reports are less likely to satisfy criteria	Broadcast allows for fewer words, as well as less time for writing, given time spent recording and editing. This would logically have a negative impact on the quality of reports. I am aware of empirical evidence that supports this hypothesis (Walsh-Childers, et al., 2016). A credible difference in the hypothesised direction would imply consumers should pay less attention to broadcast reports.
Time period	Decades (≥ 2000 ; 1990-1999; 1980-1989; etc.)	The proportion of reports satisfying criteria is likely to be similar across decades	On the one hand, news reports may have improved since the advent of “evidence-based medicine” (EBM) (Evidence-Based Medicine Working Group, 1992). On the other, there are sustained, systemic barriers to reliable health journalism such as lack of time (Larsson, et al., 2003; Pettersen, 2005), which EBM does not address. Moreover, revenue losses may have led news media to invest less in strategies that might improve the quality of their health news, such as training journalists and editors, and hiring and retaining specialised health and science reporters. However, a credible improvement for one or more criteria would be an impetus for research to explain said improvement, and that research could inform the development of interventions to improve quality overall.
Financial model	Commercial vs. non-commercial	Reports published by commercial outlets are less likely to satisfy criteria	The financial interests of commercial outlets may lead to sensationalism, to attract audiences (Larsson, et al., 2003; Pettersen, 2005; Wallington, et al., 2010). Moreover, companies that sell health interventions may be advertisers, disincentivising critical reporting. A credible difference in the hypothesised direction would imply consumers should support non-commercial outlets and pay less attention to reports published by commercial outlets.
Specialisation	Health and science journalists vs. other	Reports prepared by health or science journalists are more likely to satisfy criteria	Journalist who have specialised in health or science are more likely to have received training in critical appraisal. Moreover, compared to other journalists, they may have been in more contact with scientists, and may be more familiar with the language, content and structure of research papers. There is evidence that journalists with more advanced degrees are more likely to use scientific journal articles as sources, and that journalists with fewer years of experience are more likely to say providing entertainment is an important priority for health news (Wallington, et al., 2010). A credible difference in the hypothesised direction would imply consumers should pay more attention to reports prepared by specialised health and science journalists.
Country income level	Low-income countries vs. other (middle and high-income countries)	The proportion of reports satisfying criteria is likely to be similar across countries with different income levels	While news media in middle and high-income countries may have more resources, journalists in those countries still face systemic barriers to high-quality reporting (Larsson, et al., 2003; Pettersen, 2005; Wallington, et al., 2010). Support for this hypothesis would suggest researchers should develop interventions that can improve the quality of reporting in low-income countries and middle or high-income countries alike. It would also imply those interventions address barriers beside lack of resources.
Newspaper type	Broadsheet vs. tabloid	Reports in broadsheet newspapers are more likely to satisfy criteria	Reporting in tabloid newspapers may be less detailed and more sensational than reporting in broadsheet newspapers, and I am aware of evidence that the quality of reports about the effects of health interventions is lower in tabloids (Entwistle & Hancock-Beaulieu, 1992; Robinson, et al., 2013). A credible difference in the hypothesised direction would imply consumers should pay more attention to broadsheet newspapers, and less to tabloids.

Pilot

Having conducted the search, we will pilot forms, spreadsheets and tables (Appendixes 3 through 6) on the first five eligible studies. We will discuss any issues and make any revisions deemed necessary before completing data extraction, risk of bias assessments and analyses for remaining studies.

Rating the quality of the evidence

I have adjusted the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to rating the quality of evidence (Guyatt, et al., 2008). The GRADE approach involves considering five factors for lowering the certainty of evidence: study limitations, inconsistency of results, indirectness of evidence, imprecision, and publication bias. I have adjusted these, given that GRADE was developed for research about effects, not prevalence. Appendix 9 is the form we will use to rate the quality of the evidence.

The first adjustment is that we will not consider publication bias. I am not aware of any research documenting publication bias for prevalence or incidence studies generally, nor specifically in this area. Besides, I expect too few studies and samples that are too small for it to be meaningful to assess the risk of publication bias.

Second, we will not consider directness since it is inapplicable. I will, however, be cautious about extrapolating evidence across subpopulations specified in Table 1 (e.g. across reports from different time periods), if there is no evidence for one of the subpopulations specific (e.g. reports from before a particular decade).

This leaves considering risk of bias, imprecision and inconsistency. In terms of imprecision, we will use quartiles as a rule of thumb; if the confidence interval substantially crosses two or more quartiles, we will consider there to be important imprecision, following guidance developed by the Cochrane Effective Practice and Organisation of Care (EPOC) group (Cochrane Effective Practice and Organisation of Care, 2017). GRADE factors for increasing the certainty of the evidence are not relevant here.

Ethical considerations and conflicts of interest

The review does not involve particular ethical challenges. I declare that I have no relevant conflicts of interest.

References

- Ashorkhani, M. et al., 2012. Quality of health news disseminated in the print media in developing countries: A case study in Iran. *BMC Public Health*, 12(627).
- Austvoll-Dahlgren, A. et al., 2017. Measuring ability to assess claims about treatment effects: a latent trait analysis of items from the 'Claim Evaluation Tools' database using Rasch modelling. *BMJ Open*, Volume 7, p. e013184.
- Austvoll-Dahlgren, A., Nsangi, A. & Semakula, D., 2016. Interventions and assessment tools addressing key concepts people need to know to appraise claims about treatment effects: a systematic mapping review. *Systematic Reviews*, 5(215).
- Austvoll-Dahlgren, A. et al., 2015. Key concepts that people need to understand to assess claims about treatment effects. *Journal of Evidence-Based Medicine*, Volume 8, pp. 112-125.
- Austvoll-Dahlgren, A. et al., 2016. Measuring ability to assess claims about treatment effects: the development of the 'Claim Evaluation Tools'. *BMJ Open*, Volume 7, p. e013184.
- Berkman, N. D. et al., 2011. Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), pp. 97-107.
- Brownlee, S. et al., 2017. Evidence for overuse of medical services around the world. *Lancet*, 390(10090), pp. 156-168.
- Castell, S. et al., 2014. *Public Attitudes to Science 2014*, London: Ipsos MORI.
- Chalmers, I. et al., 2018. Key Concepts for Informed Health Choices: a framework for helping people learn how to assess treatment claims and make informed choices. *BMJ Evidence-Based Medicine*, 23(1), pp. 29-33.
- Charnock, D., Shepperd, S., Needham, G. & Gann, R., 1999. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology and Community Health*, Volume 53, pp. 105-111.
- Chew, C. & Eysenbach, G., 2010. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLoS ONE*, 5(11), p. e14118.
- Cochrane Effective Practice and Organisation of Care, 2017. *Reporting the effects of an intervention in EPOC reviews*. s.l.:Cochrane Effective Practice and Organisation of Care.
- Coulter, A., Entwistle, V. & Gilbert, D., 1999. Sharing decisions with patients: is the information good enough?. *BMJ*, Volume 318, pp. 318-322.

- Culver, J. D., Gerr, F. & Frumkin, H., 1997. Medical information on the Internet: A study of an electronic bulletin board. *Journal of General Internal Medicine*, 12(8), pp. 466-470.
- Entwistle, V. & Hancock-Beaulieu, M., 1992. Health and medical coverage in the UK national press. *Public Understanding of Science*, Volume 1, pp. 367-382.
- Evidence-Based Medicine Working Group, 1992. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA*, 268(17), pp. 2420-2425.
- Faerber, A. E. & Kreling, D. H., 2013. Content Analysis of False and Misleading Claims in Television Advertising for Prescription and Nonprescription Drugs. *Journal of General Internal Medicine*, 29(1), pp. 110-118.
- Frosch, D. L. et al., 2007. Creating Demand for Prescription Drugs: A Content Analysis of Television Direct-to-Consumer Advertising. *Annals of Family Medicine*, 5(1), pp. 6-13.
- Frosch, D. L., May, S. G., Tietbohl, C. & Pagán, J. A., 2011. Living in the “land of no”? Consumer perceptions of healthy lifestyle portrayals in direct-to-consumer advertisements of prescription drugs. *Social Science & Medicine*, Volume 73, pp. 995-1002.
- Glasziou, P. et al., 2017. Evidence for underuse of effective medical services around the world. *Lancet*, 390(10090), pp. 169-177.
- Glenton, C., Paulsen, E. J. & Oxman, A. D., 2005. Portals to Wonderland: Health portals lead to confusing information about the effects of health care. *BMC Medical Informatics and Decision Making*, 5(7).
- Groven, K. S. & Braitwaite, J., 2016. Happily-ever-after: Personal narratives in weight-loss surgery advertising. *Health Care for Women International*, 37(11), pp. 1221-1238.
- Guyatt, G. H. et al., 2008. GRADE: what is “quality of evidence” and why is it important to clinicians?. *BMJ*, Volume 336, pp. 995-998.
- Higgins, J. & Green, S. eds., 2011. Defining the scope of a review question (broad versus narrow). In: *Cochrane Handbook for Systematic Reviews of Interventions*. s.l.:The Cochrane Collaboration.
- HLS-EU Consortium, 2013. *Comparative report of health literacy in eight EU members states*, s.l.: s.n.
- Hoffmann, T. C. & Del Mar, C., 2015. Patients’ Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review. *JAMA Internal Medicine*, 175(2), pp. 274-286.
- Hoffmann, T. C. & Del Mar, C., 2017. Clinicians’ Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review. *JAMA Internal Medicine*, 177(3), pp. 407-419.
- Hoy, D. et al., 2012. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *Journal of Clinical Epidemiology*, Volume 65, pp. 934-939.

- Kable, A. K., Pich, J. & Maslin-Prothero, S. E., 2012. A structured approach to documenting a search strategy for publication: A 12 step guideline for authors. *Nurse Education Today*, Volume 32, pp. 878-886.
- Kutner, M., Greenberg, E., Jin, Y. & Paulsen, C., 2006. *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy*, Washington, DC: National Center for Education Statistics.
- Larsson, A., Oxman, A. D., Carling, C. & Herrin, J., 2003. Medical messages in the media – barriers and solutions to improving medical journalism. *Health Expectations*, Volume 6, pp. 323-331.
- Marcon, A. R., Murdoch, B. & Caulfield, T., 2017. Fake news portrayals of stem cells and stem cell research. *Regenerative Medicine*, 12(7), pp. 765-775.
- Moher, D. et al., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, 4(1).
- Moult, B., Franck, L. S. & Brady, H., 2004. Ensuring Quality Information for Patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expectations*, Volume 7, pp. 165-175.
- Moynihan, R. et al., 2000. Coverage by the news media of the benefits and risks of medications. *The New England Journal of Medicine*, 342(22), pp. 1645-1650.
- Munn, Z. et al., 2015. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *International Journal of Evidence-Based Healthcare*, Volume 13, pp. 147-153.
- Munn, Z., Moola, S., Riitano, D. & Lisy, K., 2014. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. *International Journal of Health Policy and Management*, 3(3), pp. 123-128.
- Nsangi, A. et al., 2017. Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial. *Lancet*, 390(10092), pp. 374-388.
- Optum, 2013. *Effectiveness of homeopathy for clinical conditions: Evaluation of the evidence*, s.l.: Optum.
- O'Sullivan, J. W. et al., 2018. Overtesting and undertesting in primary care: a systematic review and meta- analysis. *BMJ Open*, Volume 8, p. e018557.
- Oxman, A. D., Austvoll-Dahlgren, A., Garrat, A. & Rosenbaum, S., 2017. *Understanding of key concepts relevant to assessing claims about treatment effects: a survey of Norwegian adults. IHC Working Paper*. Oslo: Informed Health Choices.
- Oxman, A. D. et al., 1993. An Index of Scientific Quality for Health Reports in the Lay Press. *Journal of Clinical Epidemiology*, 46(9), pp. 987-1001.
- Pettersen, S., 2005. Critical Thinking in Norwegian Upper Secondary Biology Education : The Cases of Complementary-Alternative- Medicine and Health Claims in the Media. *Nordina*, 1(2), pp. 61-71.

- Pettersen, S., 2005. Norwegian health journalists' ability to report on health research: A concern to science education?. *Nordina*, 1(1), pp. 5-16.
- Pettersen, S., 2007. *Health Claims and Scientific Knowledge: A study of how students of health sciences, their teachers, and newspaper journalists relate to health claims in society*. Oslo: University of Oslo.
- Robinson, A., Coutinho, A., Bryden, A. & McKee, M., 2013. Analysis of health stories in daily newspapers in the UK. *Public Health*, Volume 127, pp. 39-45.
- Sansgiry, S., Sharp, W. T. & Sansgiry, S. S., 1999. Accuracy of Information on Printed Over-the- Counter Drug Advertisements. *Health Marketing Quarterly*, 17(2), pp. 7-18.
- Semakula, D. et al., 2017. Effects of the Informed Health Choices podcast on the ability of parents of primary school children in Uganda to assess claims about treatment effects: a randomised controlled trial. *Lancet*, 390(10092), pp. 389-398.
- Shamliyan, T., Kane, R. L. & Dickinson, S., 2010. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*, Volume 63, pp. 1061-1070.
- Shoemaker, P. J. & Cohen, A. A., 2006. Part 1: Theory and Method. In: P. J. Shoemaker & A. A. Cohen, eds. *News Around the World*. New York: Routledge, pp. 1-34.
- Smith, D. E., Wilson, A. J. & Henry, D. A., 2005. Monitoring the quality of medical news reporting: early experience with media doctor. *MJA*, Volume 183, pp. 190-193.
- Spencer, E. A., Mahtani, K. R., Goldacre, B. & Heneghan, C., 2016. Claims for fertility interventions: a systematic assessment of statements on UK fertility centre websites. *BMJ Open*, Volume 6, p. e013940.
- Starr, R. R., 2015. Too Little, Too Late: Ineffective Regulation of Dietary Supplements in the United States. *American Journal of Public Health*, 105(3), pp. 478-485.
- Stephens, M., 2007. *A History of News*. 3rd Edition ed. New York: Oxford University Press.
- Sun, X. et al., 2014. How to Use a Subgroup Analysis: Users' Guides to the Medical Literature. *JAMA*, 311(4), pp. 405-411.
- The Academy of Medical Sciences, 2016. *The Academy of Medical Sciences: Medical Information Survey*, London: The Academy of Medical Sciences.
- The Joanna Briggs Institute, 2014. *The systematic review of prevalence and incidence data*. Adelaide: The Joanna Briggs Institute.
- United States Government Accountability Office, 2010. *Herbal Dietary Supplements: Examples of Deceptive or Questionable Marketing Practices and Potentially Dangerous Advice*, s.l.: United States Government Accountability Office.

- Vandenbroucke, J. P. et al., 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Epidemiology*, 18(6), pp. 805-835.
- Wallington, S. F., Blake, K., Taylor-Clark, K. & Viswanath, K., 2010. Antecedents to Agenda Setting and Framing in Health News: An Examination of Priority, Angle, Source, and Resource Usage from a National Survey of U.S. Health Reporters and Editors. *Journal of Health Communication*, 15(1), pp. 76-94.
- Walsh-Childers, K., Braddock, J., Rabaza, C. & Schwitzer, G., 2016. One Step Forward, One Step Back: Changes in News Coverage of Medical Interventions. *Health Communication*, 33(2), pp. 174-187.
- Welch, V. et al., 2010. How effects on health equity are assessed in systematic reviews of interventions (Review). *Cochrane Database of Systematic Reviews*, Issue 12.
- Wolfe, R. M., Sharp, L. K. & Lipsky, M. S., 2002. Content and Design Attributes of Antivaccination Web Sites. *JAMA*, 287(24), pp. 3245-3248.
- Woloshin, S., Schwartz, L. M. & Welch, H. G., 2008. *Know your chances: understanding health statistics*. Berkeley: University of California Press.
- Zeraatkar, D., Obeda, M., Ginsberg, J. S. & Hirsh, J., 2017. The development and validation of an instrument to measure the quality of health research reports in the lay media. *BMC Public Health*, 17(343).