





Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study

Aran Tajika ¹, Yasushi Tsujimoto,^{1,2} Akira Onishi,³ Yusuke Tsutsumi ^{4,5}, Satoshi Funada ¹, Yusuke Ogawa,⁶ Nozomi Takeshima,¹ Yu Hayasaka,¹ Naotsugu Iwakami,¹ Toshi A Furukawa ¹

To cite: Tajika A, Tsujimoto Y, Onishi A, *et al.* Twenty-year follow-up of promising clinical studies reported in highly circulated newspapers: a meta-epidemiological study. *BMJ Health Care Inform* 2023;**30**:e100768. doi:10.1136/bmjhci-2023-100768

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2023-100768>).

Received 23 March 2023
Accepted 29 May 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Aran Tajika;
aran.tajika28@gmail.com

ABSTRACT

Objectives Researchers have identified cases in which newspaper stories have exaggerated the results of medical studies reported in original articles. Moreover, the exaggeration sometimes begins with journal articles. We examined what proportion of the studies quoted in newspaper stories were confirmed.

Methods We identified newspaper stories from 2000 that mentioned the effectiveness of certain treatments or preventions based on original studies from 40 main medical journals. We searched for subsequent studies until June 2022 with the same topic and stronger research design than each original study. The results of the original studies were verified by comparison with those of subsequent studies.

Results We identified 164 original articles from 1298 newspaper stories and randomly selected 100 of them. Four studies were not found to be effective in terms of the primary outcome, and 18 had no subsequent studies. Of the remaining studies, the proportion of confirmed studies was 68.6% (95% CI 58.1% to 77.5%). Among the 59 confirmed studies, 13 of 16 studies were considered to have been replicated in terms of effect size. However, the results of the remaining 43 studies were not comparable.

Discussion In the dichotomous judgement of effectiveness, about two-thirds of the results were nominally confirmed by subsequent studies. However, for most confirmed results, it was impossible to determine whether the effect sizes were stable.

Conclusions Newspaper readers should be aware that some claims made by high-quality newspapers based on high-profile journal articles may be overturned by subsequent studies within the next 20 years.

INTRODUCTION

As people's health awareness has increased, newspapers have covered more stories about health and medicine. These stories feature many diseases, including cancer, stroke, infectious diseases and mental disorders. Some sensationalise the fear and frustration of the disease, while others provide hope for new treatments or preventative measures. These

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ When newspapers cite the results of clinical research articles, they sometimes misrepresent the results based on exaggerated expectations.
- ⇒ Studies with higher levels of evidence may overturn the results of clinical research.

WHAT THIS STUDY ADDS

- ⇒ The results of clinical research articles were relatively stable in papers in which the citation source was properly listed in the newspaper article.
- ⇒ However, the results of approximately one-third of the papers were overturned in the following two decades.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Journalists should be careful in accurately reporting clinical research articles and stating the sources of their citations.
- ⇒ Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may still be overturned by subsequent studies.

stories are often based on articles published in medical journals. The important points of these articles are summarised and presented clearly in newspaper stories for the general public.

However, the media coverage often exaggerates fear and hopes.¹ For example, a phase I uncontrolled study of a new cancer drug published in the *New England Journal of Medicine* showed some effects in one subgroup. Newspapers reported that this treatment produced highly promising results.¹ However, studies cited in newspaper stories are sometimes overturned. Gonon² investigated the 'top 10' most frequently reported studies on attention deficit hyperactivity disorder (by newspapers) and compared these results with

those of subsequent studies. Two studies were confirmed, four attenuated, three refuted and one was neither confirmed nor refuted.

When the strength of the research design is considered, randomised controlled trials (RCTs) and their meta-analyses provide the strongest evidence for treatment decisions. However, newspapers are more likely to report observational studies (OSs) than RCTs.³ Notably, exaggeration often begins with medical journal articles themselves.¹ One problem with studies with weak evidence is that the reproducibility of the results is low. Ioannidis conducted a simulation study and noted that a meta-analysis of good-quality RCTs and adequately powered RCTs assumed a reproducibility of 85%, but only 23% for underpowered RCTs and approximately 20% for adequately powered OSs.⁴ Ioannidis⁵ identified studies cited more than 1000 times in high-impact factor (IF) journals in general and internal medicine. When these studies were compared with subsequent studies that theoretically had better-controlled designs, only half of the RCTs and none of the OSs were replicated. Furthermore, when statistically significant and extremely favourable initial reports of intervention effects were examined, it was found that the majority of such large treatment effects emerged from small studies. When additional trials were performed, the effect sizes typically became much smaller.⁶ When newspapers report and overestimate the results of these initially promising studies, the information that reaches the public may be doubly overstated.

This study investigated the trustworthiness of medical news. We examined whether newspaper reports were confirmed through subsequent studies that examined the same clinical questions. In other words, we examined how much caution general readers need to exercise when reading newspaper reports on medical research.

METHODS

Selection of newspaper stories and original studies

We selected four quality papers (two from the USA and two from the UK) and four non-quality papers (two from the USA and two from the UK) with the highest circulation according to the Audit Bureau of Circulations⁷ and Alliance for Audited Media.⁸ We examined these two newspaper types for several reasons. Generally, quality papers are believed to have higher quality reporting than non-quality papers,⁹ which tend to focus on readers' emotions rather than on the veracity of the reports.¹⁰ However, when we consider the respective circulations of the two types of papers, non-quality papers have as many readers as quality papers; they sometimes have more power to lead public opinion.¹¹

We selected newspaper articles that quoted main medical journals. First, we selected medical journals from the following two fields: 'general and internal medicine' and 'public, environmental and occupational health' according to their journal IF on Journal Citation Reports. In addition, we selected the 20 journals in each

field with the highest IFs for 2000. We ultimately selected 40 medical journals as an ad hoc set of representative medical journals that might meet the public interest. Next, we searched the LexisNexis database,¹² which contains stories from prominent newspapers worldwide. We used the names of 40 medical journals as search words and selected newspaper stories:

- ▶ Printed in 2000 in the four above-mentioned quality and four non-quality newspapers.
- ▶ That quoted articles that were published in the above-mentioned 40 journals.
- ▶ In which we could identify the original medical journal article.
- ▶ That mentioned the effectiveness, recommendation of treatment or prevention at that time.

Pairs of independent investigators (AT, YO, NT, YH and NI) selected eligible newspaper articles for analysis. Disagreements were resolved through discussions between the two investigators and, when necessary, in consultation with a third author (TAF). We found the original articles quoted in these newspapers. When two or more articles were quoted in a newspaper story, we selected all the articles. When the number of eligible studies was greater than 100, 100 studies were randomly selected. Original articles were classified into the following categories:

- ▶ Animal or laboratory study.
- ▶ Clinical study.
 - Case reports or case series.
 - OS.
 - RCT.
 - Systematic review (SR) of OSs with or without meta-analysis.
 - SR of RCTs with or without meta-analysis.
 - Other reviews (eg, narrative reviews).
- ▶ Others (eg, comment, letter).

We excluded studies in which specific clinical questions were not identifiable (eg, health economics studies) because we could not search for corresponding subsequent studies in the next step.

Selection of subsequent studies on the same clinical questions

For each original article, we searched for subsequent studies that examined the same clinical questions using 'stronger' research designs. The evidence levels of all the studies were classified according to the following hierarchy:

1. SR of RCTs.
2. Single RCT.
3. SR of OSs/single OS.
4. Case series/a case study.

The characteristics of 'stronger design' are as follows^{5 13}:

- ▶ The subsequent study used a design with a higher level of evidence hierarchy than the original study.
- ▶ If studies had the same level of evidence hierarchy, a study with a larger sample size constituted stronger evidence.

- ▶ If the design of the original study was an SR of an RCTs, we searched for the latest SR for the RCTs.
- ▶ If the design of the original study was the SR of OSs or other reviews, we searched for the largest RCT or the latest meta-analysis of RCTs. If we could not find these studies, we searched for the latest OS meta-analyses.
- ▶ If the original study was an animal or laboratory study, we searched for the most appropriate clinical study asking the same clinical question according to the evidence hierarchy.

First, two authors (AT, YaT, AO, YuT and SF) independently searched the Web of Science for potential new papers in which the original paper was cited through December 2021. Subsequently, to prevent search omissions, AT conducted a PubMed search through June 2022 to search for anything more valid than the candidates' new articles on the Web of Science. If new candidate papers were found, the authors discussed them in pairs to identify the new papers. The PubMed search was conducted using the most comprehensive terms possible, and the search formula was documented.

Comparisons of original and subsequent studies

We extracted the data when the original study authors presented their primary outcomes. If the authors failed to designate their primary outcome(s), the outcome described first was considered the primary outcome. Next, we extracted the outcomes of the subsequent studies, which were as similar as possible to those of the original studies.

We conducted the following two-step comparison. First, we compared the effectiveness of the original studies with that of newer studies and classified each comparison into one of three categories: 'unchallenged', 'contradicted' or 'confirmed'.^{5 13}

- ▶ Unchallenged: when there was no subsequent study with a higher level of evidence.
- ▶ Contradicted: when a subsequent study denied the effectiveness of the original study.
- ▶ Confirmed: The original and subsequent studies concluded that the intervention was effective, regardless of the effect size difference.

When we could not compare these outcomes, we compared the benefits and applicability of both studies and made qualitative judgements.

Furthermore, among 'confirmed' cases, when the outcomes of both original and subsequent studies were exactly comparable (ie, when a new paper was a meta-analysis, the original paper was included in the funnel plot of the new paper, and accurate effect size comparison was possible), we compared the effect sizes of both studies. Outcomes were extracted as continuous or dichotomous data such as standardised mean difference (SMD), OR, risk ratio (RR) or HR. We gave preference to continuous data. We compared these values when the SMD was shown in the subsequent meta-analysis, and when the SMD of the original paper was shown in that study. When studies showed effectiveness using only dichotomous data, the

OR was calculated first. We then converted OR into SMD using the following formula¹⁴:

$$\text{SMD} = \frac{\sqrt{3}}{\pi} \ln \text{OR}$$

We classified 'confirmed' cases into one of two categories: 'initially stronger effects' or 'replicated'.¹³

- ▶ Initially stronger effects: when the point estimate of the original study was not included in the 95% CI of the SMD of the subsequent study or the SMD of the original study was 0.2 SD units or greater than that of the subsequent study (0.2 SD units would signify a small effect difference according to Cohen's rule of thumb).¹⁵
- ▶ Replicated: when the point estimate of the original study was included in the 95% CI of the SMD of the subsequent study, and the two SMDs were within 0.2 SD units apart, or the effect size of the subsequent study was larger than that of the original study.

When the SMD could not be calculated from the RR or the study showed only the HR, as it could not be converted into SMD, we directly compared only the RRs or HRs. Their 95% CI was presented in the papers without considering the difference of 0.2 SD units of SMDs.

Outcomes

Primary outcome

We defined the primary outcome, 'the proportion of confirmed studies', as follows:

$$\text{Proportion of confirmed studies} = \frac{\text{Confirmed studies}}{\text{Total studies} - \text{Unchallenged studies}} \times 100 (\%)$$

Secondary outcomes

We classified the original studies according to their research design and medical fields and examined the differences between quality and non-quality papers. The proportion of confirmed studies in each subgroup was calculated.

Analyses

Statistical analyses were performed using STATA V.17.0. Statistical differences among subgroup categories were tested using the χ^2 test, and SMD was compared using the Wilcoxon signed-rank test. The level of significance was set at $p < 0.05$ (two tailed).

Patient and public involvement

No patients or public members were involved in conducting this research.

RESULTS

Characteristics of newspaper stories, original studies and subsequent studies

Figure 1 illustrates the details of the search. The eight newspapers selected were the *New York Times* (USA, quality), *Washington Post* (USA, quality), *Daily Telegraph* (UK, quality), *Times* (UK, quality), *USA Today* (USA,

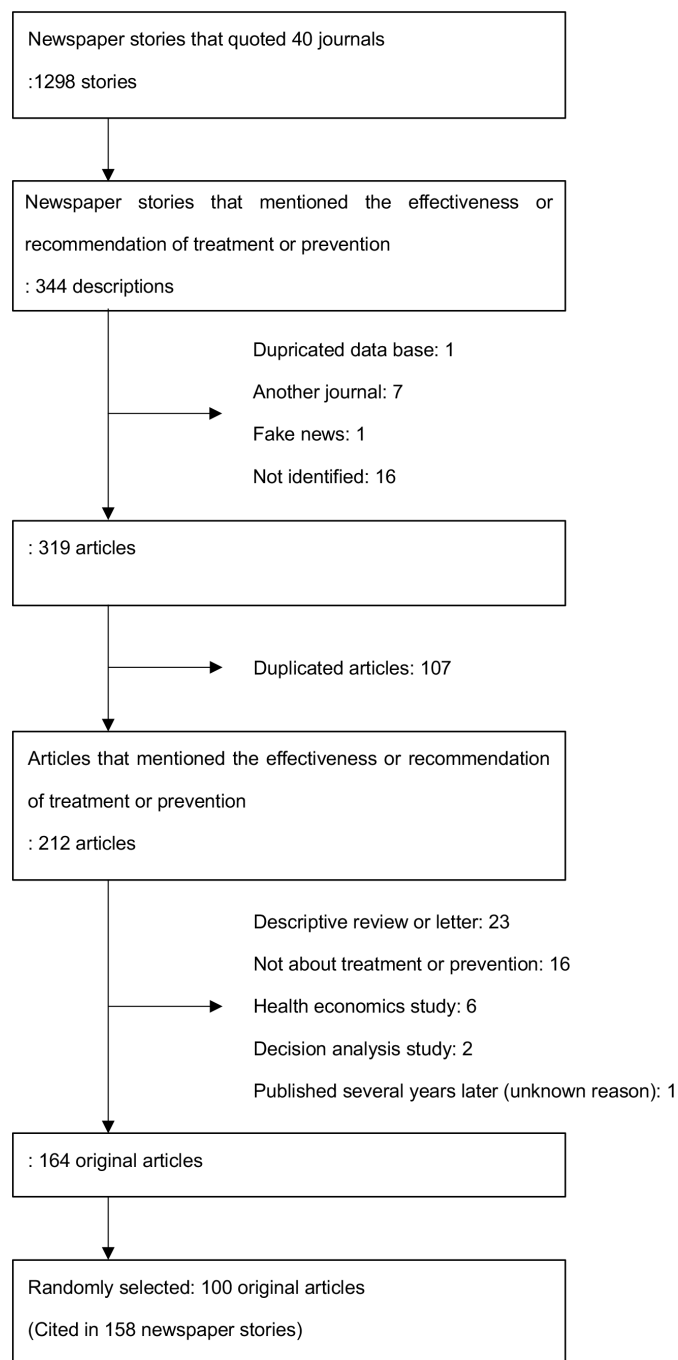


Figure 1 Flow chart of original study identification process.

non-quality), *Daily News* (USA, non-quality), *Daily Mail* (UK, non-quality) and *Daily Mirror* (UK, non-quality). When searching for journal names in newspaper stories, we found 1298 newspaper stories, of which 344 described the effectiveness of or recommended certain treatments or preventive measures ($\kappa=0.73$) (table 1). Online supplemental eTable 1 lists the names of 40 medical journals.

A total of 344 newspaper stories were referred to in 319 scientific journal articles. After excluding duplicates, we identified 212 articles that mentioned the effectiveness of the recommended treatment or prevention. We excluded 48 articles because the research questions

could not be identified. Finally, we identified 164 original articles and randomly selected 100 of them. These were cited in 158 newspaper articles. The journals in which the 100 original articles were published were as follows: *New England Journal of Medicine (NEJM)*, 39; *Journal of the American Medical Association (JAMA)*, 21; *Lancet*, 16; *British Medical Journal (BMJ)*, 9; *Archives of Internal Medicine*, 8; *Annals of Internal Medicine*, 3; *American Journal of Epidemiology*, 1; *American Journal of Public Health*, 1; *Infection Control and Hospital Epidemiology*, 1; and *Mayo Clinic Proceedings*, 1. Approximately three-quarters of these articles were published in three major journals (*NEJM*, *JAMA*, *Lancet*).

Of the 100 articles, 58 were RCTs and 31 OSs. A few other designs corresponded to various ICD-10 categories. Of the 158 newspaper stories, two-thirds were in quality papers and the rest in non-quality.

For four of the 100 original studies, the newspapers stated their effectiveness, but the primary outcome of those studies did not indicate their effectiveness. Therefore, these were excluded from this study. In the remaining 96 studies, 104 effective treatments were identified. Subsequent studies on each treatment were searched. We identified relevant subsequent studies for 86 of these 104 treatments. The 18 others remained unchallenged (table 2). Of the 86 subsequent studies, 83 were SR (SR of RCTs, $n=45$; SR of OSs, $n=23$; SR of RCTs and OSs, $n=15$), followed by RCT ($n=2$) and OS ($n=1$). The PubMed search formulae are listed in online supplemental eTable 2.

Comparisons of original and subsequent studies

Table 2 shows the proportions of the confirmed studies. A total of 69% (59/86) (95% CI 58.1 to 77.5) of the original studies were confirmed in subsequent studies. Furthermore, of the 59 confirmed original studies, 16 were comparable to subsequent studies in terms of effect size. Among these 16, 13 were replicated and three reported effect sizes larger than the corresponding subsequent studies. Of these 16 studies, 11 compared SMDs. The median SMDs of the original and subsequent studies were 0.23 (0.18, 0.45) and 0.25 (0.15, 0.32), respectively ($p=0.34$, Wilcoxon signed-rank test). However, for the remaining 43 studies, strict comparisons of effect sizes were not possible because the outcomes were not fully matched between the original and subsequent studies. Details of the original and subsequent studies are presented in online supplemental eTable 3.

We conducted subgroup analyses on the proportions of confirmed studies for each research design in the original articles (online supplemental eTable 4). The proportions of confirmed OS and RCT studies (of which there was a relatively large number) were 61.3% (19/31) and 70.5% (31/44), respectively. Other designs included fewer studies, and we found no significant differences in the research design ($p=0.74$, χ^2 test). For the ICD-10 categories, the differences according to disease were not significant ($p=0.67$, χ^2 test). The proportion of confirmed studies cited in quality papers (56/88, 63.6%) was lower

Table 1 Characteristics of included newspaper stories

Newspaper	Country	Newspaper type	Newspaper stories that quoted 20 general and internal medicine journals	Newspaper stories that quoted 20 public, environmental and occupational health journals	Total
<i>New York Times</i>	USA	Quality	258	13	271
<i>Washington Post</i>	USA	Quality	279	22	301
<i>Daily Telegraph</i>	UK	Quality	28	5	33
<i>Times</i>	UK	Quality	191	18	209
<i>USA Today</i>	USA	Non-quality	122	11	133
<i>Daily News</i>	USA	Non-quality	65	7	72
<i>Daily Mail</i>	UK	Non-quality	173	9	182
<i>Daily Mirror</i>	UK	Non-quality	91	6	97
Total			1207	91	1298

than that in non-quality papers (31/44, 70.5%); however, the difference was not statistically significant ($p=0.42$, χ^2 test).

Example 1: contradicted

A prospective cohort study published in *BMJ* in 2000, covered by *Daily Mail*, suggested that drinking fluoridated water significantly reduced hip fractures.¹⁶ Neither the subsequent matching study, meta-analysis of 14 observational studies, nor the original study¹⁷ found any significant risk reduction in hip fractures.

Example 2: confirmed

One RCT published in the *JAMA* in 2000 and covered by the *Washington Post* suggested that sertraline was more effective than a placebo in patients with post-traumatic stress disorder (PTSD). The subsequent matching study was a meta-analysis comparing pharmacotherapies for PTSD, published in 2022.¹⁸ In the subgroup analysis, which included the original RCT, sertraline was compared with placebo. The authors concluded that sertraline was effective. Therefore, the effectiveness reported in the original study was confirmed in a subsequent study. Furthermore, the point estimate of the original study's RR described in the subsequent study's forest plot was 0.70, and the point estimate and 95% CI of the RR of the new article was 0.68 (0.56 to 0.81). After calculating the SMD from these values, the original study had an SMD of 0.26, and the new study had a value of 0.27 (95% CI 0.15 to 0.40). We categorised this finding as not only 'confirmed' but also 'replicated'.

Example 3: unchallenged

Examples included in the unchallenged studies are as follows: Most studies have investigated unique interventions (eg, short nails for preventing infection, anti-digoxin fab for cardiac arrhythmia, horse chestnut seed extract for chronic venous insufficiency, beta-sheet breaker peptides for prion-related disorders, the Krukenberg procedure for double-hand amputees and yoga for carpal tunnel syndrome), and several studies have examined the effects of special drug use (eg, ondansetron for bulimia nervosa, growth hormone for Crohn's disease and combination therapy with old antidepressants, nefazodone and psychotherapy for chronic depression). However, these findings are difficult to validate using well-designed studies. The details are shown in online supplemental eTable 3.

DISCUSSION

This is the first study to examine a 20-year course of treatment or prevention recommended by newspaper articles in various medical fields. We selected newspaper stories that recommended certain treatments or preventions in 2000 and compared their results with those in the original research articles and compared the original studies with newer ones with better-controlled designs. Sixty-nine per cent (59/86) of the original studies were confirmed by subsequent studies. Among the confirmed studies, 13 of the 16 studies replicated both the direction and magnitude of the treatment effect. In studies in which the effects were confirmed, the effect sizes were relatively stable. However, the results of the remaining 43 studies were not comparable.

Table 2 Main analyses of the proportion of confirmed studies

	Total	Unchallenged	Contradicted	Confirmed	Proportion of confirmed studies, 95% CI (%)
Original studies	104*	18	27	59	68.6 (58.1 to 77.5)
*104 comparisons from 96 original studies (including duplicates).					



As far as we know, few studies investigated the replicability of articles quoted in daily newspapers.^{2 19} One is about attention deficit hyperactivity disorder studies, and the other is about risk factor studies; the proportions of 'confirmed' studies according to their definitions were 20% and 49%, respectively. The proportion of confirmed cases in our study (68.6%) was higher than those in these studies. The reasons for this may be as follows. Previous studies have not focused on treatment or prevention. Therefore, these proportions could not be compared. Furthermore, the definition of 'confirmed' in these studies was stricter than in our study. However, even in well-known newspapers, one-third of the stories may have been overturned by subsequent studies. Several studies have reported that the reporting standard in quality newspapers is significantly higher than that in non-quality papers.^{9 20 21} In this study, the proportion of confirmed studies in quality newspapers was slightly lower than that in non-quality newspapers; however, this difference was not statistically significant. There may not be much of a difference between highly circulated quality papers and low-quality papers.

This study had some limitations. First, newspaper story authors often do not provide details about their information sources. It is often claimed that the best journalists are those with the most sources'.²² In these cases, we could not find any articles quoted in newspapers. Therefore, for convenience, we used the journal names as search words. Consequently, only better-quality newspaper stories, in which journal names were written, were included. This may have led to the discovery of higher quality stories. Consequently, the proportion of quoted RCT may be higher than that of other standard newspaper stories. The credibility of studies cited in newspaper articles that do not list the sources of citations remains unclear. Second, an increasing number of SRs have been published in recent years, and several similar SRs can often be found on any research topic. Therefore, it is difficult to select the most appropriate option. To find the optimal subsequent study, two independent researchers checked the full paper and selected the best study from among several candidates. This reduced the number of arbitrary choices as much as possible. Third, we assumed that most subsequent study designs would be SR. Therefore, we searched the Web of Science for new studies that cited the original paper, and compared them with the effect sizes shown in the forest plot. However, the authors of subsequent SRs did not always cite the original articles for various reasons (eg, subtle differences in the type of outcome or timing of measurement). If cited, they were excluded from forest plots. Only 11 studies compared SMDs and 43 studies, although found to be effective, were unable to compare effect sizes. It is possible that the original studies reported a very large effect size, while the subsequent studies were only marginally significant. Based on these results, it is impossible to determine whether the SMDs are stable. Future studies should rigorously compare effect sizes by aligning outcomes. Fourth, 18 unchallenged studies

focused on unique topics. Our definition of primary outcome excluded these numbers from the denominator, which makes the proportion of confirmed studies appear higher than it is. If these were included in the denominator, the proportion of confirmed cases would have been much lower.

However, this study has several strengths. This is the first study to examine the veracity of newspaper stories on treatment and prevention in various medical fields. Second, we followed up on each treatment over a 20-year period and took relevant subsequent studies with stronger designs as the gold standard. Although we cannot rule out the possibility that the results of subsequent studies may be reversed in the future, we believe that the results obtained over the past 20 years are generally robust. Third, to find the most appropriate subsequent study, we reviewed and discussed many SRs using the Web of Science and PubMed. We spent a lot of time carefully going through this process to make sure we did not miss any relevant papers.

CONCLUSION

The results for clinical research articles were relatively stable for papers in which the citation source was properly listed in newspaper articles. Journalists should provide information on the source studies to enable researchers to identify them. However, the results of approximately one-third of these studies were overturned over the following two decades. Readers should be aware that more than a few claims made in highly circulated newspapers based on high-profile journal articles may be overturned in subsequent studies.

Author affiliations

¹Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan

²Oku Medical Clinic, Osaka, Japan

³Department of Advanced Medicine for Rheumatic diseases, Graduate School of Medicine, Kyoto University, Kyoto, Japan

⁴Department of Emergency Medicine, National Hospital Organization Mito Medical Center, Ibaraki, Japan

⁵Department of Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan

⁶Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

Contributors AT and TAF conceived the study. AT and TAF designed the study.

AT, YaT, AO, YuT, SF, YO, NT, YH and NI did the literature search and extracted the data. AT did the analyses. AT and TAF wrote the first draft of the manuscript. All authors contributed to the interpretation of the findings and subsequent edits of the manuscript. TAF provided overall supervision to the project. AT is the guarantor and accepts full responsibility for the work and the conduct of the study, had access to the data and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests AT received lecture fees from Sumitomo Dainippon Pharma, Eisai, Janssen Pharmaceutical, Meiji-Seika Pharma, Mitsubishi Tanabe Pharma, Otsuka and Takeda Pharmaceutical. AO received research grants and/or speaker fees from Pfizer, Bristol-Myers Squibb, Advantest, Asahi Kasei Pharma, Chugai Pharmaceutical, Eli Lilly Japan K. K., Ono Pharmaceutical Co., UCB Japan Co., Mitsubishi Tanabe Pharma Co., Eisai, Abbvie, Takeda Pharmaceutical, and Daiichi

Sankyo, SF received a research grant from JSPS KAKENHI (Grant Number JP 20K18964), the KDDI Foundation and the Pfizer Health Research Foundation. TAF reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi, and SONY, and a grant from Shionogi outside the submitted work; TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Aran Tajika <http://orcid.org/0000-0003-3926-8867>

Yusuke Tsutsumi <http://orcid.org/0000-0002-9160-0241>

Satoshi Funada <http://orcid.org/0000-0002-8925-2348>

Toshi A Furukawa <http://orcid.org/0000-0003-2159-3776>

REFERENCES

- 1 Woloshin S, Schwartz LM, Kramer BS. Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 2009;101:1596–9.
- 2 Gonon F, Konsman J-P, Cohen D, *et al*. Why most biomedical findings echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. *PLoS One* 2012;7:e44275.
- 3 Selvaraj S, Borkar DS, Prasad V, *et al*. Media coverage of medical journals: do the best articles make the news *PLoS ONE* 2014;9:e85355.
- 4 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
- 5 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218.
- 6 Pereira TV, Horwitz RJ, Ioannidis JPA. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676.
- 7 Circulations ABO. n.d. Available: <http://www.auditbureau.org>
- 8 Media AfA. n.d. Available: <http://www.auditedmedia.com>
- 9 Wilson A, Bonevski B, Jones A, *et al*. Media reporting of health interventions: signs of improvement, but major problems persist. *PLoS One* 2009;4:e4831.
- 10 Bell L, Seale C. The reporting of cervical cancer in the mass media: a study of UK newspapers. *Eur J Cancer Care (Engl)* 2011;20:389–94.
- 11 Dean M. Tabloid campaign forces UK to reconsider sex-offence laws. *Lancet* 2000;356:745.
- 12 Lexisnexis. n.d. Available: <https://www.lexisnexis.com/en-us/home.page>
- 13 Tajika A, Ogawa Y, Takeshima N, *et al*. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry* 2015;207:357–62.
- 14 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127–31.
- 15 Cohen J. *Statistical power analysis in the behavioral science*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
- 16 Phipps KR, Orwoll ES, Mason JD, *et al*. Community water fluoridation, bone mineral density, and fractures: prospective study of effects in older women. *BMJ* 2000;321:860–4.
- 17 Yin X-H, Huang G-L, Lin D-R, *et al*. Exposure to fluoride in drinking water and hip fracture risk: a meta-analysis of observational studies. *PLoS One* 2015;10:e0126488.
- 18 Williams T, Phillips NJ, Stein DJ, *et al*. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev* 2022;3:CD002795.
- 19 Dumas-Mallet E, Smith A, Boraud T, *et al*. Poor replication validity of biomedical association studies reported by newspapers. *PLoS One* 2017;12:e0172650.
- 20 Cooper BEJ, Lee WE, Goldacre BM, *et al*. The quality of the evidence for dietary advice given in UK national newspapers. *Public Underst Sci* 2012;21:664–73.
- 21 Robinson A, Coutinho A, Bryden A, *et al*. Analysis of health stories in daily newspapers in the UK. *Public Health* 2013;127:39–45.
- 22 Weitkamp E. British newspapers privilege health and medicine topics over other science news. *Public Relations Review* 2003;29:321–33.