

RESEARCH

The association between exaggeration in health related science news and academic press releases: retrospective observational study

 OPEN ACCESS

Petroc Sumner *professor*^{1,2}, Solveiga Vivian-Griffiths *research assistant*^{1,2}, Jacky Boivin *professor*², Andy Williams *lecturer*³, Christos A Venetis *senior lecturer*⁴, Aimée Davies *research assistant*², Jack Ogden *research assistant*², Leanne Whelan *research assistant*², Bethan Hughes *research assistant*², Bethan Dalton *research assistant*², Fred Boy *senior lecturer*⁵, Christopher D Chambers *professor*^{1,2}

¹Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff CF10 3AT, UK; ²School of Psychology, Cardiff University, UK; ³School of Journalism, Media & Cultural Studies, Cardiff University, UK; ⁴School of Women's and Children's Health, University of New South Wales, and Graduate School of Medicine, Faculty of Science, Medicine and Health, University of Wollongong, Australia; ⁵Department of Psychology, Swansea University, UK

Abstract

Objective To identify the source (press releases or news) of distortions, exaggerations, or changes to the main conclusions drawn from research that could potentially influence a reader's health related behaviour.

Design Retrospective quantitative content analysis.

Setting Journal articles, press releases, and related news, with accompanying simulations.

Sample Press releases (n=462) on biomedical and health related science issued by 20 leading UK universities in 2011, alongside their associated peer reviewed research papers and news stories (n=668).

Main outcome measures Advice to readers to change behaviour, causal statements drawn from correlational research, and inference to humans from animal research that went beyond those in the associated peer reviewed papers.

Results 40% (95% confidence interval 33% to 46%) of the press releases contained exaggerated advice, 33% (26% to 40%) contained exaggerated causal claims, and 36% (28% to 46%) contained exaggerated inference to humans from animal research. When press releases contained such exaggeration, 58% (95% confidence interval 48% to 68%), 81% (70% to 93%), and 86% (77% to 95%) of news stories, respectively, contained similar exaggeration, compared with exaggeration rates of 17% (10% to 24%), 18% (9% to 27%), and 10% (0% to 19%) in news when the press releases were not exaggerated. Odds ratios for each category of analysis were 6.5 (95% confidence interval 3.5 to 12), 20 (7.6 to 51),

and 56 (15 to 211). At the same time, there was little evidence that exaggeration in press releases increased the uptake of news.

Conclusions Exaggeration in news is strongly associated with exaggeration in press releases. Improving the accuracy of academic press releases could represent a key opportunity for reducing misleading health related news.

Introduction

The framing of health related information in the national and international media, and the way in which audiences decode it, has complex and potentially powerful impacts on healthcare utilisation and other health related behaviour in many countries.¹⁻⁶ The media also demonstrably influences the behaviour of scientists and doctors.^{3,4} Such impacts may often be beneficial, but misleading messages can have adverse effects (even if these effects may be difficult to predict and prove because the responses of audiences are complex and multiply determined).⁶ This problem is not restricted to rare dramatic cases such as vaccination scares^{7,8}; the cumulative effect of everyday misreporting can confuse and erode public trust in science and medicine, with detrimental consequences.⁹⁻¹¹

"Information subsidies" such as university press releases have long been used to deliver salient aspects of selected research,^{12,13} and as journalists are increasingly expected to produce more copy in less time^{14,15} these press releases have become the dominant link between academia and the media.^{16,17} As such,

Correspondence to: P Sumner or C D Chambers sumnerp@cardiff.ac.uk or chambersc1@cardiff.ac.uk

Extra material supplied by the author (see <http://www.bmj.com/content/349/bmj.g7015?tab=related#datasupp>)

Supplementary information

information included in press releases is highly likely to be included in news stories.¹⁸ Although accurate information, alone, is not sufficient for clear public understanding and informed behaviour,¹⁹ it is nevertheless important that health and science news is not misleading, especially when it includes health advice for readers. News pieces have a different purpose to, and readership from, journal articles and are not expected to reproduce them or express claims in the same way. However, given that news is often explicitly or implicitly blamed for distorting and exaggerating scientific findings,⁹ it is pertinent to determine the sources of such misreporting. In fact there is little evidence on how often news stories go beyond what scientists state in peer reviewed journal articles, and, when they do, whether misrepresentation is already present in the un-peer reviewed sources supplied by scientists and press offices.

Previous research suggests that press releases can be a source of misinformation. Of 200 randomly selected medical press releases in 2005, 29% were rated as exaggerated and less than half provided appropriate caveats to their claims.²⁰ In a study of 23 press releases and 71 associated news stories about cancer genetics, two thirds of claims in the press release were at least as deterministic as the claims in the news.²¹ However, since these studies did not compare press releases with statements made in the abstracts or discussions of the associated peer reviewed journal articles, they may not be examples of exaggeration beyond what journal articles routinely include themselves. Indeed, in a study on “spin” in the reporting of randomised controlled trials (70 press releases and associated journal abstracts, 41 news stories), in only four cases the news contained spin where the associated journal abstract did not.²²

We aimed to clarify how often news contains claims or advice from health related research that go beyond those in the peer reviewed journal articles, and to identify the likely source of these exaggerations (press releases or news). Furthermore, we tested whether exaggerations in press releases were associated with a higher likelihood of news coverage, compared with press releases without exaggeration.

Methods

From publicly accessible university repositories we identified all the press releases based on published studies with possible relevance for human health (biomedical and psychological sciences; fig 1) issued in 2011 by the Russell Group universities (the 20 leading UK research universities). We selected these universities as a clearly defined group with international prominence; we did not expect differences between this sample and other UK or international press releases (see for example^{20 21}). For each relevant press release (n=462) we sourced the associated peer reviewed journal article and print or online news stories (n=668) from national press using the Nexis database, BBC, Reuters, and Google (we did not include broadcast news; the number of news stories per press release ranged from 0-10). We coded each journal article, press release, and news set using a detailed protocol available online (<http://dx.doi.org/10.6084/m9.figshare.903704>; supplementary information SI sections 1-3 provide full details of our sample and methods). Each set took on average 3-4 hours to code. We double coded 27% of press releases and journal articles and 21% of news stories (concordance rate 91%, mean $\kappa=0.88$; given the large number our simulations show that 10% disagreement would not influence our conclusions, see supplementary section SI7).

Taking the peer reviewed paper as a baseline (which is not to assume that peer reviewed publications are true; many already

contain exaggeration), we sought cases where news stories offered advice to readers, made causal claims, or inferred relevance to humans beyond (or different to) that stated in the associated peer reviewed paper. Given the likelihood that some statements in journal articles themselves would be considered exaggerated by other scientists in the specialty, our overall levels of measured exaggeration are likely to be underestimates. We then asked whether such discrepancies were already present in the corresponding press release. For example, if a study reported a correlation between stress and wine consumption and the news story claimed that wine causes stress, what did the press release say? Similarly, if a news story claimed a new treatment for humans but the study was on rodents, what did the press release say?

We focused our study on analysing advice to readers to change behaviour, causal statements drawn from correlational results (cross sectional and longitudinal observational data), and inference to humans from animal research.²³ Explicit advice clearly has the potential to influence behaviour, as do causal claims about what factors influence health. It is notoriously difficult to ascertain cause from correlational results. For example, a correlation between consuming wine and a disease could occur because wine increases the risk of the disease, the disease increases the consumption of wine, or the consumption of wine correlates with another factor that is associated with the increased risk. For animal research, it is estimated that less than 10% of non-human investigations ever succeed in being translated to human clinical use.²⁴ Over-selling the results of non-human studies as a promised cure potentially confuses readers and might contribute to disillusionment with science.¹¹

Advice

We coded each journal article, press release, and news story for the maximum level of advice it contained using four levels based on explicitness and directness: no advice, implicit advice (for example, “these findings suggest that mid-late childhood may be the best bet for childhood obesity prevention”, “simply exercising with a best friend or having a friend who is a good exercise role model increases the chance of a child keeping fit and active”), explicit advice, but not to the reader or general public (for example, “I think we now have enough evidence to say that pulse oximetry screening should be incorporated into everyday clinical practice”, “ambulatory monitoring is recommended for most patients before the start of hypertensive drugs”), and explicit advice to the reader or general public (for example, “children who are thirsty should be encouraged to drink water”, “for anyone considering taking aspirin I would recommend . . .”). Relevant samples for analysis of exaggeration of advice were those containing at least one implicit or explicit advice statement anywhere in the journal article or press release or news (n=213 press releases, n=116 press releases with news; n=360 news stories).

Causal statements from correlational results

For journal articles, press releases, and related news stories associated with correlational results we coded for the strength of the main statements of the findings. For press releases and news we used the title and first two sentences as their main statements, since nearly all follow the “inverted pyramid” structure of stating their main claims first.²⁵ For journal articles we used the abstract and discussion. We used a seven point scale to rate increasing levels of determinism, where the presence of stronger statements trumped weaker ones: no statement (in which case no further comparison was possible), explicit statement of no relation, correlational (for example, “drinking

wine is associated with increased cancer rates”), ambiguous (for example, “drinking wine linked to cancer risk”), conditional causal (for example, “drinking wine might increase cancer risk”), can cause (for example, “drinking wine can increase cancer risk”), and unconditionally causal (for example, “drinking wine increases cancer risk”). For analysis of causal claims we focused on correlational research, which we defined as observational cross sectional and longitudinal designs. We did not analyse qualitative, interventional, or simulation designs. We coded the first claim statement for our primary analysis (relevant samples for analysis were 182 press release, 95 with news; 261 news stories). Where a second statement occurred about a different variable pair, we also coded these for replication (see supplementary section SI5 for analysis).

Conclusions for humans from studies in non-humans

For each non-human study (animals, cells, or simulations), we coded whether the main statements of press release and news were phrased as explicitly non-human, implicitly human (for example, “a pregnant mother’s stress level affects the brain of her unborn baby”), or explicitly human (for example, “a pregnant woman’s stress . . .”). For journal articles we searched the discussion section and abstract for any statements about human relevance. Relevant samples for analysis were 105 press releases, 48 with news; 115 news stories.

Caveats and justifications

We searched the whole press release and news stories for any caveats stated for the advice, causal claims, or inference to humans (for example, “This is a population study. It cannot say definitively that sugary drinks raise your blood pressure, but it’s one piece of the evidence in a jigsaw puzzle”, “The scientists who carried out the study emphasized that they could not say for certain . . .”). Similarly, we searched for justifications of the advice, claims, or inference (for example, “even after taking into account the effect of extra body weight on blood pressure, there was still a significant link with sweetened drinks”).

Study facts and quotes

We also coded facts about the study and press release, including sample size, duration, completion rate, and the source of quotes. These are analysed in section SI11 of the supplementary file.

Further details of the coding methodology are given in section SI2 of the supplementary file. All coding sheets (n=462), full instructions for coding, and data analysis files and programs are available online (<http://dx.doi.org/10.6084/m9.figshare.903704>).

Statistical analysis

We used generalised estimating equations to calculate percentages and 95% confidence intervals for news with exaggeration relative to what was present in the journal article, while adjusting for the clustering of several news articles to one press release (using an exchangeable working correlation). The generalised estimating equations framework was also employed to estimate the association (in odds ratios) between exaggeration in the press release and exaggeration in the news. Note that these analyses included only those journal articles and press releases for which there was at least one news story (and the news could be appropriately coded for the relevant analysis). We compare the characteristics of press releases with and without associated news, using bootstrapped 95% confidence intervals and standard inferential statistical tests.

Results

Exaggeration rates in press releases

For our analysis of advice we found that 40% of the press releases contained more direct or explicit advice than did the journal article (bootstrapped 95% confidence interval 33% to 46%). For our analysis of statements based on correlational results (cross sectional or longitudinal) we found that 33% of primary claims in press releases were more strongly deterministic than those present in the associated journal article (bootstrapped 95% confidence interval 26% to 40%). For studies on animals, cells, or simulations, 36% of press releases exhibited inflated inference to humans compared with the journal article (bootstrapped 95% confidence interval 28% to 46%). Given the likelihood that some statements in journal articles themselves would be considered exaggerated by other scientists in the speciality, our levels of measured exaggeration are likely to be underestimates.

Association of news exaggeration with press release exaggeration

Figure 2 [↓](#) summarises the rates of exaggeration in news for press releases that did or did not already contain exaggeration. For advice, overall 36% of news (95% confidence interval 29% to 44%) contained more direct or explicit advice than did the journal article. The odds of exaggerated advice in news was 6.5 times higher (odds ratio 6.5, 95% confidence interval 3.5 to 12.4) when the press release contained exaggerated advice (58%, 95% confidence interval 48% to 68%; see table [↓](#) for the odds) than when it did not (17%, 10% to 24%; difference 41%, 95% confidence interval 28% to 53%).

For main news statements about correlational results, 39% (95% confidence interval 31% to 49%) were more strongly deterministic than those present in the associated journal article. The odds of exaggerated statements in news was 20 times higher (95% confidence interval 7.6 to 51) when press release statements were exaggerated (81%, 95% confidence interval 70% to 93%) than when they were not (18%, 9% to 27%; difference 63%, 95% confidence interval 49% to 78%).

For non-human studies, 47% of news contained inflated inference to humans. The odds of exaggeration in news was 56 times higher (95% confidence interval 15 to 211) when press release statements were exaggerated (86%, 95% confidence interval 77% to 95%) than when they were not (9.6%, 0% to 19%; difference 76%, 95% confidence interval 63% to 89%). See supplementary section SI5-7 for further details of these results.

Effect of exaggeration in press releases on news uptake

A key motivation for inflating advice, causal inference, or inference to humans in press releases may be the assumption that it greatly increases news uptake. Contrary to our expectations, however, the proportion of press releases with at least one associated news story did not differ significantly between press releases with exaggeration and those without for any of our three analyses (figure 3 [↓](#) and table), although in this dataset we cannot assess what the news uptake would have been for identical press releases with and without exaggeration. While there was a small numerical increase in news uptake with exaggerated press releases, any real effect is unlikely to be greater than the upper confidence intervals. For advice, 66/128 (52%) press releases without exaggeration had news uptake compared with 50/85 (59%) press releases with exaggerated

advice (bootstrapped 95% confidence intervals of the difference -6.4% to 21%). For causal claims from correlation, 61/122 (50%) press releases without exaggeration had news uptake compared with 34/60 (57%) press releases with exaggerated claims (95% confidence intervals of the difference -9% to 22% ; see supplementary SI5 for secondary statements). For inference to humans, 29/67 (43%) press releases without exaggeration had news uptake compared with 19/38 (50%) press releases with exaggerated advice (95% confidence intervals of the difference -13% to 27%).

Further, there was no statistical support for the idea that when press releases do successfully generate news, exaggeration would be linked with more associated news stories. As for percentage news uptake, any real effect is unlikely to be greater than the upper confidence intervals. Non-exaggerated advice was associated with 2.8 news stories per press release, whereas exaggerated advice was associated with 3.4 news stories per press release (95% confidence intervals of difference -0.3 to 1.5). Non-exaggerated main causal claims were associated with 2.8 news stories per press release, whereas exaggerated causal claims were associated with 2.7 news stories per press release (95% confidence intervals of difference -1.0 to 1.0). Non-exaggerated inference to humans was associated with 2.3 news stories per press release, whereas exaggerated inference was associated with 2.5 news stories per press release (95% confidence intervals of difference -0.8 to 1.1).

Between universities there was also no evidence that higher rates of inflated claims in press releases attracted more news uptake. The percentage of inflated advice, causal statements, or inference to humans in press releases varied from 11% to 50%, while the proportion of press releases with news varied from 8% to 87%, but these did not significantly correlate ($r=0.13$; see supplementary section SI9).

We also tested whether explicit caveats mentioned about advice, causal statements, or inference to humans from animal research in press releases were associated with reduced news uptake, as many scientists and press officers might fear. Overall, caveats were rare in press releases, and there was a clear association between their presence or absence in press releases and in news (see supplementary section SI8). But we found no evidence for an effect on uptake; if anything, caveats to causal statements might be associated with higher uptake (69% v 51%, bootstrapped 95% confidence intervals of difference -0.1% and 35% ; note that numbers are small for caveats). We also coded the presence of justifications for advice, causal claims, and inference to humans that would help readers judge which statements are warranted; these were also rare, highly associated between press release and news, but with no evidence for an effect on news uptake (see supplementary section SI10).

Discussion

Although it is common to blame media outlets and their journalists for news perceived as exaggerated, sensationalised, or alarmist, our principle findings were that most of the inflation detected in our study did not occur de novo in the media but was already present in the text of the press releases produced by academics and their establishments. Among biomedical and health related press releases issued by Russell Group universities in 2011, 33% to 40% contained exaggerated statements compared with the corresponding peer reviewed journal articles. Moreover, when press releases contained exaggeration it was likely that the news would too (58% for advice, 81% for causal claims, and 86% for inference to humans, fig 2), but when press releases did not contain exaggeration, rates of exaggeration in

news were only 17%, 18%, and 10%, respectively. Therefore the odds of exaggerated news were substantially higher when the press releases issued by the academic institutions were exaggerated (odds ratios 6.5, 20, and 56, respectively).

Caveats for our observational approach

Our study was correlational, so does not demonstrate a causal relation between inflated statements in press releases and inflated news. For example, if journalists did not read the press releases, associated exaggeration could nevertheless emerge between the press release and news because of features in the journal articles that might naturally lead to such exaggerations. However, many sources of converging evidence point to press releases as the main source of science news,¹⁷⁻²⁷ including the quotes and study facts analysed from our data (see supplementary section SI11). Although some of the studies will have had press releases released from both university and journal, this could only increase the proportion of occasions when exaggeration is already contained in at least one important press release source. It is not yet known whether exaggeration rates in press releases issued by journals differ noticeably from those issued by universities; our ongoing research is exploring this further.

Changes in presentation style between peer reviewed papers and press releases are expected in order to spark the interest of journalists. But seeking simplification and stimulating interest does not justify exaggeration. Moreover, contrary to common assumption, we did not find evidence that exaggerated statements in press releases are more likely to attract news uptake or substantially increase the number of news articles when they do occur. We also found no indication that caveats in press releases reduce uptake, although presumably the fear that they do is the reason caveats are so rare. These aspects of our results should be clarified by further research. It may not be simply the case that similar press uptake would be achieved with non-exaggerated headlines and inclusion of caveats. For example, press releases with exaggeration may not be based on journal articles with news value equal to those without exaggeration. Similarly, caveats may have been included in our sample of press releases only where likely press interest was already judged to be sufficiently strong.

Using journal articles as the baseline

Since we are not experts in every discipline (and experts also disagree), we did not attempt to code whether changes to advice, causal claims, and inference to humans from animal research were scientifically justified. It is possible that some journal articles are worded over-cautiously, and in these cases stronger or more direct statements in press releases might be justifiable (although our results showed that they are rarely explicitly justified in press releases, see supplementary section SI10). However, we assume that pressure to publish means that most journal articles already contain the highest level (at least) of justifiable inference and advice; if further inflation occurs in a press release, it is thus likely to go beyond what a consensus of scientific opinion would find acceptable. Consistent with this interpretation, a preliminary survey (see supplementary section SI12 and figure S2) revealed that a surprising number of scientists were willing to say that their press releases were exaggerated (relative to their own judgment of what was scientifically justified). Furthermore, given the imperfections of peer review, many journal articles may contain statements that are already exaggerated relative to a consensus of scientific opinion, or at least spun to emphasise positive findings,²² and thus our measured level of within university exaggeration is likely to underestimate the extent of the problem.

Implications for practice

It is important that these results are not perceived as simply shifting the blame from one group of non-scientists (journalists) to another (press officers). Most press releases issued by universities are drafted in dialogue between scientists and press officers and are not released without the approval of scientists²⁰ (and confirmed in our survey, see supplementary section SI12), and thus most of the responsibility for exaggeration must lie with the scientific authors. At the other end of the chain, journalists have a continuing responsibility to cross check their sources even if their working conditions make that increasingly difficult. The blame—if it can be meaningfully apportioned—lies mainly with the increasing culture of university competition and self promotion, interacting with the increasing pressures on journalists to do more with less time. It is interesting in this context that news outlets were broadly similar in the degree of exaggeration between press release and news (see supplementary section SI13).

Our findings may seem like bad news but we prefer to view them positively: if the majority of exaggeration occurs within academic establishments, then the academic community has the opportunity to make an important difference to the quality of biomedical and health related news. Arguably it would be far more difficult to change the working practices and cultures of journalists at independent news organisations. Furthermore, we are not arguing that accurate (or appropriately cautious) claims are sufficient for the public readership to make well informed choices in health related issues (that is, the discredited information deficit model).¹⁹ The potential influence of the media on the opinion and behaviour of different publics is complex and other factors are involved.¹⁻⁶ What we do argue is that appropriate claims are a necessary starting point, that misleading claims can do harm, and that since many such claims originate within universities, the scientific community has the ability to improve this situation.

We thank our professional advisory group for input into this research, including Rob Dawson (BBSRC), Fiona Fox (Science Media Centre), Ruth Francis (Biomed Central), James Gallagher (BBC), Helen Jamison (Science Media Centre), Alok Jha (*The Guardian*), Emma Little (*The Sun*), Carmel Turner (Medical Research Council), Bob Ward (Grantham Institute), and Ed Yong (*National Geographic*); Ananyo Bhattacharya and Simon Dymond for helpful discussions; and Eleanor-Rose Corney and Caitlin Argument for research assistance.

Contributors: PS, CDC, AW, JB, and FB conceived and planned the research. SV-G, AD, JO, LW, BH, and BD carried out the research, supervised by PS and CDC and funded by grants awarded to PS and CDC. SV-G, PS, CDC, CAV, and JB analysed and presented the data. PS, CDC, JB, and AW wrote the paper. PS, CDC, JB, and CAV revised the paper. PS and CDC contributed equally to this work and are the guarantors.

Funding: This study was supported by grants from the British Psychological Society, Experimental Psychology Society, Wales Institute of Cognitive Neuroscience, the Wellcome Trust, the Economic and Social Research Council, the Biotechnology and Biological Sciences Research Council, and Cardiff University.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that the study was

supported by grants from the British Psychological Society, Experimental Psychology Society, Wales Institute of Cognitive Neuroscience, the Wellcome Trust, the Economic and Social Research Council, and the Biotechnology and Biological Sciences Research Council, and Cardiff University; no financial relationships with any organisations, except universities, that might have an interest in the submitted work in the previous three years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: All coding sheets (n=462), full instructions for coding, summary data files, and analysis programs are available online (<http://dx.doi.org/10.6084/m9.figshare.903704>).

Transparency: The lead authors (PS and CDC) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

- Lewis J. Constructing Public Opinion. Open University Press, 2001.
- Grilli R, Ramsay C, Minozzi S. Mass media interventions: effects on health services utilisation. *Cochrane Database Syst Rev* 2002;1:CD000389.
- Phillips DP, Kanter EJ, Bednarczyk B, Tastad PL. Importance of the lay press in the transmission of medical knowledge to the scientific community. *N Engl J Med* 1991;325:1180-3.
- Sharma V, Dowd MD, Swanson DS, Slaughter AJ, Simon SD. Influence of the news media on diagnostic testing in the emergency department. *Arch Pediatr Adolesc Med* 2003;157:257-60.
- Boyce T. Health, risk and news: the MMR vaccine and the media. Peter Lang, 2007.
- Kitzinger J. Framing abuse. Pluto Press, 2004.
- Ramsay ME. Measles: the legacy of low vaccine coverage. *Arch Dis Childhood* 2013;98:752-4.
- Deer B. How the case against the MMR vaccine was fixed. *BMJ* 2011;342:77-82.
- Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Med* 2008;5:e95.
- Cooper BEJ, Lee WE, Goldacre BM, Sanders TA. The quality of the evidence for dietary advice given in UK national newspapers. *Public Underst Sci* 2012;21:664-73.
- Leveson B. An inquiry into the culture, practices and ethics of the press. Stationery Office, 2012: pp22, 80, 690-1, 1803. ISBN 9780102981100.
- Gandy OH. Beyond agenda setting. Ablex Publishing, 1982.
- Nelkin D. Selling science. W H Freeman, 1995.
- Lewis J, Williams A, Franklin B. A compromised fourth estate? *Journal Stud* 2008;9:1-20.
- Bauer MW, Bucchi M. Journalism, science, and society. Routledge, 2008.
- Kiernan V. Embargoed science. University of Illinois Press, 2006.
- Williams A, Clifford S. Mapping the field: specialist science news journalism in the UK national media. Science and the Media Expert Group, Department of Business, Innovation and Skills, 2009.
- Schwartz LM, Woloshin S, Andrews A. Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study. *BMJ* 2012;344:d8164.
- Sturgis P, Allum N. Science in society: re-evaluating the deficit model of public attitudes. *Public Underst Sci* 2004;13:55-74.
- Woloshin S, Schwartz LM, Casella SL. Press releases by academic medical centers: not so academic? *Ann Intern Med* 2009;150:613-8.
- Brechman J, Lee CJ, Cappella JN. Lost in translation?: a comparison of cancer-genetics reporting in the press release and its subsequent coverage in the press. *Sci Commun* 2009;30:453-74.
- Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med* 2012;9:e1001308.
- Science Media Centre. 10-best-practice-guidelines-for-science-and-health-reporting. 2012. www.sciencemediacentre.org.
- Van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? *PLoS Med* 2010;7:e1000245.
- Pottker H. News and its communicative quality: the inverted pyramid—when and why did it appear? *Journal Stud* 2003;4:501-11.
- Bartlett C, Sterne, J, Egger, M. What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *BMJ* 2002;325:81.
- De Semir V. Press releases of science journal articles and subsequent newspaper stories on the same topic. *JAMA* 1998;280:294-5.

Accepted: 5 November 2014

Cite this as: *BMJ* 2014;349:g7015

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

What is already known on this topic

Health related news has widespread potential to influence health related behavior but often misreports the science
 It is not known whether exaggerations originate in the news stories themselves or in press releases issued by academic institutions producing the research

What this study adds

Most exaggeration in health related science news was already present in academic press releases
 Exaggeration was not significantly associated with increased news coverage, relative to press releases without overstatement
 Press releases could be a primary target to improve the accuracy of science news, with potential benefit for public health

Table

Table 1 | Summary of results for analyses of advice, primary claims from correlational data (causal claims), and human inference from non-human studies (human inference)

Variables	No	PR with news	No with news	Odds news uptake	Odds ratio (95% CI)	Odds news exaggerated	Odds ratio (95% CI)
Advice:							
PR not exaggerated	128	66	188	1.1	1.3 (0.8 to 2.4)	0.2	6.5 (3.5 to 12.4)
PRs exaggerated	85	50	172	1.4		1.4	
Total	213	116	360	—		—	—
Causal claims:							
PR not exaggerated	122	61	169	1	1.3 (0.6 to 3.0)	0.2	19.7 (7.6 to 51.4)
PRs exaggerated	60	34	92	1.3		4.3	
Total	182	95	261	—	—	—	—
Human inference							
PR not exaggerated	67	29	68	0.8	1.3 (0.7 to 2.5)	0.1	56.1 (14.9 to 211)
PRs exaggerated	38	19	47	1		5.9	
Total	105	48	115	—	—	—	—

PR=press release.

The key results are that odds ratios for the dependence of news uptake on PR exaggeration are indistinguishable from 1, whereas odds ratios for the dependence of news exaggeration on PR exaggeration are much larger and clearly distinguished from 1. See the results section and figures 2 and 3 for further information, including percentages and 95% confidence intervals.

Figures

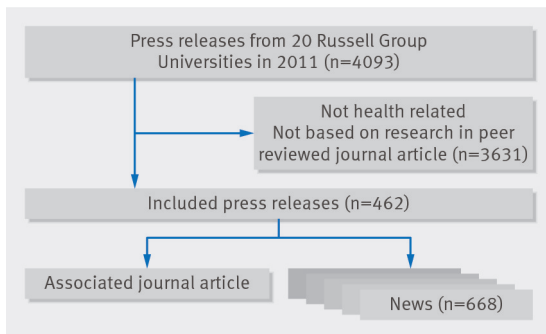


Fig 1 Identification of press releases based on published studies with possible relevance for human health (biomedical and psychological sciences)

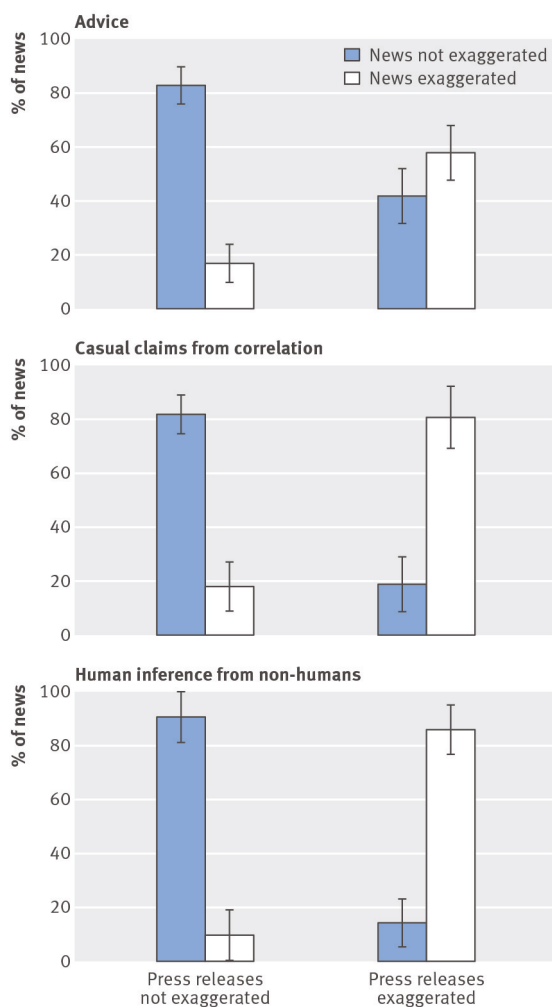


Fig 2 Proportions of news with exaggerated advice, causal statements from correlational research, or inference to humans from non-human studies were higher when the associated press releases contained such exaggeration. Error bars are 95% confidence intervals. See table for odds ratios

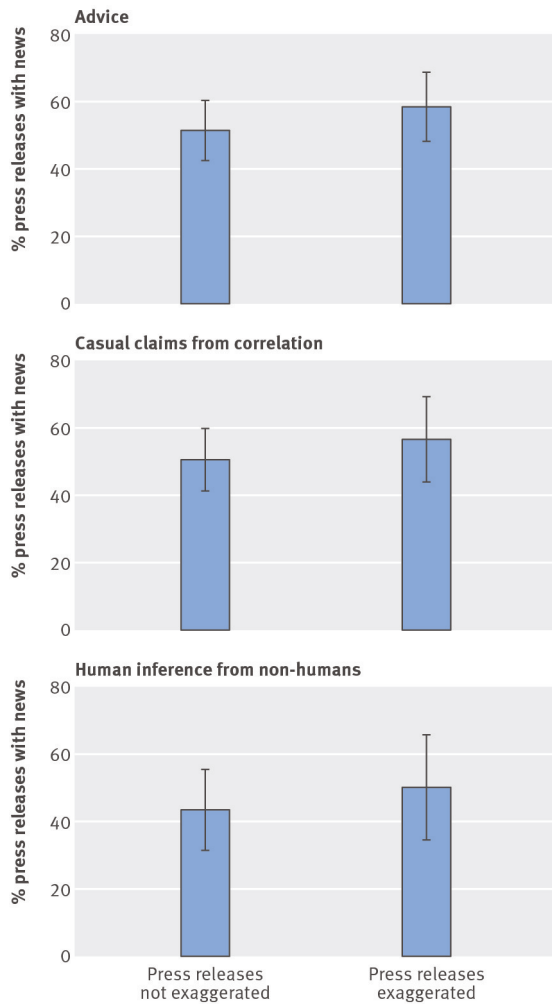


Fig 3 The proportion of press releases with some news uptake (at least one news story) was not statistically distinguishable regardless of whether the press release did or did not contain exaggerated advice, causal statements, or inference to humans from animal research. Furthermore, the mean number of news stories per press release did not significantly differ with exaggeration (data not in figure, see text). Error bars are bootstrapped 95% confidence intervals. See table for odds ratios

The association between exaggeration in health-related science news and academic press releases: retrospective observational study

Supplementary Information

List of acronyms

DV, Dependent variable; IV, Independent variable; JA, Journal article; PR, Press release.

SI1. Materials

Press releases (PR). PRs from Russell Group members (in 2011) were obtained from the online press site of each university (University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, University of Edinburgh, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Liverpool, London School of Economics, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, Queen's University Belfast, University of Sheffield, University of Southampton, University College London, University of Warwick). All press releases published in 2011 were inspected, and those not relating to human health, psychology, or neuroscience were excluded. Also excluded were those about future research not yet performed.

Journal articles (JA). The JAs corresponding to each press release were obtained in several ways. Some PR contained a full link; others reported the lead author name, and either the article title, journal name or both. In all such cases, an online search identified the article. In cases where only the lead author name was mentioned, the author's full publication list was found on the university web page, and each publication in 2010, 2011 and 2012 was inspected until the JA corresponding to the PR could be identified. If the JA could not be identified (usually after 2-4 hours of search), that PR was excluded from the analysis. For 6 PRs there were two associated JAs, and in 1 case 3 JAs.

News stories. News sources included were: *Associated Press, BBC News, Daily Mail, Daily Star, Daily Star Sunday, Economist, Express, Financial Times, Guardian, Guardian.co.uk, Independent, Mail on Sunday, Mail Online, Metro, Mirror, New Scientist, Observer, People, Press Association, Racing Post, Reuters, Scotsman, Sun, Sunday Express, Sunday Mail, Sunday Mirror, Sunday Post, Sunday Sun, Sunday Telegraph, Sunday Times, Telegraph, Times.*

First, the key words of the press release were searched on the Nexis® database (<http://www.lexisnexis.com/uk/nexis/>). The search was limited to UK national newspapers and articles published within 30 days of the PR. Each search was performed several times with different key words, from specific to more general. For example, for the PR titled "Scientists demonstrate potential new treatment for most common form of infant leukaemia", the search would be performed with key words "BET proteins AND MLL", "proteins AND leukaemia", "new treatment AND leukaemia", and "treatment AND cancer". In the same way, BBC news webpages (<http://www.bbc.co.uk/news/>) and Reuters webpages (<http://uk.reuters.com/>) were searched. Finally, a google search was performed using the key words and the name of each individual newspaper. In cases where several news articles were found from the same newspaper, the longer article was selected for coding.

SI2. Coding

The coding template and full instructions for coders are available online (<http://dx.doi.org/10.6084/m9.figshare.903704>). These were designed by the authors to address the three main questions of interest (advice to readers, causal conclusions from correlation, and generalizations from non-human samples to humans). It was developed using a pilot batch of materials by iterative expansion and modification in order to capture aspects relevant to our main questions that were not adequately coded initially. Pilot testing was used to clarify potential ambiguity, and to solve coding difficulties and instances of low inter-rater reliability.

Each set of JA, PR and news took 3-4 hours to code on average, but varied depending on clarity of the materials, number and length of news articles and complexity of the study.

Advice. The PR, news stories and JA (abstract and discussion) were read for statements of implicit or explicit advice. Coding levels were:

- 0, *No advice* (including advice to researchers, for example to perform further study).
- 1, *Implicit advice* (e.g. “Eating chocolate might be beneficial for...”).
- 2, *Explicit but not to reader or general public* (e.g. “Doctors should advise patients to...”)
- 3, *Explicit to reader / general public* (e.g. “Expectant mothers should...”).

A range of examples of inflation is given below (note that not all would be considered inappropriate if they are just changes to the intended audience; our purpose is not to evaluate each inflation, but to find the source given that inflation is generally blamed on journalists). PR: *Mothers who want to breastfeed should be given all the support they need* (code 2); News: *Mums should breastfeed for at least four months to avoid having naughty kids, experts now advise* (code 3). PR: *If these weather patterns continue, both forage and dairy management will have to adapt to maintain current milk quality* (code 1); News: *spend 9p extra a pint and save Daisy the Dairy Cow, in her straw hat* (code 3). JA: *the data we present add to growing justification to monitor the health of preterm men and women beyond infancy and childhood* (code 1); PR: *we need to monitor the health of premature babies beyond infancy and childhood* (code 2). JA: *These specific defects should be included in public health educational information to encourage more women to quit smoking* (code 2); PR: *women should quit smoking before becoming pregnant, or very early on, to reduce the chance of having a baby with a serious and lifelong physical defect* (code 3). PR: *It is possible that good nutrition during the first three years of life may encourage optimal brain growth* (code 1); News: *People should seek advice from a registered dietician, but simply it's a message of moderating fat intake, five fruit and veg a day and whole grain starchy foods* (code 3). PR: *Our findings support the concept of more widespread HIV testing* (code 1); News: *if you've been at risk for HIV, get tested now* (code 3).

Causal statements from correlational research. For each PR and news story, the IV (or pseudo IV in correlational designs), DV and stated relationship between them (if any) were extracted from the main claims, which were operationalized as the title plus first two sentences in PRs and news. For the JA, main claims were defined within the abstract and discussion sections. If there were claims about more than one set of IV and DV in the PR or news, a second set was also coded and the same sets were identified in the JA, allowing us to test whether the findings for the main statements are replicated in the second statements (SI5).

In order to consistently code the 6 levels of relationship statement we drew up a table of examples from the first stage of coding. These were:

0. *No relationship stated (but could have been):* The study must have contained at least two variables (IV and DV, or pseudo IV and DV) between which a relationship could have been stated. If there were not two suitable variables, the code 'not applicable' was used.
1. *Statement of NO relationship/cause:* e.g. 'no difference'; 'persists without'; 'does not result in'; 'no significant extra risk'; 'added no benefit'.
2. *Statement of correlation:* Meaning remains if the variables in the statement are reversed; e.g. 'associated with'; 'related to'; 'varies with'; 'increases as'; 'decreases as'; 'have higher/lower rates of'; 'linear relationship';
3. *Ambiguous statement of correlation/cause:* Strictly does not state cause, but contains stronger implication of direct relationship than level 2; e.g. 'linked to'; 'predicts'; 'connected to'; 'more successful'; 'had difficulty';
4. *Conditional cause:* Causal statement (examples in level 6) preceded by conditional, such as 'might', 'could', 'may', 'appears'; or a statement that implies 'might cause', e.g. 'likely to be a critical event', 'demonstrates potential'.
5. *Can cause:* Causal statement (examples in level 6) preceded by 'can'. 'Can cause' is more deterministic than 'may cause' because it denotes ability, and implies 'does sometimes', while 'may' denotes a potential not yet realised. Note that although 'could' and 'might' are grammatically related to 'can' and 'may', respectively, 'could' is closer in meaning to 'might' in common usage than 'can' is to 'may'. Thus we grouped 'could' with 'might' in category 4.

6. *Causal statement*: Direct causal statement; e.g. 'enhances', 'increases' [as transitive verb - X increases Y], 'decreases', 'cuts', 'ameliorates'; 'influences'; 'prevents'; 'raises/reduces'; 'enables'; 'determines'; 'is vital for'; 'boosts/improves'; 'drives'; 'leads to'; 'transforms'; encourages' etc.

Where statements of different levels were made within the analysed segments of text, stronger statements trumped weaker ones. Separately, we also coded whether or not the statement of relationship was explicitly probabilistic – for example, 'correlated with the risk of...' (correlational probabilistic); 'raises the chance of...' (causal probabilistic); Further probabilistic words/phrases included: 'likelihood'; 'makes more likely'; 'tendency'; 'rate'.

For analysis of causal claims we focused on correlational research; we coded type of study design using 6 categories: Qualitative; Correlational cross-sectional; Correlational longitudinal; Intervention (not full RCT); Full randomised controlled trial (RCT); Modelling / Simulation. We did not detect any differences in the distribution of causal statement levels between cross sectional and longitudinal correlational designs; therefore we grouped these together into a single correlational category for further analysis. We did not analyse qualitative, interventional or simulation studies further. We checked whether the IVs and DVs themselves got distorted, changed or generalised in the progression of claims. We find that this happened in PR for IVs in only 11/573 samples, and for DVs in only 6/554; similarly in news only 21/726 for IVs and 11/740 for DVs.

A range of examples of inflation is: JA: *This observational study found significant associations between use of antidepressant drugs and several severe adverse outcomes in people aged 65 and older with depression* (code 2); PR: *New antidepressants increase risks for elderly* (code 6). JA: *Reported flooding experiences had a significant relationship with perceptions relating to climate change* (code 2); PR: *Direct experience of extreme weather events increases concern about climate change* (code 6). JA: *A brief TCBT or exercise program was associated with substantial, significant, clinically meaningful improvements in self-rated global health* (code 2); PR: *Talking therapy over the phone improves symptoms of chronic widespread pain* (code 6). JA: *deregulation of a single kinase in two distinct cellular compartments... is intricately linked to implantation failure and miscarriage* (code 3); News: *The protein SGK1 in the lining of the womb makes it harder to get pregnant* (code 6). JA: *bisphosphonate use is associated with a significantly lower rate of revision surgery of up to about 50% ... in patients without a previous fracture* (code 2); News: *Bisphosphonates 'extend hip replacement life'* (code 6). JA: *human orbital volume significantly increases with absolute latitude* (code 2); News: *... gives you a bigger brain* (code 6). JA: *... association between RXRA chr9:136355885+ methylation and mother's carbohydrate intake* (code 2); PR: *During pregnancy, a mother's diet can alter the function of her child's DNA* (code 5).

Human conclusions from non-human studies. We coded the explicit or implicit study sample, population type or experimental participants of the main claims in JA, PR and news. We used the same code to separately identify the actual sample, population type or experimental participants of the study. If there was more than one type (e.g. rodent and human) in a JA, it was excluded from the analysis of human inference from non-human studies. The coding levels were:

1. Explicitly human: explicit mention of e.g. *humans, people, participants, women, men, girls, boys*; pronouns e.g. *we, our, your*; mention of specific study designs e.g. *genome-wide association study, prospective cohort study*; mention of activities that relate to humans only, e.g. *mobile phone use, reading books, A levels*.
2. Implicitly human: no explicit mention of sample, yet the sentences are judged to refer to humans, e.g. *eating chocolate causes cancer, or new drug found for Alzheimer's*
3. Non-human primates: explicit mention of *chimpanzees, monkeys*, etc.
4. Rodents: explicit mention of *mice, rats, rodents*
5. Other animals/organisms: explicit mention of *fruit flies, bacteria, worms, birds, fish*, etc.
6. Cells *in vitro*: explicit mention of cell cultures, e.g. *stem cells, T cells*, etc.
7. Simulations: explicit mention of *mathematical modelling or computer simulations* etc.

A range of examples of inflation is: JA: *We have developed a broadly effective strategy to polymerize monomeric Fc-fusion proteins...* (code 6); PR: *Better and more affordable treatments for sufferers of autoimmune diseases* (code 1). JA: *An animal model of this important component of memory consolidation...* (code 4); PR: *Scientists have shed light on why it is easier to learn about things related to what we know* (code 2). JA: *In mouse tooth development, Barx1*

expression is restricted to presumptive molar mesenchyme and throughout tooth development to molar mesenchyme cells (code 4); PR: Researchers have uncovered a novel mechanism they have termed 'developmental stalling', that might explain how errors in the development of human embryos are naturally corrected to prevent birth defects (code 2).

Caveats and justifications. For each section, we searched the whole PR and news stories for any caveats stated for the advice or claims (e.g. "This is a population study. It can't say definitively that sugary drinks raise your blood pressure, but it's one piece of the evidence in a jigsaw puzzle"; "The scientists who carried out the study emphasised that they could not say for certain..."). Similarly, we searched for justifications of the advice or claims (e.g. "even after taking into account the effect of extra bodyweight on blood pressure, there was still a significant link with sweetened drinks").

Study facts and quotes. We also coded various facts about the study and PR, including sample size, duration, completion rate and the source of quotes. These are analysed in section SI11 (Indicators of news sources).

SI3. Inter-rater reliability.

We double-coded 27% of PR and associated JA, and 21% of news stories. This difference is due to the fact that the PRs randomly selected for double coding had lower than average number of news stories. Inter-rater concordance was 90.5% ($\kappa = .87$) for cells relevant for the advice analysis; 86.3% ($\kappa = .84$) for cells in the analysis of causal claims and 94.4% ($\kappa = .93$) for cells analysed for human inference from non-human research. We analyzed the distribution of coding disagreements where they arose in the double-coded samples (i.e. whether each disagreement was between a code 1 and 2, or between a code 2 and 3 etc). Then within each round of the simulations in section SI7, 10% of the samples were by chance changed to another code in line with the observed distribution of coding disagreement in the double coded samples. This had a negligible effect on our results.

SI4. Association between advice, causal statements and human inference.

Of the studies contributing to the analysis of advice, 110 were included in the analysis of causal claims from correlation, while 19 were non-human studies included in the human inference analysis. There were only 14 studies that were both non-human and correlational. Thus while the analyses of advice and causation share many PRs, JAs and News, the analysis of nonhuman studies is on a largely independent sample of PRs, JAs and news.

Within the 110 correlational studies included in the advice analysis (because some level of advice was offered somewhere in JA, PR or news), the level of advice was not correlated with the level of causal claim within JA, PR or news ($r=0.02$, 0.05 and -0.003 , respectively). Within the 19 non-human studies included in the advice analysis (because some level of advice was offered somewhere in JA, PR or news), the level of advice was not significantly associated with the level of human inference ($r=0.07$, $p=0.78$; $r=0.29$, $p=0.23$; $r=-0.29$, $p=0.12$; although note that N in this analysis is small).

SI5. Secondary Statements (i.e. about a second set of variables in correlational studies)

For the secondary statements 25% (95% CI: 18-34%) were more strongly deterministic than those present in the associated JA. The odds of exaggerated statements in news were 36 times higher (OR=36, 95% CI: 7.8-148) when PR statements were exaggerated (83%, 95% CI 65-100%) than when the PR was not exaggerated (12%, 95% CI: 3.2-22%; difference=70%, 95% CI: 51-90%). Thus while secondary statements tended to be exaggerated less often (presumably because they are not the leading eye-catching statement), the association between exaggeration in PR and news is still very strong, replicating the results for main statements.

For rates of news uptake, 44/76 (58%) PRs without exaggeration had news uptake vs 13/26 (50%) PRs with exaggerated claims (bootstrapped 95% confidence intervals of the difference are -30% to +15%). Non-exaggerated secondary causal claims were associated with 3.0 news stories per PR, while exaggerated causal claims were associated with 2.2 news stories per PR (confidence intervals of the difference are -1.8 to +0.3).

SI6. Breakdown of PR exaggeration for exaggerated news

In the main analysis we categorized news and PR as exaggerated or not relative to the JA. This simple categorization did not distinguish between PRs that are exaggerated to the same extent as news and PRs that are exaggerated a bit, while the news is exaggerated further. In fact the latter case was relatively rare, and the most common scenario was for an identical level of exaggeration in PR and news. In the cases where news went beyond what was written in the JA, Figure S1 shows the proportions of cases when the associated PR contained no exaggeration relative to the JA (left solid bars in each plot, labeled $PR \leq JA$) or when the PR did contain exaggeration relative to the JA (the three rightward solid bars in each plot, labeled $PR > JA$). Within the cases where the PR went beyond the JA ($PR > JA$), we plot the proportions when the news was further exaggerated from the PR ($N > PR$), when the news had equivalent statements to the PR ($N = PR$) and when the news was deflated again from the PR ($N < PR$, but remember news is still inflated relative to JA in order to qualify for this analysis). The key results are that we consistently found the largest category to be 'N=PR'; in other words, when the news was inflated relative to the JA, the most likely scenario for the PR was that it said the same as the news.

By adding this category ($N = PR$) to cases where PR was even more inflated than news ($N < PR$), we find that the PR was at least as inflated as the news on 70% (advice), 48% (causal claims from correlation) and 75% (human inference from non-humans) of cases. The by adding in the cases where there was some inflation from JA to PR, and then some more from PR to news (second bar, $N > PR$ and $PR > JA$), we find the overall inflation rates occurring between JA and PR (78%, 75%, 90%). On the other hand, the inflation occurring between PR and news (30%, 52%, 25%) can be obtained from adding the two leftward columns: $PR \leq JA$ (remember all cases in this analysis have inflation from JA to news) and $N > PR$. Thus the rate of inflation between JA and PR consistently outweighs the rate of inflation between PR and news.

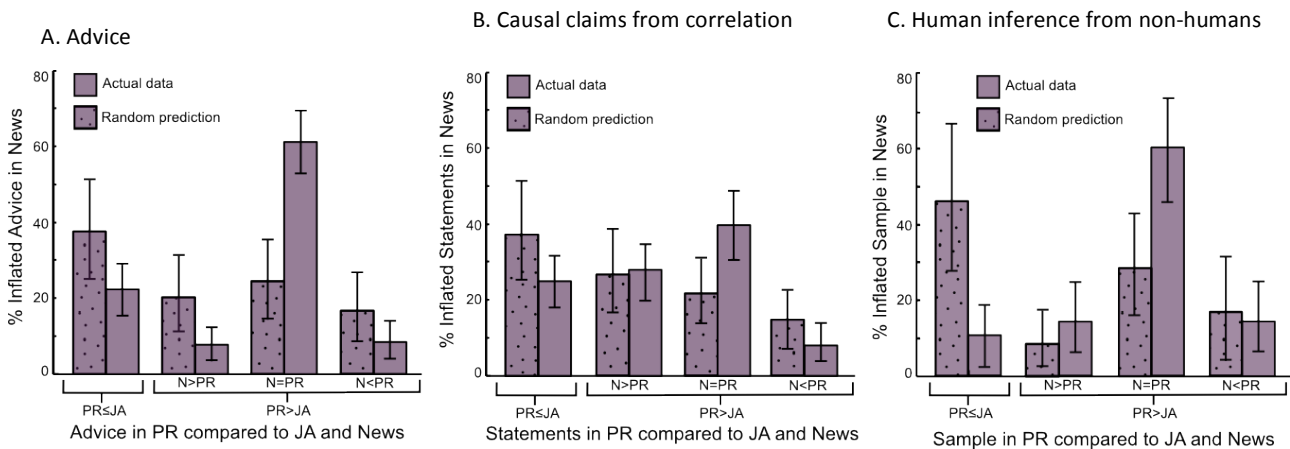


Figure S1. PR content where news contained exaggerated statements relative to the JA ($N=131, 173, 49$, respectively). In each plot, left bars ($PR \leq JA$) indicate the cases where the PR contained nothing stronger than the JA. The other bars ($PR > JA$) indicate the cases where the PR contained inflated advice or statements relative to JA, in which case there could be further inflation in the news ($N > PR$), the same strength in news and PR ($N = PR$), or occasionally, deflation from PR to news ($N < PR$). Error bars show bootstrap-estimated 95% confidence intervals (the bootstrapping preserved the clustering structure of news to PR). The consistently most frequent situation in each analysis (**A-C**) was that the PR and news were equivalent, occurring much more often than chance prediction (dotted bars and associated error bars; advice, $p < 0.001$; causal claims, $p < 0.001$; human inference, $p < 0.001$), estimated through simulating how often the observed distributions of coded levels in PR and news if written independently would produce each category plotted (see SI7). Adding the two rightmost bars together gives the proportion of cases where the PR was at least as inflated as the news (70%, 48% and 75% for A, B and C), while adding all three $PR > JA$ bars together gives the proportion of occasions that there was some degree of inflation in the step from JA to PR (78%, 75%, 90%). For comparison, adding the two left bars together gives the total proportion of cases where there was some inflation from PR to news (30%, 52%, 25%).

SI7. permutation simulation of chance associations

Table S1 presents the distributions of coded advice levels for each category of outlet. We simulated the expected number of times that chance selection from these distributions would lead to the four categories displayed in Figure S1: no inflation in PR but inflation in news; PR inflated from journal article and news inflated further; the same level of inflation in both PR and news; and news inflated relative to journal article but deflated relative to PR. For each of 10000 iterations, the JAs, PRs and news were randomly reordered with respect to each other, but preserving the distributions shown in Table S1 and the clustering structure of news to PRs (i.e. that more than one news article can come from the same PR), and the analysis was rerun to categorize the inflation level in PR when there was inflation in news, just as for the analysis of the actual data in Figure S1. We also incorporated an estimate for the effect of coding dis-concordance (see SI3). To do this, we analyzed the distribution of coding disagreements where they arose in the double-coded samples (i.e. whether each disagreement was between a code 1 and 2, or between a code 2 and 3 etc). Then within each round of the simulation 10% of the samples were by chance changed to another code in line with the observed distribution of coding disagreement in the double coded samples. Adding this effect of coding dis-concordance had a negligible effect on results.

Similarly for correlational/causal claims and for human/non-human claims, we performed equivalent simulations based on the distributions of each statement level found in each outlet (Tables S2 and S3), the clustering structure of news to PR and the observed coding disagreement distributions in the double coded samples.

Note that the comparison of the actual number of cases where PR=news to these permutation analyses is conservative, since the simulations are likely to overestimate the chance expectation of PR=news. This is because they are based on distributions for each outlet that are not, in fact, independent. If they were independent, the similarity between the distributions would likely be reduced and this in turn would reduce the estimate of the associations that would occur by chance. In the extreme of non-independence, where most news stories were to copy a restricted range of phrases in PR, the estimated chance overlap would be very high due to the paucity of potential alternative options for the random sampling. In other words, since the occurrence of coding levels is not evenly distributed, as the real overlap between PR and news becomes larger, this simulation approach stacks the cards against finding differences between the data and the simulation. Thus we can be confident that where a statistically significant difference between the data and the chance simulations can be detected despite this bias, that difference is meaningful.

Table S1. The distribution (absolute N) of coded advice levels for each category of outlet, where there was some advice in at least one of JA, PR and news.

	No Advice	Implicit	Explicit not to the reader	Explicit to the reader
Journal article	75	79	52	7
PR	42	87	48	36
News	111	93	62	71

Table S2. The distribution (absolute N) of coded levels of causal/correlational statements about correlational studies for each category of outlet.

	No statement	No relationship	Correlational	Ambiguous	Conditional	Can	Causal
JA	14	9	96	18	7	6	64
PR	29	12	44	29	20	11	69
News	24	17	58	28	35	16	135

Table S3. The distribution (absolute N) of coded levels of human inference in non-human studies for each category of outlet.

	Non-human	Implicit Human	Explicit human
Journal article	96	3	6
PR	59	13	33
News	58	27	43

SI8. Predictors of news uptake

As shown in Figure 3 (main manuscript), inflation of advice, causal claims or human inference was not reliably associated with a higher proportion of PRs attracting news or a higher mean number of news per PR. That analysis compared inflated to non-inflated statements irrespective of the actual coded level of those statements (many strong statements are not inflated because they are also contained in the JA). While inflation was our main interest, we can also analyze whether simply the coded level of PR statements (irrespective of whether they were inflated relative to JA) was associated with news uptake. The proportion of PRs with news appeared to be about 15% greater where explicit advice was present, though this was not statistically significant even without correction for multiple comparisons ($\chi^2(3)=6.1, p=0.11$). There was even less indication that the proportion of PRs with news was predicted by the strength of main causal claims ($\chi^2(6)=2.6, p=0.86$), or human inference ($\chi^2(2)=3.2, p=0.2$). Relatedly, for the mean number of news stories per PR, explicit advice was associated with an increase only if uncorrected for multiple comparisons ($F(3,458)=2.95, p=.03$; no advice 1.3, implicit advice 1.4, explicit advice not to the reader 2.0, explicit advice to the reader 2.1). Likewise, for stronger human inference there was a possible increase that would not survive the appropriate correction for three comparisons ($F(2,405)=3.86, p=0.02$ uncorrected). There was no indication of an increase with stronger causal claims ($F(6,207)=0.7, p=.7$ uncorrected). Thus overall there is some suggestion, as would be expected, that stronger advice with relevance to humans attracted more news coverage, but these effects were, perhaps surprisingly, not strong enough to be clearly significant.

SI9. Comparison between universities

Table S4 shows the number of PRs from each university included in our analyses of advice, causal claims and human inference (without double counting those included for more than one analysis), as well as the percentage of claims that were un-inflated and the percentage that had news uptake. Due to the low numbers once broken down by university, advice, main causal claims, secondary causal claims and human inference are added together to form one category of 'statements', which is why N statements differs from N PR. Note that the percent of statements *without* inflation is given (rather than *with* inflation) for straightforward multiplication with % news uptake in the combined score that estimates the % of non-inflated PR attracting news. The table is ordered by % uninflated PR claims in order to illustrate the lack of any correlation with the % news uptake ($r=-.13$). Note that in 10 cases identical press releases were released by two universities on the same research; these are included here for each university, but they were not double counted in Figure 1, or in the analyses for Figs 2 and 3.

The ranks of inflation and uptake are also shown, as well as a rank for the estimated % non-inflated PR attracting news. However, it is important to note the relatively large confidence intervals on these ranks. Ranks alone often leave it difficult for the reader to discern whether a rank order is clear-cut or largely due to small differences and random variation. We estimated the confidence intervals using the following procedure: For Birmingham, we drew [Birmingham N] times with replacement from the pool of [Birmingham N] relevant Birmingham PRs and calculated the percent inflation and uptake; then for Bristol we drew [Bristol N] times with replacement from the pool of [Bristol N] relevant Bristol PRs and calculated percent inflation and uptake; and so on for each university. Rank orders for inflation, uptake and combined scores were then found to produce a table for that round of resampling. This procedure was repeated 100000 times to create 100000 tables, from which the 95% confidence intervals

for inflation rank, uptake rank and overall rank were estimated. The CIs are generally wide, partly because of the low N for some universities.

In the case of animal research, one possible reason for PRs to generalise to humans might be to avoid advertising animal research facilities. There is currently an impetus for scientists and institutions to be more open about animal research. Recently, a 'concordat on openness on animal research' was launched, supported by 16/20 of our sample universities: University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, University of Edinburgh, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, Queen's University Belfast, University of Sheffield, and University College London.

Table S4. Rates of inflation and news uptake for health-related PRs from each Russell Group university in 2011 (see text above for explanation).

	N PR	N statements	N inflated statements	% without inflation	Inflation rank (95% CI)	% PR with news	Uptake rank (95% CI)	Combined score	Combined rank (95% CI)
University of Oxford	28	28	3	89%	1 (1-4)	54%	11 (5-16)	48%	3 (1-10)
University of Warwick	16	24	4	83%	2 (1-7)	56%	6 (4-16)	47%	4.5 (1-13)
University of Manchester	28	30	7	77%	3 (2-11)	43%	16 (8-18)	33%	11 (5-17)
University of Liverpool	12	15	4	73%	4 (1-17)	8%	20 (19-20)	6%	20 (18-20)
London School of Economics	5	10	3	70%	5 (1-19)	80%	3 (1-13)	56%	2 (1-13)
University of Leeds	17	26	8	69%	6 (3-17)	47%	14 (5-18)	33%	12 (4-18)
University of Sheffield	11	12	4	67%	7 (2-19)	55%	9 (3-18)	36%	8 (2-17)
Imperial College London	41	55	19	65%	8 (5-17)	54%	10 (5-15)	35%	10 (5-15)
University of Southampton	13	20	7	65%	9 (3-19)	62%	5 (3-16)	40%	7 (2-16)
University College London	15	17	6	65%	10 (3-19)	87%	1 (1-4.5)	56%	1 (1-8)
University of Nottingham	29	42	15	64%	11 (5-18)	38%	17 (10-19)	24%	16 (9-19)
University of Cambridge	45	53	19	64%	12 (5-17)	56%	7.5 (5-14)	36%	9 (5-14)
University of Bristol	52	73	27	63%	13 (6-17)	31%	18 (15-19)	19%	18 (14-19)
University of Edinburgh	36	40	15	63%	14 (5-18)	75%	4 (2-6.5)	47%	4.5 (2-10)
University of Glasgow	29	44	17	61%	15 (6-18)	52%	12 (5-16.5)	32%	13 (6-17)
University of Birmingham	23	24	10	58%	16 (5-19)	30%	19 (13-19)	18%	19 (13-20)
King's College London	27	40	18	55%	17 (9-20)	48%	13 (6-17)	26%	15 (8-18)
Cardiff University	18	22	10	55%	18 (7-20)	56%	7.5 (4-16)	30%	14 (5-18)
Queen's University Belfast	7	8	4	50%	19.5 (4-20)	86%	2 (1-7)	43%	6 (1-18)
Newcastle University	20	22	11	50%	19.5 (9-20)	45%	15 (6-18.5)	23%	17 (9-19)

SI10. Caveats and justifications

The overall number of caveats for advice, causation (main and secondary statements) and human inference was low in PRs (14%, 11%, 4%) and news stories (16%, 14%, 0%), and was strongly associated between news and PR: of the news with caveats, 57% and 50% had PRs with caveats. This can be compared to simulated levels of chance association of 1.3% ($\pm 0.9\%$ 95% CI) and 2.2% ($\pm 0.7\%$ 95% CI), derived similarly to the simulations described in SI7, above. Arguably, such context is always relevant, to draw attention to the research being correlational or based on animal research, and to allow readers to distinguish, for example, between a simple correlation on a small sample and a large study that controlled for many other factors. The frequency of caveats we found was even lower than that reported by Woloshin et al. (2009; ref 20 in main text) in PRs issued by 20 academic medical centres,

where 42% did not provide any relevant caveats and 90% about animal or laboratory studies lacked caveats about extrapolating to humans.

Caveats for advice were not significantly associated with news uptake: 7/12 PRs with caveats had news (58%); 46/78 PRs without caveats had news (59%); confidence intervals on the difference are -29% to +32%; $\chi^2(1)=0.002$, $p=0.97$. Caveats for causation were possibly associated with increased news uptake from 94/190 (49%) for PR without caveats to 17/24 (71%) for PR with caveats ($\chi^2(1)=3.9$, $p=0.05$ uncorrected; bootstrapped 95% confidence intervals of the difference are 1% to 40%). Caveats for human inference were too few to further analyse. Overall, Ns for caveats are too small for strong conclusions about news uptake.

Similarly, the number of justifications for advice, causal claims and human inference was low both in PR (13%, 14%, 11%) and in news (17%, 10%, 3%), and strongly associated between news and PR: of the news with justifications, 45%, 87% and 50% had PRs with justifications compared to simulated chance associations of 1.2% ($\pm 0.8\%$), 1.5% ($\pm 0.6\%$) and $<1\%$. There were no significant associations between presence/absence of justifications and news uptake, though Ns for justifications are too small for strong conclusions. For advice 7/11 [64%] with justifications had news, 46/79 [58%] without justifications had news, CIs of the difference are -27% to +26%; $\chi^2(1)=0.1$, $p=0.7$. For causation 11/25 [44%] with justifications had news, 97/177 [55%] without justifications had news, CIs of the difference are -11% to +32%; $\chi^2(1)=1.0$, $p=0.3$. For human inference 1/5 with justifications had news, 21/41 without justifications had news, CIs of the difference are -20% to +63%; small N violates χ^2 assumptions.

SI11. Indicators of news sources.

To estimate the relative importance of PR as the main source for the science stories in our sample, independently from the factors analysed for our main questions, we used dates of release, quotes and study details. Note that these estimates do not necessarily reflect all science news, given that the news stories in our sample were purposely selected to be on the same studies as those in our PR materials.

Dates. For selecting news stories, we used a criterion of release date being within 30 days of the PR. In fact, 580/668 (87%) of news stories were released within a day of the PR release date.

Quotes. We coded up to four quoted sources in news stories. Of the 668 analyzed news stories, 592 (89%) had quotes; 427 (72%) of these stories contained quotes identical to those included in the PR; 263 (44%) had alternative or additional quotes from the authors of the associated peer-reviewed journal article; 29 (5%) contained quotes identical to text in the journal article; 50 (8%) had quotes from other sources (e.g. funders) related to the research; and 179 (30%) had quotes from independent scientists or 'experts'.

Study details. We coded whether and how accurately/precisely each PR and news story reported sample size (N), completion rate, length of study, and number of time points for longitudinal studies. Of these data, we asked how often news stories provide details that were not contained in the associated PR (i.e. as evidence that the journalist has used a source additional to the PR). There were 89 news stories for which the PR did not give any indication of N where N could have been reported. Of these, only 2 stories reported N exactly, 7 reported N approximately (within 20%), and 6 reported N with more than 20% discrepancy with the actual N. Similarly, only 5% (16) news stories mentioned completion rate out of the 346 news stories where it was relevant but the PR did not provide it. Likewise, only 3% (9) of 280 news stories gave study duration where PR did not and only 6% (2/36) gave the number of time points in longitudinal studies where the PR did not.

For comparison, when PRs do provide these numbers, journalists often use them. N was reported exactly in the news in 57% of cases (89/157), and approximately in a further 13% of stories. Completion rate was reported in 53% (67/127) of news, while duration was reported in 67% (118/176). Similarly, of 81 PRs that mentioned time points for longitudinal studies, 57 news stories mentioned them as well (70%).

SI12. Scientist doublethink

Whilst instigating the main study, we performed an online survey of scientists' attitudes toward science in the media, and their experiences with PR. We advertised the survey via the *Guardian*, the BBSRC, and social media. The sample is self-selected and likely biased towards pre-existing interest in the topic of science news and by the subject area distribution in our advertising routes. As expected, the respondents (N=248) blamed journalists more than any other party for misreporting in science news. However, 79% of scientists who had PRs about their work reported involvement with those PRs, and despite this involvement, 32% acknowledged that their PRs were exaggerated (Figure S2). Thus it appears that some scientists do have awareness that PRs are a source of misreporting, but as a group we appear to engage in doublethink - colluding in producing exaggerated PRs but mainly blaming the media for the shortcomings of science news.

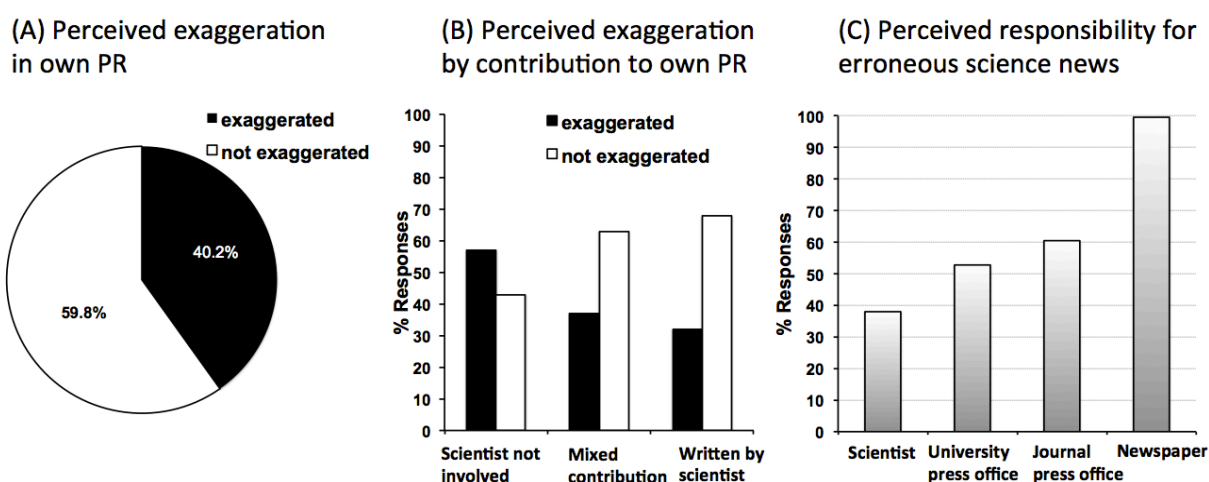


Figure S2. We performed an online survey in March 2012 to gather a sample of scientist's experiences and opinions on PR and science news (N=248). The key results were that 40% (N=43) of respondents with experience of PR (N=107) perceived that their most recent PR was exaggerated (A). Unsurprisingly, this proportion decreased with greater levels of declared involvement in the preparation of PRs, but still remained above 30% even for those scientists who reportedly wrote the PR themselves (B). When asked who was responsible for erroneous science news (C), 30-60% attributed some responsibility to scientists and press offices; this may reflect awareness of some PR exaggeration. However, 100% of respondents attributed responsibility to newspapers. The survey and accompanying data can be downloaded from <http://dx.doi.org/10.6084/m9.figshare.903704>. We advertised the survey via the *Guardian*, the BBSRC, and social media. The sample is self-selected (likely due to pre-existing interest in the topic of science news) and possibly biased by the subject area distribution in our advertising routes.

SI13. Comparison between news outlets and journalist type

Table S5 shows the rates of inflation from PR to news and from JA to news for different outlets. As in Table S4, advice, causal claims and human inference are combined to form one category of 'statements', which is why N statements differs from N PR. Note that the percent of statements *without* inflation is given (rather than *with* inflation). Some news outlets had too few N to be included individually: The Mail on Sunday (N=1) and Mail Online (N=2) have been combined with The Daily Mail (N=89), The Sunday Sun (N=1) has been combined with The Sun (N=44), The Sunday Telegraph (N=1) has been combined with The Telegraph (N=80), The Sunday Times (N=1) has been combined with The Times (N=31); The Daily Star, The Economist, the New Scientist, and the Press Association (each N<6) have been excluded. Note that the reason N statements differ slightly between the comparisons to PR and to JA is because in some cases comparison could not be made, for example if the PR does not say anything upon which to base a code for animal vs human, then human/animal claims in news could not be compared to PR.

The table is ordered by % claims without inflation from PR to news, though we do not make any conclusions from this order, given the very wide confidence intervals on the ranks (calculated as for Table S4). The more appropriate conclusion appears to be that news outlets do not differ from each other as much as might be generally assumed.

We also coded whether the journalist for each news story was a generalist or health/science specialist. Counter to expectation, we detected no differences between these categories for inflation rates. For advice, there were 23 inflations from PR in 182 news stories for specialists (13%) compared to 21/179 (12%) for generalists (difference =0.9%, with 95% CI of -5.8% to 7.6%). For causal claims from correlational results, there were 71/201 (35%) for specialists vs 83/244 (34%) for generalists (difference =1.3%, with 95% CI of -7.7% to 10.1%). For human inference from non-human studies, there were 5/57 (9%) for specialists vs 12/95 (13%) for generalists (difference =-3.9%, with 95% CI of -13.7% to 6.7%). It may be noteworthy that specialists wrote about non-human studies less frequently than did generalists, which may indicate differing knowledge about the difficulties of translating animal results into treatments for humans.

Table S5. Rates of inflation for news outlets in our study. Inflation is listed relative to PRs and relative to JAs (see text above for further explanation).

	N stories	N statements	N inflated PR to news	% without inflation	Rank (95% CI)	N statements JA to news	N inflated to news	% without inflation	Rank (95% CI)
The Mirror	50	47	5	89%	1 (1-7)	52	14	73%	1.5 (1-8)
The Independent	40	55	8	85%	2 (1-9)	53	21	60%	8 (2-12)
The Scotsman	23	32	5	84%	3 (1-11)	32	11	66%	5 (1-12)
BBC news online	95	120	20	83%	4.5 (1-9)	120	46	62%	7 (3-11)
The Express	57	66	11	83%	4.5 (1-10)	67	32	52%	12 (6-12)
The Times	32	35	6	83%	6 (1-11)	36	12	67%	4 (1-11)
Reuters	49	68	13	81%	7 (2-11)	66	21	68%	3 (1-9)
Metro	49	56	11	80%	8 (1-11.5)	52	14	73%	1.5 (1-7)
The Telegraph	81	110	28	75%	9 (5-12)	112	39	65%	6 (2-10)
The Guardian	44	65	17	74%	10 (4-12)	67	28	58%	9 (3-12)
The Sun	45	48	13	73%	11 (4-12)	48	22	54%	10 (4-12)
The Daily Mail	92	118	34	71%	12 (7-12)	118	55	53%	11 (6-12)