# Generic Language in the Communication of Health Research

Harry T. Clelland, PhD [1,2 *]

https://orcid.org/0000-0001-7525-007X

Matthew Haigh, PhD [1]

https://orcid.org/0000-0003-1549-697X

*Corresponding author

[1] Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom.

[2] Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

Email: harry.t.clelland@northumbria.ac.uk

Word count (excluding Abstract, Tables and Figures): 4177

Pages (excluding Title Page, Abstract, References): 18

Tables: 2

Figures: 1

## Abstract

Generic claims imply universal rules about categories or groups by glossing over individual variability and exceptions. Consider, for example, the generic news headline ‘Exercise helps teens quit smoking’ relative to the non-generic ‘Exercise helps *some* teens quit smoking’. When generic language is used in primary reporting of scientific research (i.e., journal articles) the reported claims are perceived to be more important and more generalisable than when non-generic language is used. This study aimed to establish whether the same effects would be present within secondary reporting of research (i.e., news articles aimed at the public), with a particular focus on personal health claims. Participants read a series of genuine news headlines in either their original generic format (e.g., ‘Exercise helps teens...’) a past-tense non-generic format (e.g., ‘Exercise help*ed* teens...’) or a qualified non-generic format (e.g., ‘Exercise helps *some* teens...’). Generic headlines were rated as slightly more important and more generalisable than headlines qualified with the word ‘*some*’. In contrast, we found no differences in perceived importance or generalisability between generic headlines and past-tense non-generic headlines. Our results suggest that writers must be explicitly non-universal when summarising research in order to accurately communicate constraints on generality.



**Key words:** Generics; Language; Inference; Pragmatics; Health Communication

**Generic Language in the Communication of Health Research**

Generic generalisations (aka. generics) make category wide claims (Cimpian et al., 2010). Take for example statement (1):

(1) Women give vaginal birth

The generic statement (1) expresses a characteristic (giving vaginal birth) on the basis of a group categorisation (women) and few people would argue the truthfulness of the statement. A key characteristic of generics is that they have a tolerance to exceptions (Krifka et al., 1995). In other words, they are difficult to prove false. For example, consider generic (1) in comparison to its universal equivalent (2):

(2) *All* women give vaginal birth

The generic statement (1) can be asserted truthfully even in the face of obvious exceptions, such as the common knowledge that some women do not give birth and that some women give birth via cesarean section. By contrast, the non-generic universal statement (2) cannot be true in the face of exceptions, as it can be falsified with a single counterexample. A single woman who gives birth by cesarean section is sufficient to conclude that 'not all women give vaginal birth' (Lazaridou-Chatzigoga, 2019).

Another notable characteristic of generic statements is that they require minimal evidence to be accepted as true. Across four experiments, Cimpian et al. (2010) compared interpretations of generic statements about a fictional creature (e.g., "Lorches have purple feathers") to a quantified version of that statement, modified with the term 'most' (e.g., "Most Lorches have purple feathers"). Even when a property was uncommon (i.e., less than 50% of Lorches have purple feathers) participants often judged the generic statement to be true. Moreover, when asked to judge proportion based on the generic statement, it was inferred that the property occurs in

over 90% of category member cases (i.e., over 90% of Lorches have purple feathers). In comparison, no such finding was present for the quantified statement using '*most*'. That is, for the quantified statements there was no such asymmetry between truth judgements and implied proportion prevalence. In short, generics have three notable characteristics: they require minimal evidence to be accepted as true, they are difficult to prove false and they may imply universal rules on the basis of not-so-universal evidence.

One area of concern with regards to generalisability relates to how scientific findings are communicated (e.g., journal articles, university press releases, social media posts). The findings of research with humans, which is always based on a limited sample, tends to be generically generalised to 'humans' or 'people', despite Henrich et al. (2010) highlighting an over-reliance on 'WEIRD' samples (Western, Educated, Industrialised, Rich, Democratic) that are not necessarily representative of the global population.

There is growing evidence suggesting that generic statements can be used to unwittingly imply that research findings are both more important and more generalisable than what can truly be inferred from the data. DeJesus et al. (2019) sampled research papers and found that almost 90% summarised their results with a generic statement in either the title, highlights or abstract (e.g., '*group discussion improves lie detection*') (see also DeJesus et al., 2023 for similar findings). They then experimentally manipulated journal article titles that contained generics to test how changes would be interpreted by readers. Specifically, bare generic headlines were altered to create non-generic versions for comparison (e.g., group discussion improv*ed* lie detection; *Some* group discussion improv*ed* lie detection). Overall, the results of multiple experiments revealed that generic statements were perceived as more important than their non-

generic counterparts. Moreover, generic statements were perceived to be more generalisable and more conclusive than non-generics, but only under particular circumstances.

DeJesus et al. (2019) highlight a seemingly discipline-wide tendency to over-generalise psychological research findings by making generic claims in article titles and abstracts. Their work was limited to primary source journal articles, written for research professionals. In the study below we aim to extend the findings of DeJesus et al. by investigating how generics are interpreted when used by the media to report research to the public. We focus specifically on the communication of health-related research as health-related research findings are frequently overgeneralised by academics (Peters et al., 2024), are frequently reported by the media (Bossema et al., 2019; Schat et al., 2018) and may inform consequential health-related behaviours by members of the public.

**The Present Research**

The aim of the current study was to extend the findings of DeJesus et al. (2019) and test whether generic statements would be interpreted as more important and more generalisable than non-generic statements when reported through a secondary source (i.e., the headline of a news article). The reason for our decision is two-fold.

First, this approach has wider applicability because a lay person is more likely to read news headlines that present short summaries of research than they are to read a primary source. In this study, headlines were rated by the same audience that will read them in the real world (i.e., the public), which was not the case in DeJesus et al. (2019), as non-academic participants evaluated the importance and generalisability of research papers written for an academic

audience[1]. This distinction has important implications because academics (via their research methods training) should have an implicit understanding of the limitations of a sample (i.e., they should be aware that a generic claim does not apply across-the-board). This is not necessarily true of the public however, meaning that overgeneralisations caused by generics may have greater consequences in this context.

Second, generic claims may be more widespread in the media, as journalists are more motivated to utilise them. Generics are inherently short, simple and easy to understand, all of which are desirable properties for news article writers. While academics should fulfil their obligation to rigour and factual accuracy, secondary reporters are tasked primarily with translating the findings to a wider audience in an appealing manner. For this reason, generics may be more widespread in secondary reports because journalists are more motivated by appeal (and less motivated by precision) than the original authors (Haigh, Birch & Clelland, 2024).

We chose to focus specifically on reports of health research because these reports could inform consequential health behaviours made by the public. If research findings presented in a generic manner are perceived as more important and more generalisable, then individuals may be more likely to adopt associated lifestyle changes (for better or worse). There is therefore a potential danger that generic summaries of nuanced health research will cause some people to make health decisions based on the ill-informed assumption that research findings apply across-the-board.

The experiment reported below examines whether the use of generic phrases in the secondary reporting of health research increases the perceived importance and generalisability of

[1] Note that one sample reported in DeJesus et al. (2019) was introductory psychology students (i.e., not 'non-academics'), however, it remains true that this population are not the primary target audience of research papers.

findings. We selected 18 genuine generic news headlines (e.g., '*Online therapy helps teens recover from chronic fatigue syndrome*'). For each headline we created a non-generic past-tense version (e.g., '*Online therapy helped teens….*') and a non-generic version qualified with 'some' (e.g., '*Online therapy helps some teens….*'). Participants rated each headline on importance and then rated generalisability in three ways; implied sample diversity, implied sample size and implied efficacy.

We expected that the effects observed by DeJesus et al. (2019) relating to summaries of primary data would carry over to secondary reporting of data. We hypothesised therefore that bare generic headlines (i.e., X helps Y) would be perceived as more important and more generalisable than both past-tense non-generic headlines (i.e., X helped Y) and qualified non-generic headlines (i.e., X helped some Y).

## Method

This study was preregistered prior to data collection (osf.io/wgz7d/). Materials, raw data and the analysis script can be found on the Open Science Framework (OSF; see osf.io/krpab/).

### Design

This experiment used a one-factor counterbalanced (Latin Square) repeated measures design. The independent variable (fixed factor) was the Headline Format with three levels (Bare Generic, Past-Tense Non-Generic, Qualified Non-Generic). Participants were randomly assigned to one of three Latin Square presentation lists, each of which contained 18 health research headlines presented in a different random order to each participant (see osf.io/6jgb7/ for Latin Square design). Participants were exposed to all three conditions (six bare generic headlines, six

past-tense, six qualified), but saw only one version of each headline. After reading each headline, they answered four questions (described below).

**Participants**

We conducted a prospective power analysis for a one factor counterbalanced repeated measures design (one fixed effect, two random effects) with 18 items. To detect a 'medium' sized effect 80% of the time (d = 0.5, $\alpha$ = 0.05) we required a minimum of 69 participants (Westfall et al., 2014). This was our minimum sample size target. An initial sample of 102 consenting participants were recruited via Prolific, an online recruitment platform. One participant was excluded for non-completion. All remaining participants declared that they took the survey seriously. The final sample was 101 English-speaking adults (38 male, 63 female), aged between 18 and 77 (mean age = 37.37, SD = 14.08). Participants were paid £1.

**Materials**

We selected 18 online news headlines from the Google News archive that summarised research findings using a generic statement. Headlines were selected on two key characteristics. First, the headline must make a generic statement. Second, the headline must summarise primary research that reports improvements in an aspect of personal health (e.g., '*online therapy helps teens recover from chronic fatigue syndrome*').

The headlines varied in content, but each can be reduced to a generic 'X helps Y' format, whereby 'X' is a treatment or intervention that 'helps' (e.g., 'improves', 'boosts', 'beneficial to') a specific group of people 'Y' (e.g., 'patients', 'men', 'children'). One headline followed the inverse of this format by starting with the group (e.g., 'stroke patients [Y] recover arm use with

virtual reality [X]') but served the same functional purpose as an 'X helps Y' statement. The headlines related to six categories of people (patients, children, women, men, elderly, teenagers). Full details of all 18 items (as well as their past-tense and qualified versions) can be found on the OSF (see osf.io/3mav5/).

### *Headline Format Manipulation*

Each bare generic headline was manipulated to create two distinct non-generic versions. One version was edited to become past tense (e.g., 'helps' edited to be 'helped'), while the other version was edited to include a qualifier (i.e., 'men' edited to become 'some men'). Examples of our manipulations are presented in Table 1.

**Table 1**

*Three Examples of Bare Generic, Past-Tense and Qualified Headlines used in the Experiment.*

| | |
|---|---|
| ***a) "Stroke Patients Recover Arm Use With Virtual Reality"*** | |
| *Bare Generic* | *Stroke Patients Recover Arm Use With Virtual Reality* |
| *Past-Tense* | *Stroke Patients **Recovered** Arm Use With Virtual Reality* |
| *Qualified* | ***Some** Stroke Patients Recover Arm Use With Virtual Reality* |
| ***b) "Plasma Therapy Helps Critically Ill Covid-19 Patients"*** | |
| *Bare Generic* | *Plasma Therapy Helps Critically Ill Covid-19 Patients* |
| *Past-Tense* | *Plasma Therapy **Helped** Critically Ill Covid-19 Patients* |
| *Qualified* | *Plasma Therapy Helps **Some** Critically Ill Covid-19 Patients* |
| ***c) "Testosterone Therapy Beneficial to Men with Heart Disease"*** | |
| *Bare Generic* | *Testosterone Therapy Beneficial to Men with Heart Disease* |
| *Past-Tense* | *Testosterone Therapy **was Beneficial** to Men with Heart Disease* |
| *Qualified* | *Testosterone Therapy Beneficial to **Some** Men with Heart Disease* |

### *Measures*

Measures were based on those used by DeJesus et al. (2019) study 2. We constructed our questions with the intention of replicating fundamental elements of their wording, whilst making each question specific to each individual headline. For example, when measuring efficacy DeJesus et al. broadly asked '*what percentage of those in the world today would show the effect described*'. By contrast, our questions were specifically tailored to each headline (e.g., 'what percentage of *stroke patients* would *recover arm use with virtual reality*').

#### Importance (DV1)

Subjective ratings of importance were established by asking participants how important they thought the research findings were. Importance was rated on a scale ranging from 0 (Not at all Important) to 7 (Extremely Important).

#### Diversity (DV2)

Subjective ratings of diversity were established by asking participants to estimate the likelihood that the research findings would extend to those from diverse backgrounds. Diversity was rated on a 0 (Not at all Likely) to 7 (Extremely Likely) scale.

#### Sample Size (DV3)

Subjective ratings of sample size were established by asking participants to guess how many people took part in the research. Estimated sample size was chosen from 1 of 7 options (1-10 people, 11-50 people, 51-100 people, 101-250 people, 251-500 people, 501-1000 people, 1001+ people). These are the same options used by DeJesus et al. (2019).

### Efficacy (DV4)

Subjective ratings of efficacy[2] were established by asking participants to estimate the percentage of those described in the headline (e.g., stroke patients) that would show the effect described (e.g., recover arm use with virtual reality). Efficacy was rated as from 0 to 100%.

## Procedure

This online study was run online using Qualtrics. Participants accessed the study anonymously via the Prolific recruitment platform. They were then randomly allocated to one of three item lists. Once allocated, a series of 18 headlines were shown in a different random order to each participant. For each, they indicated their interpretations of importance, diversity, sample size and efficacy. Afterwards, participants were asked whether they took the survey seriously (Aust et al., 2013). Median completion time for the entire survey was 8.1 minutes. A .qsf file for the survey, which allows for direct replication, can be accessed via the OSF (see osf.io/azjmp/).

## Statistical Analysis

All questions required a response, so there were no missing data. No data points were excluded. Analyses were conducted using R version 4.0.2 and RStudio version 1.3.959. Each dependent variable was analysed using a linear mixed effects model, fitted via the 'lme4' package (Bates et al., 2015). The fixed factor was Headline Format which had three levels (bare generic vs. past-tense vs. qualified). We obtained *p* values for our fixed effects using 'lmerTest' (Kuznetsova et al., 2017), which uses Satterthwaite's method for approximating degrees of

[2] This DV is named 'generalisability' in our pre-registration, but we favoured 'efficacy' here as it is more reflective of the question that was asked.

freedom. Participants and items were entered into the model as crossed random factors. Following the recommendations of Barr et al. (2013), each model began with a maximal random effects structure (i.e., random intercepts and slopes for participants and items). If the model did not converge, we systematically reduced the random effects structure until the model did converge. The first reduction method was removing the random correlations. The next reduction method was running a random slopes only model, followed by a random intercepts only model if still unsuccessful. For all of our measures, the most maximal converging model was a random intercepts only model (i.e., maximal model with random slopes removed). Post-hoc tests were conducted using the 'emmeans' package (Lenth, 2022) using Satterthwaite's method for approximating degrees of freedom. Bonferroni correction (adjusted $\alpha$ = .0167) was used for each model when comparing condition means.

## Results

Condition means for all four DVs are displayed in Figure 1. For each type of headline (Bare Generic, Past-Tense, Qualified), the figure plots average ratings of importance (i.e., how important the summarised finding is perceived to be; upper left quadrant), estimations of how the summarised finding may extend to a diverse background (upper right quadrant), estimations of the sample size from which the summarised finding was derived (lower left quadrant), and estimations of the summarised finding's applicability to the wider population (lower right quadrant). Inferential statistics from the linear mixed models are presented in Table 2.

**Figure 1**

*Density plots showing mean ratings by condition. Surrounding bands represent 95% CI.*

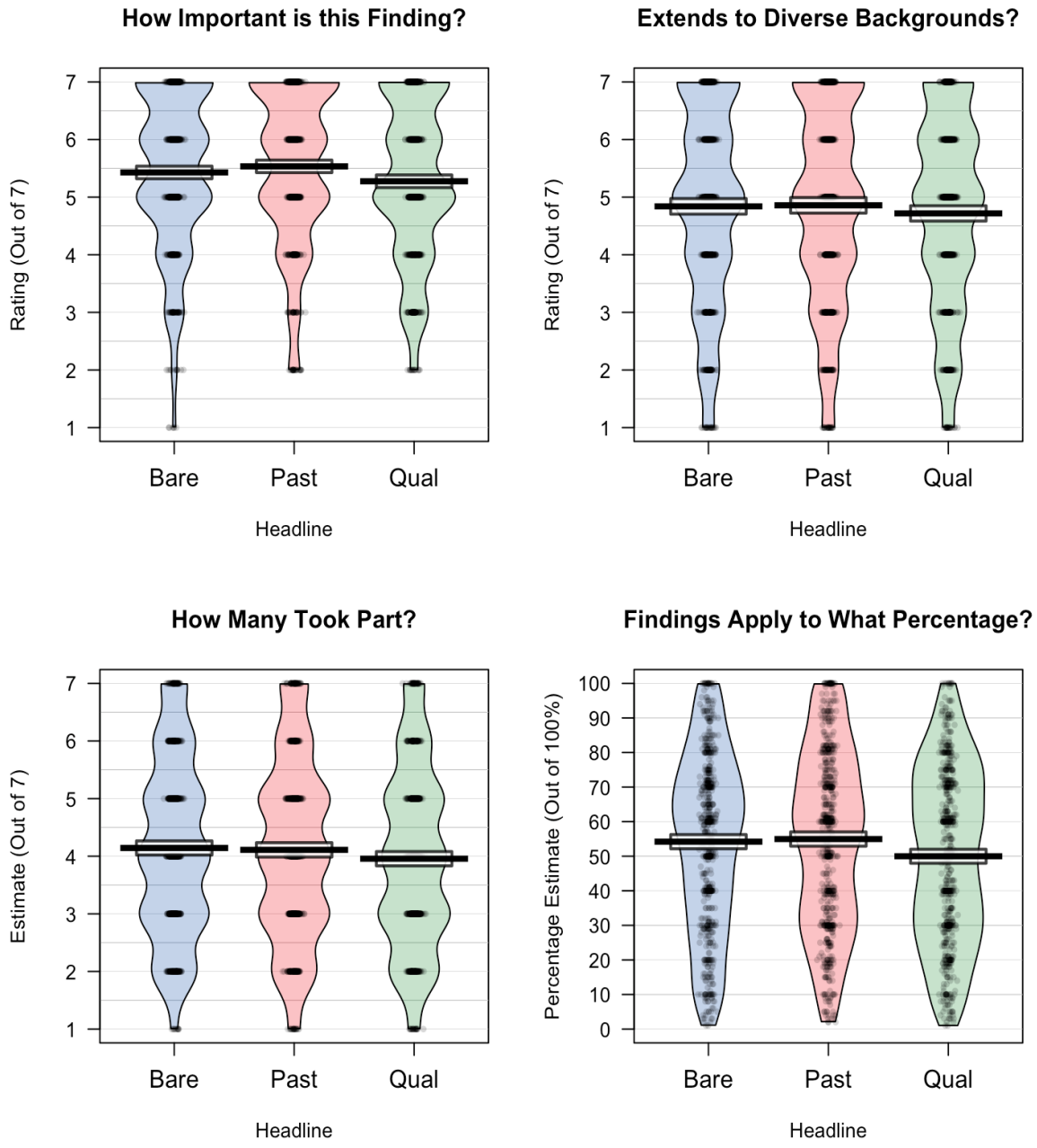

**Table 2**

*Parameter estimates for each dependent variable.*

| Condition | *b* | *SE* | *t* | *p* |
|---|---|---|---|---|
| **Importance** | | | | |
| Bare (Intercept) | 5.430 | 0.150 | 36.302 | <.001 |
| Past-Tense | 0.105 | 0.060 | 1.750 | .080 |
| Qualified | -0.157 | 0.060 | -2.618 | .009 |
| **Diversity** | | | | |
| Bare (Intercept) | 4.841 | 0.149 | 32.518 | <.001 |
| Past-Tense | 0.017 | 0.067 | 0.250 | .802 |
| Qualified | -0.124 | 0.067 | -1.855 | .064 |
| **Sample Size** | | | | |
| Bare (Intercept) | 4.145 | 0.152 | 27.350 | <.001 |
| Past-Tense | -0.035 | 0.061 | -0.569 | .569 |
| Qualified | -0.188 | 0.061 | -3.092 | .002 |
| **Efficacy** | | | | |
| Bare (Intercept) | 54.225 | 2.446 | 22.172 | < .001 |
| Past-Tense | 0.738 | 1.089 | 0.677 | .498 |
| Qualified | -4.243 | 1.089 | -3.895 | < .001 |

## Importance

Headline format significantly predicted importance rating (see Table 2). Post-hoc comparisons revealed that Qualified headlines (mean = 5.27) were rated as significantly less important than Bare generic (mean = 5.43; $t(1700) = 2.618$, $p = .009$) and Past-Tense (mean = 5.54; $t(1700) = 4.368$, $p < .001$) headlines. There was no difference between the Bare and Past-Tense conditions ($t(1700) = -1.750$, $p = .080$).

**Diversity**

Headline format significantly predicted diversity rating (see Table 2). However after making the Bonferroni correction to $\alpha$ (using $\alpha = .0167$), post-hoc tests revealed no significant differences between Bare (mean = 4.84) and Past-Tense (mean = 4.86; $t(1700) = -0.250$, $p = .802$), Bare and Qualified (mean = 4.72; $t(1700) = 1.855$, $p = .064$) and Past-Tense and Qualified conditions ($t(1700) = 2.106$, $p = .035$).

**Sample Size**

Headline format significantly predicted sample size estimates (see Table 2). Post-hoc comparisons revealed that Qualified headlines (mean = 3.96) were estimated to have significantly less participants in the sample than Bare (mean = 4.14; $t(1700) = 3.092$, $p = .002$) and Past-Tense (mean = 4.11; $t(1700) = 2.523$, $p = .012$) headlines. There was no difference between the Bare and Past-Tense conditions ($t(1700) = 0.569$, $p = .569$).

**Efficacy**

Headline format significantly predicted estimates of efficacy (see Table 2). Post-hoc comparisons revealed that Qualified headlines (mean = 50.0) implied that the results were applicable to a significantly lower proportion of people than Bare (mean = 54.2; $t(1700) = 3.895$, $p < .001$) and Past-Tense (mean = 55.0; $t(1700) = 4.573$, $p < .001$) headlines. There was no difference between the Bare and Past-Tense conditions ($t(1700) = -0.677$, $p = .498$).

## Discussion

The key findings of this experiment are two-fold. First, bare generic headlines (e.g., '*exercise helps men regain bone mass*') were perceived as more important and generalisable than qualified non-generic headlines (e.g., '*exercise helps some men regain bone mass'*). Second, bare generic headlines were perceived as no more important and no more generalisable than a past-tense non-generic (e.g., '*exercise helped men regain bone mass'*). In other words, 'X helps Y' was perceived no differently to 'X helped Y'.

As predicted, headlines including generic statements were considered to summarise findings that were more important than the same headlines qualified with the word 'some'. In other words, 'X helps Y' was considered to be more important than 'X helps *some* Y'. The same pattern of results was also found for estimates of sample size and ratings of percentage efficacy. Generic headline summaries of research findings were judged to have been concluded from a larger sample of participants than the qualified version (i.e., it was perceived that a larger sample was used to conclude 'X helps Y' than the sample used to conclude 'X helps *some* Y'). Finally, the same pattern of results was found for estimates of efficacy. The benefit described in a generic headline was perceived as applicable to greater proportion of relevant people than the same benefit described in a qualified headline (i.e., 'X helps Y in men' was considered to be applicable to a greater proportion of men than 'X helps Y in *some* men').

Taken together, these findings largely align with our understanding of generics to date, in that they can exaggerate claims and imply universal rules (Cimpian et al., 2010; Gelman & Roberts, 2017; Leslie, 2007; Leslie, 2008). Importantly, these findings develop our understanding by evidencing that the effects of generic language that have been previously observed in primary reporting (i.e., journal articles), are also observed in secondary reports

designed for public consumption. This may be consequential because implying that a research finding is more important and more generalisable than it really is could directly impact the health beliefs and behaviours of the reader.

An increased demand for shorter, more concise distributions formats encourages authors of journal articles to make enticing claims about their findings (Weinstein & Sumeracki, 2017), and the most exaggerated (or enticing) claims are favoured by the news media (Sumner et al., 2014). Whilst DeJesus and colleagues (2019) effectively evidence a widespread tendency for authors to use generics in the shortest sections of articles (titles, highlights and abstracts), and also demonstrate the over-importance and over-generalisability achieved by using generics; the present data (using genuine news articles titled with generics) indicate that the same effects carry over into interpretations of secondary reports (i.e., media headlines). This may be particularly concerning, because arguably, the dangers associated with generics are more pronounced within this context. Media headlines are predominantly read by the general public, who by-and-large are not trained in research methods like academics are. There may therefore be greater potential for generic claims to be understood as universals by lay readers, leading to the conclusion that a statement applies to *all* members of a category (e.g., all men) (e.g., Haigh et al., 2020; Haigh et al., 2024; Leslie et al., 2011).

Perhaps our most notable finding was that there was no difference between the generic headlines and the past-tense non-generic headlines on ratings of importance or generalisability. In other words, across all of our measures, generic statements (i.e., X helps Y) produced the same ratings of importance and generalisability as the past-tense non-generic version (i.e., X helped Y). A potential explanation for this could be that both statements make a generalised (i.e., non-specific) claim about Y and are therefore perceived in the same way. By contrast, the

qualified statement ‘X helps some Y’ makes an explicitly non-universal claim about Y that is limited to a certain number of individuals, hence is interpreted differently. Interestingly, the pattern of results described here are mirrored in some (but not all) of the studies reported in DeJesus et al. (2019). For example, their studies 3a to 3d consistently show no differences in perceived importance between bare generic and past-tense non-generic summaries, but the original effect returns in study 4a when participants are given a direct comparison (i.e., they are asked to choose the most important from two versions of the same summary). Taken with the current findings, it is plausible that any interpretive differences observed between bare generics and past-tense non-generics are detectable only when the two formats are presented in close proximity (i.e., easily contrasted).

Nonetheless, this unexpected finding may have significant implications, as it suggests the requirement for writers to be explicitly non-universal when looking to accurately convey limits on generality. DeJesus et al. (2019) found that their participants were more sensitive to language when journal article titles had multiple cues to non-generality. For example, a title that is both qualified *and* past-tense (i.e., X *helped* Y, *under certain circumstances*). They concluded that in order to accurately communicate constraints on generality (i.e., acknowledging the limitations of the sample from which results are drawn from), writers may be required to take an explicit linguistic approach, rather than a reliance on subtle signals. Our results support this interpretation. Despite the fact that a generic statement phrased in the past-tense no longer fits the definition of a ‘generic’, the evidence here implies that there are no meaningful differences in how they are interpreted. This aligns with the notion that a writer cannot successfully acknowledge sampling limitations by taking a non-explicit approach to language (i.e., only slightly altering the generic statement format). Rather, substantial linguistic changes to the

generic formula may be required to avoid unintentionally misleading readers. We contend that replacing a generic (X helps Y) with a past-tense (X helped Y) is not substantial enough to accurately consider the limitations of a sample. Writers must instead be explicitly non-universal to achieve this (e.g., by explicitly qualifying their statement).

## Implications for Health Behavior Research

This and other related studies (e.g., DeJesus et al., 2019, 2023) have implications for academic research into health behaviours, media reporting of health research and the public understanding of science. Academic researchers should reflect on their own use of generic language. Generics come across as appealing and conclusive, which may help research to be published and disseminated. However, this appeal can come at the expense of accuracy. We urge academic writers to identify generics in drafts of their work (i.e., manuscripts, press releases, social media posts) and consider whether a non-generic phrasing might be a more accurate summary of the results. Health interventions rarely help 100% of participants in a trial, even if there is a significant effect on average, so consider whether the intervention truly helps a whole group (e.g., men, women, people) or whether it is more accurate to say that it helps some or most of that group (e.g., some men). This approach would be complemented by the inclusion of a 'Constraints on Generality' statement, as recommended by Simons et al. (2017). A reduction in the use of generic claims by researchers may feed through to a reduction in their use by the media, as articles written by journalists and popular science writers tend to closely follow press releases (Sumner et al. 2014), but we also urge media and social media content creators to reflect on their own use of generic language. Finally, public understanding of health research could be

improved by promoting the message that generic claims are not necessarily based on larger or more diverse samples than non-generic claims (e.g., Peters et al., 2024).

The next step is to examine whether the effects reported above differ for reports of negative health outcomes (e.g., X harms Y). The findings reported in this study are limited to headlines reporting health benefits, but the effects may be amplified when the reported outcome is harmful (see Baumeister et al., 2001). The second and potentially most impactful line of research should focus on whether the effects of generic language on perceived importance extend to consequential health decisions and behaviours. The current study is limited by a focus on interpretation, which fails to shed light on whether health-related decision making can be altered by generics. For example, are people more likely to accept a medication or make a lifestyle change if the efficacy of that intervention is described using generic language?

## Conclusions

Generic language is common in the reporting of health research to the public. The choice to make a general claim may lead to an increased perception of importance and generalisability in the mind of a reader, so academics and journalists should be mindful as to whether the evidence is truly strong enough to warrant an unqualified claim.

## Discussion Question

Generic claims published by the media often derive from generic claims made by researchers (e.g., in papers or press releases). Reflect on your own use of generics and discuss the potential impacts (both positive and negative) of avoiding these generalisations.

## Ethical approval

The project was approved through the Northumbria University Ethical Approval System (REF 26549).

## Declaration of Conflicting Interests

This work was partially supported by a Leverhulme Trust Research Project Grant (RPG-2019-158) awarded to the second author.

## References


Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, 45(2), 527-535. https://doi.org/10.3758/s13428-012-0265-2

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370. https://doi.org/10.1037/1089-2680.5.4.323

Bossema, F. G., Burger, P., Bratton, L., Challenger, A., Adams, R. C., Sumner, P., Schat, J., Numans, M. E., & Smeets, I. (2019). Expert quotes and exaggeration in health news: a retrospective quantitative content analysis. *Wellcome Open Research*, 4, 56. https://doi.org/10.12688/wellcomeopenres.15147.2

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, 34(8), 1452-1482. https://doi.org/10.1111/j.1551-6709.2010.01126.x

DeJesus, J. M., Callanan, M. A., Solis, G., & Gelman, S. A. (2019). Generic language in scientific communication. *Proceedings of the National Academy of Sciences*, 116(37), 18370-18377. https://doi.org/10.1073/pnas.1817706116

DeJesus, J. M., Callanan, M. A., Umscheid, V. A., & Gelman, S. A. (2023). Generic Language and Reporting Practices in Developmental Journals: Implications for Facilitating a More Representative Cognitive Developmental Science. *Journal of Cognition and Development*, 1-23. https://doi.org/10.1080/15248372.2023.2290504

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, 114(30), 7900-7907. https://doi.org/10.1073/pnas.1621073114

Haigh, M., Birch, H. A., & Clelland, H. T. (2024). "Experts think…" The production and comprehension of propositional attitude generics. *Thinking & Reasoning*, 1-26. https://doi.org/10.1080/13546783.2024.2370924

Haigh, M., Birch, H. A., & Pollet, T. V. (2020). Does 'Scientists Believe…'Imply 'All Scientists believe…'? Individual Differences in the Interpretation of Generic News Headlines. *Collabra: Psychology*, 6(1). https://doi.org/10.1525/collabra.17174

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Krifka, M., Pelletier, F. J., Carlson, G., Ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: an introduction. In G.N. Carlson & F.J. Pelletier (Eds.), *The Generic Book* (pp. 1-124). University of Chicago Press.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*. 82(13), 1-26: https://doi.org/10.18637/jss.v082.i13

Lazaridou-Chatzigoga, D. (2019). Genericity. In C. Cummins & N. Katsos (Eds.). *The Oxford Handbook of Experimental Semantics and Pragmatics* (pp. 156-177). https://doi.org/10.1093/oxfordhb/9780198791768.013.12

Lenth R (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.5, https://CRAN.R-project.org/package=emmeans

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, 21, 375-403.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, 117(1), 1-47. https://doi.org/10.1215/00318108-2007-023

Leslie, S.-J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language*, 65(1), 15-31. https://doi.org/10.1016/j.jml.2010.12.005

Peters, U., Sherling, H. R., & Chin-Yee, B. (2024). Hasty generalizations and generics in medical research: A systematic review. *Plos one*, 19(7), e0306749. https://doi.org/10.1371/journal.pone.0306749

Schat, J., Bossema, F. G., Numans, M. E., Smeets, I., & Burger, P. (2018). Exaggerated health news: association between exaggeration in university press releases and exaggeration in news media coverage. Nederlands Tijdschrift Voor Geneeskunde, 162(1).

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123-1128. https://doi.org/10.1177/1745691617708630

Sumner P., Vivian-Griffiths S., Boivin J., Williams A., Venetis C. A., Davies A. Ogden, J., Whelan, L., Hughes, B., Daton, B., Boy, F., & Chambers C. D. (2014). The association

between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ*, 349. https://doi.org/10.1136/bmj.g7015

Weinstein, Y., & Sumeracki, M. A. (2017). Are twitter and blogs important tools for the modern psychological scientist? *Perspectives on Psychological Science*, 12(6), 1171-1175. https://doi.org/10.1177/1745691617712266

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of *stimuli. Journal of Experimental Psychology: General*, 143(5), 2020. https://doi.org/10.1037/xge0000014