

Message-level Claims Require Message-level Data Analyses: Aligning Claims and Evidence in Communication Research

Daniel J. O'Keefe

To cite this article: Daniel J. O'Keefe (04 May 2024): Message-level Claims Require Message-level Data Analyses: Aligning Claims and Evidence in Communication Research, Communication Methods and Measures, DOI: [10.1080/19312458.2024.2349987](https://doi.org/10.1080/19312458.2024.2349987)

To link to this article: <https://doi.org/10.1080/19312458.2024.2349987>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 04 May 2024.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

Message-level Claims Require Message-level Data Analyses: Aligning Claims and Evidence in Communication Research

Daniel J. O'Keefe 

Department of Communication Studies, Northwestern University, Evanston, IL, USA

ABSTRACT

Researchers often invoke individual-level correlations (correlations between properties of individuals) as a basis for message-level claims (claims about properties of messages). For example: “People who are more transported by a narrative message are generally also more persuaded by it; individuals’ transportation and persuasion scores are positively correlated. Therefore narratives that are more transporting will be more persuasive than narratives that are less transporting.” But that inference is mistaken. The reasoning mistake is not specific to that example, but rather is a common mistake in communication research. This article explains the reasoning mistake, identifies multiple examples of the mistake, and discusses implications of and remedies for this circumstance.

Consider the following reasoning: “Narrative transportation and persuasion are positively correlated: the more individuals are transported by a narrative (carried away by the story, immersed in the story world, etc.), the more persuaded they are. Hence when creating narrative messages meant to persuade, message designers should make the messages highly transporting, because messages that are more transporting will be more persuasive.” This reasoning sounds quite sensible, but in fact is deeply flawed. The reasoning mistake here is not specific to narrative messages or transportation or persuasion, but instead is a common and general mistake evident in several aspects of communication research.

In what follows, this article first specifies and explains the mistake in this reasoning. A second section provides multiple examples of the mistake. A concluding section discusses implications of, and remedies for, this circumstance.

The reasoning mistake

To continue the example: It seems obvious that if individuals who are more transported by a narrative are more persuaded (than individuals not so transported), then narratives that are more transporting will be more persuasive (than narratives that are not so transporting) and hence persuaders should use those narratives that are more transporting. For example, Green and Clark (2013) noted that “individuals who are more highly transported into narratives show greater attitude, belief and behavior change” (p. 477); one implication drawn was that “if a smoking scene occurs in a highly transporting movie, it should have a stronger effect on individuals’ attitudes and behaviors than a smoking scene in a less transporting film” (p. 481). Similarly, A. J. Dillard et al. (2018) found that among people who read a narrative about skin cancer, greater

CONTACT Daniel J. O'Keefe  d-okeefe@northwestern.edu  Department of Communication Studies, Northwestern University, 999 Michigan Ave Apt. 1C, Evanston, IL 60202, USA

Daniel J. O'Keefe is the Owen L. Coon Professor Emeritus in the Department of Communication Studies at Northwestern University

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

transportation was associated with stronger behavioral intentions; this led to the question of “what types of cancer narratives may lead to the greatest transportation” (p. 588), because of the inference that narratives producing greater transportation would be more persuasive than narratives that are less transporting.

But in fact such inferences are mistaken. This can straightforwardly be illustrated through a hypothetical dataset.

The mistake concretely illustrated

This hypothetical dataset arises from an experiment comparing the persuasiveness of three narratives (A, B, and C). Each message condition has four participants (so $N = 12$), with two dependent variables: a measure of transportation (e.g., Green & Brock, 2000) and a measure of persuasion (such as attitude, intention, or behavior). Participants' (transportation, persuasion) scores are as follows. For message A: (56,48), (66,58), (76,68), and (86,78). For message B: (40,40), (62,62), (72,72), and (94,94). For message C: (78,86), (68,76), (58,66), and (48,56).

The correlation between individuals' transportation scores and persuasion scores is strongly positive, .90. The rank-order of messages on transportation is A, B, C (means of 71.0, 67.0, and 63.0, respectively). The rank-order of messages on persuasion, however, is C, B, A (means of 71.0, 67.0, and 63.0, respectively). That is, the messages' ranking on transportation is the opposite of their ranking on persuasion: the message that is *most* transporting is the message that is *least* persuasive.

Obviously, then, even if *individuals* who are more transported are more persuaded, that does not necessarily show that *messages* that are more transporting will be more persuasive. As the hypothetical dataset illustrates, it is possible – even in the same dataset – for individuals' transportation and persuasion scores to be strongly positively correlated (+.90) while messages' transportation and persuasion scores are strongly negatively correlated (-1.00). And hence it is a mistake to reason that “because individuals who are more transported by a narrative are more persuaded than individuals less transported, therefore narratives that are more transporting will be more persuasive than narratives that are less transporting.”

This hypothetical dataset illustrates the possibility of divergence between (on the one hand) how transportation and persuasion are related when the unit of analysis is individuals' scores and (on the other hand) how those two variables are related when the unit of analysis is messages' scores. But notice: the *possibility* of divergence. It is an empirical question whether messages that are more transporting are also more persuasive (that is, whether messages' transportation and persuasion scores are positively related).

To be clear: The claim being advanced here is a methodological one, not a substantive one. The claim is not (e.g.) that narratives evoking greater transportation are no more persuasive than narratives evoking less transportation. The claim concerns what counts as evidence relevant to the assertion that narratives evoking greater transportation are more persuasive than narratives evoking less transportation. And the argument advanced here is that positive correlations between individuals' transportation scores and persuasion scores are not relevant evidence, because – as the hypothetical dataset illustrates – it is possible for individuals' transportation and persuasion scores to be strongly positively correlated even if narratives evoking greater transportation are *less* persuasive than narratives evoking less transportation.

So with respect to the question of whether narratives evoking greater transportation are more persuasive than narratives evoking less transportation, the relevant evidence is how *messages'* transportation and persuasion scores are related, not how *individuals'* transportation and persuasion scores are related. One cannot draw any warranted conclusions about how messages' transportation and persuasion scores are related just from seeing how individuals' transportation and persuasion scores are related. Thus it is a mistake to use individual-level transportation-persuasion correlations as a basis for conclusions about message-level transportation-persuasion correlations.

Some clarifications

The problem described abstractly

The discussion to this point has relied on the concrete example of narrative messages' transportation and persuasiveness. But it is important to see that the possibility of divergence (between individual-level correlations and message-level correlations) does not arise from the particular variables used in that example, but rather is a consequence of the mathematics of correlation coefficients – and hence can arise whenever the unit of analysis changes from a lower-level unit (such as individual person) to a higher-level unit (such as message) based on aggregating cases from the lower-level unit.

For example: A manufacturing plant has three machines that make widgets. The quality of a widget is thought to vary depending on the molecular density of the widget. So a sample of four widgets is drawn from the output of each machine, and these are assessed for molecular density and for widget quality. For machine A, the widgets have the following (density, quality) pairs of scores: (56,48), (66,58), (76,68), and (86,78); for machine B, the scores are (40,40), (62,62), (72,72), and (94,94); for machine C, the scores are (78,86), (68,76), (58,66), and (48,56). (Yes, these are the same numbers as above.) The widget-level density-quality correlation is .90 ($N = 12$). This naturally invites the conclusion that whichever machine produces widgets with the highest density is the machine that should be preferred, because it will produce the widgets with the highest level of quality. But, as above, the rank-order of machines on the density of their output (A, B, C) is the opposite of their rank-order on quality (C, B, A); the machine-level density-quality correlation is negative (-1.00), that is, the machine that produces the highest average density produces the lowest average quality. The widget-level correlation is positive, but the machine-level correlation is negative.

The point of this widget example is to illustrate the generality of the phenomenon under discussion. In the narrative transportation example, the possibility of divergence between individual-level results and message-level results does not arise from the fact that individual people are involved, or from something about the nature of narrative messages, or from properties of the transportation and persuasion variables. The possibility of divergence arises because correlation coefficients can vary when the unit of analysis changes from a lower-level unit (individual people, individual widgets) to a higher-level unit (message, machine).

Simpson's paradox

Some readers have thought that the point under discussion is an example of Simpson's paradox (Simpson, 1951; see Malinas & Bigelow, 2016). It is not.

As applied to the circumstance of interest here, Simpson's paradox involves comparing correlations computed either across the entire set of observations or within subsets of those observations. For example, higher doses of a medication might be associated with higher disease recovery rates across all participants (a positive correlation between dosage and recovery), but a different picture can emerge when subgroups are analyzed separately. The dose-recovery correlation could be negative among men, and also negative among women (see Kievit et al., 2013). That is, the correlations within subgroups (subsets of the data) can be different from the correlation based on the entire dataset – hence the apparent paradox.

Notice: Those different correlations were all computed using the same unit of analysis, namely, the individual person. That is, all those correlations were computed by looking at the relationship between individuals' dosage scores and recovery scores. Simpson's paradox arises when the individual-level correlations in subgroups are different from the individual-level correlation computed using all the data.

The phenomenon of interest here is different. The circumstance of interest does not involve finding different results when examining subsets of cases as opposed to when examining an entire set of cases. The circumstance of interest involves finding different results when examining a given set of cases – the entire set of cases – using different units of analysis. In the hypothetical examples above, all 12 pairs

of observations were included in the dataset; what differed was whether those observations were analyzed at the individual level (person, widget) or at some aggregate level (message, machine).

The atomistic and ecological fallacies

However, the circumstance of interest here *is* related to what is called the atomistic fallacy – and thus to its better-known relative, the ecological fallacy (Firebaugh, 2015).¹ The classic form of the ecological fallacy is interpreting correlations using aggregate (group) units of analysis as if they were correlations using individual units of analysis.

As a simplified example: Imagine a researcher wants to examine the relationship between individuals' education and income (whether people who have more education make more money). The researcher does not have information about individual people, but does have US state-level data about education and income – the average amount of education and the average amount of income for residents of each of the 50 states. So the researcher computes the state-level correlation between education and income ($N = 50$) and interprets it as speaking to the question of whether individuals who have more education make more money.

But, as has been recognized for some time, this reasoning is fallacious (the classic discussion is Robinson, 1950). The correlation between income and education at some aggregate level (e.g., state-level) is not necessarily indicative of the correlation between income and education at the individual level. To support claims about individual-level relationships, one wants individual-level data.²

The atomistic fallacy (also called the individualist fallacy) is the inverse of the ecological fallacy (Atomistic Fallacy, 2008; Mackenbach, 2000). Where the ecological fallacy is the mistake of drawing conclusions about lower-level units of analysis on the basis of correlations using aggregate-level units of analysis, the atomistic fallacy is the mistake of drawing conclusions about aggregate-level units of analysis on the basis of correlations using lower-level units of analysis. But the two fallacies have the same underlying basis: a failure to recognize that as the unit of analysis changes, the correlation between two variables can be quite different – and hence correlations using one unit of analysis are not a dependable basis for conclusions about correlations using a different unit of analysis.

The reasoning mistake of interest here thus is an example of the atomistic fallacy. The specific mistake of interest is that of drawing conclusions about messages (an aggregate-level unit of analysis) on the basis of correlations using individual people (a lower-level unit of analysis).

Summary

To sum up to this point: (1) The correlation between two variables can be affected by the unit of analysis. When data are aggregated into higher-level units, the correlation can change – not only in magnitude but even in direction. Thus it is a mistake to use correlations based on lower-level units of analysis to support claims about higher-level units of analysis. (2) This is a property arising from the mathematics of correlation coefficients. That is, it doesn't matter what the substantive variables are (persuasion, widget quality, income, etc.), and it doesn't matter what the units are (people, messages, US states, etc.). Whenever data are aggregated, the two correlations can diverge. (3) The reasoning mistake does not reflect Simpson's paradox, but is an example of the atomistic fallacy. (4) The fact that the correlations *can* diverge when data are aggregated does not mean that the correlations *do* diverge. Whether divergence actually happens is an empirical question – but divergence is always possible and hence one cannot underwrite conclusions about higher-level correlations on the basis of lower-level correlations. For communication research, then, the relevant point is: It's a mistake to draw conclusions about message-level correlations on the basis of individual-level correlations.

¹At this writing, Wikipedia has an entry for "ecological fallacy" but not for "atomistic fallacy."

²The ecological fallacy is not specific to cases in which individual people are one of the units of analysis. The point applies generally to lower-level and higher-level (aggregated) data, no matter what the specific units of analysis are. So, for example, researchers interested in the relationship of educational spending and student achievement might find different results depending on whether the unit of analysis is the individual school or the school district (that is, aggregated across individual schools).

Illustrations of the mistake in communication research

The purpose of this section is to show that researchers do in fact make the sort of reasoning mistake under discussion: inappropriately taking evidence about the individual-level relationship of variables to support conclusions about the message-level relationship of those variables.

Identifying psychological targets for persuasive messages

One common way of identifying psychological states that might be plausible targets for persuasive messages – and thereby justifying designing messages focused on evoking a given state – is by examining individual-level correlations between psychological states and persuasive outcomes. If psychological state *S* is positively correlated with persuasive outcomes, the inference is that messages evoking higher levels of *S* will be more persuasive than messages evoking lower levels of *S*. That sort of reasoning underlies the earlier transportation-persuasion example, in which individual-level correlations between transportation and persuasion were invoked as a justification for the belief that narrative messages that are more transporting will be more persuasive than narrative messages evoking less transportation.

Such reasoning is quite common. For example, Z. Xu and Guo's (2018) meta-analysis reported a substantial positive mean correlation (.54) between individuals' reported guilt and their health-related attitudes and intentions, concluding that "the guilt appeal is a powerful tool for public health" (p. 524). But even if individuals' guilt feelings are positively correlated with their health attitudes and intentions, that does not imply that messages arousing greater guilt will be more persuasive (in effects on attitudes and intentions) than messages arousing lesser amounts of guilt.³

As another example: Tal-Or and Cohen (2016) carefully distinguished transportation and character identification as distinct processes underlying narrative effects. In noting that the two states had different relationships with various outcomes (as indicated by different patterns of individual-level correlations), they pointed out that "while changes in attitudes seem to be affected by both processes, changes in the self tend to be more influenced by identification" (p. 56). The proffered implication for message design was that in contexts such as health communication in which "it is often desirable to increase feelings of vulnerability or to create some change in how people see themselves," "message producers would do well to focus on creating characters with whom audiences will identify" (p. 56). But even if individuals who identify more with a narrative character are more persuaded, that does not imply that messages containing characters with whom audiences identify (that is, messages evoking greater identification) will be more persuasive than messages without such characters (messages evoking less identification).

As another example: Ma and Ma (2022) reported that flu and COVID vaccination intentions were positively correlated with the degree to which individuals were focused on long-term (but not short-term) consequences; the implication drawn was that "vaccination promotion campaigns would benefit from a focus on promoting future positive outcomes of vaccination . . . rather than minimizing immediate concerns" (p. 959). But even if individuals' degree of concern about long-term consequences is positively correlated with their vaccination intentions, that does not imply that messages evoking greater concern about long-term consequences will be more persuasive (in effects on intentions) than messages evoking less concern about such consequences.

As these examples suggest, it is common for researchers to have discovered dependable individual-level relationships between a psychological state (transportation, guilt, identification, concern about future consequences) and an outcome of interest (e.g., persuasion) – and then to have concluded that messages that are more effective (than other messages) in evoking the

³If a concrete demonstration is needed, the same hypothetical dataset can serve. Imagine a study with three persuasive messages (A, B, and C) and assessments of both guilt and intention, with participants' guilt and intention scores as above. The individual-level guilt-intention correlation is +.90, but the message that scored highest on the guilt outcome scored lowest on the intention outcome; the message-level guilt-intention correlation is -1.00. (This same hypothetical dataset can serve as a similar concretization of all the following examples.)

psychological state will also be more effective in producing the outcome. But that reasoning is defective: the existence of positive individual-level correlations between a psychological state and an outcome is not good evidence for the claim that messages varying in the elicitation of the psychological state will correspondingly vary in the elicitation of the outcome. To underwrite a claim that messages varying in the elicitation of the psychological state correspondingly vary in the elicitation of the outcome, the kind of evidence needed is message-level evidence: direct evidence that messages that vary in their elicitation of the psychological state also vary in their elicitation of the outcome.

So if a researcher wants to know whether messages that are more successful in evoking a given psychological state are more persuasive than messages not so successful in evoking that state, the relevant evidence consists of appropriate message-level analyses – analyses that compare messages that vary in evocation of the psychological state of interest. And in fact, in the expectation that messages arousing greater fear or guilt or hope would be more persuasive (compared to messages arousing less fear or guilt or hope), many researchers have conducted appropriate message-level analyses by obtaining experimental data comparing the persuasiveness of messages confirmed to differ in the amount of fear or guilt or hope aroused. For example, Krisher et al. (1973) compared persuasive appeals that differed in the amount of fear aroused, but found no differences in intention. Bozinoff and Ghingold (1983) compared persuasive messages that differed in the amount of guilt aroused, but found no differences in attitude or intention. And Panagopoulos (2014) found that treatments successful in inducing variations in hope did not vary in persuasive effects.

Indeed, when researchers have pursued both individual-level and message-level analyses, disjunctions between the two sets of results have sometimes appeared – exemplifying the potential pitfalls associated with using individual-level associations as a basis for message design choices. For example, Lee et al. (2016) experimentally tested whether messages addressing beliefs known to be more strongly associated with intentions (i.e., stronger individual-level correlations) would be more persuasive than messages addressing beliefs not so strongly associated with intentions; but the message-level analysis found that “contrary to expectations, all messages increased intentions” (p. 433). Similarly, in J. Xu’s (2022) study 1, measures of individuals’ shame were positively correlated with charitable donation intentions; however, study 2—a message-level analysis of shame appeals – found that shame arousal was not associated with such intentions.

Justifying website design choices

The discussion to this point has (for convenience) been phrased in terms of messages, but the same considerations apply to other objects of interest in communication research such as websites. One common way of underwriting website design choices is to examine individual-level correlations between perceptions of website properties. Abstractly expressed, the reasoning is as follows: “individuals’ perceptions of website property X and website property Y are positively correlated; therefore, if you want your website to be perceived as high in X, design it so it is perceived as high in Y.”

For example, Choi et al. (2022) found that (inter alia) perceived interactivity was strongly associated with the perceived credibility of health websites and concluded that “the current findings imply that interactive design features . . . raise the credibility of websites as an information source” (p. 132). But even if individuals’ perceived interactivity scores and perceived credibility scores are positively correlated, that does not show that websites that have (or are perceived as having) more interactive design features will be perceived as more credible than websites lacking (or perceived as lacking) such features.

Koranteng et al. (2022) found perceived task support to be “the most relevant determinant” of perceived website credibility and so recommended that “designers must therefore continue to update their systems with new and relevant features that support user core tasks” (p. 3626). But even if individuals’ scores on perceived task support and perceived credibility are strongly positively correlated, that does not imply that website scores on perceived task support and perceived credibility will be positively correlated.

In these examples, individual-level correlations are taken to underwrite claims about website-level relationships, but this is not sound reasoning; even if the individual-level correlation is strongly positive, that does not imply that the website-level correlation will be positive. To underwrite a claim that websites varying in the elicitation of one psychological state correspondingly vary in the elicitation of another, the kind of evidence needed is website-level evidence: direct evidence that websites that vary in their elicitation of the first psychological state also correspondingly vary in their elicitation of the second.

Choosing contexts for advertising

The program context in which advertising appears has been extensively studied as a possible influence on ad-related outcomes (e.g., ad memory, ad liking, ad effectiveness). One form of evidence has been individual-level correlations between perceptions of media programs and advertising outcomes.

For example, Kwon et al.'s (2019) meta-analysis reported that viewers' memory for advertising was positively correlated with (inter alia) viewers' liking of the media program in which the ad appeared. The implication drawn was that "media users will more likely recognize and recall advertisements when they are placed in media contexts associated with . . . higher program liking" (p. 107). But even if individuals' program liking is positively correlated with their memory for ads, that does not imply that ads in well-liked programs will be better remembered than ads in less-well-liked programs.

As another example, Malthouse et al. (2007) found that individuals' feelings of being absorbed by the stories in a magazine was positively correlated with individuals' liking of the ads in the magazine. The inference drawn was that "an advertisement in a magazine with absorbing stories is worth more to the advertiser than the same ad in a magazine that provides lower levels of this experience" because the former sort of magazine would generate greater ad liking (p. 14). But even if individuals' absorption into stories is positively correlated with their liking for ads, that does not imply that ads in magazines with absorbing stories will be better liked than ads in magazines with less absorbing stories.

In these examples, individual-level correlations are taken to underwrite claims about context-level relationships, but this is not sound reasoning; even if the individual-level correlation is strongly positive, that does not imply that the context-level correlation will be positive. To underwrite a claim that advertising contexts varying in some property (e.g., program liking) correspondingly vary in ad outcomes, the kind of evidence needed is context-level evidence: direct evidence that advertising contexts that vary in that property also correspondingly vary in ad outcomes. For an example of a context-level analysis, see Malthouse and Calder (2010, esp. p. 222).

Justifying proxy outcomes

In communication research aimed at seeing whether different messages produce different effects (e.g., whether gain-framed and loss-framed appeals differ in persuasiveness), researchers sometimes use one outcome as a proxy for a different outcome.⁴ This is routinely justified by pointing to positive correlations between the two outcomes – more carefully, positive individual-level correlations. Expressed abstractly, the reasoning is: "Individuals' scores on outcome X and outcome Y are positively correlated. Therefore the relative standing of messages on outcome X will be a good indicator of the relative standing of messages on outcome Y." But this reasoning is defective, because positive individual-level correlations do not guarantee parallel positive message-level correlations.

In what follows, three examples of such reasoning are discussed: using attitude and intention outcomes as proxies for behavioral outcomes, using measures of perceived message effectiveness as proxies for measures of actual effectiveness, and using neuropsychological assessments as proxies for attitude measures.

⁴This is not a "proxy variable" in the customary technical sense, because the outcome being proxied is not intrinsically unobservable or unmeasurable. But it is a proxy in the sense of being used as a stand-in.

Justifying using attitude and intention outcomes in persuasion effects research

In research comparing the persuasiveness of two messages (e.g., a strong fear appeal and a weak fear appeal), it is common for researchers to use outcomes measures that assess non-behavioral outcomes such as attitude or behavioral intention (e.g., “I intend to get a flu shot”) rather than measures of behavioral performance (actually getting a flu shot). Behavioral outcome assessments are often difficult to obtain, but attitude and intention measures are relatively easy to deploy. So researchers compare the persuasiveness of messages on non-behavioral outcomes rather than behavioral ones, even though the eventual conclusions of interest concern the relative persuasiveness of messages on behavioral outcomes. This practice of using non-behavioral outcomes is often justified by pointing to positive correlations between individuals’ scores on non-behavioral measures and their scores on behavioral measures.

As examples: Anderson et al.’s (2019) study of the effects of educational contraceptive posters used (inter alia) intentions as an outcome variable; they acknowledged that “our study does not assess the impact of these posters on actual behaviors” but noted that “we did measure contraceptive intentions, which have been shown to be a good predictor of behavior” (p. 61). Roberto et al.’s (2017) study of the effects of a cyberbullying prevention intervention used intention measures as outcomes, but pointed to the positive correlations between intentions and behaviors as a rationale: “while we acknowledge that behavior is the gold standard when assessing the effects of any intervention, we are encouraged by the numerous meta-analyses that have consistently found medium to large effect sizes between intentions and behavior in a wide variety of contexts” (p. 6). In their study of the effects of narrative and nonnarrative pandemic messages, Gong et al. (2022) noted as a limitation that “our study assesses intentions rather than actual behaviors,” but also pointed out that “intentions can be strong predictors of actual behaviors” (p. 856). Nan et al. (2015) acknowledged that their study of narrative HPV messaging assessed intentions rather than vaccination behaviors, but noted that “intentions strongly predict actual behaviors” (p. 306). Ratcliff et al.’s (2019) study of gain-loss framing messages used physical activity attitudes and intentions as outcome measures, arguing that “attitudes and intentions are useful indicators of whether a person is likely to continue or adopt an advocated behavior” (p. 2644).

But the reasoning behind such justifications is defective. It could simultaneously be true that (a) *individuals’* scores on non-behavioral and behavioral measures are positively correlated and (b) *messages’* scores on non-behavioral and behavioral measures are negatively correlated. So, for example, even if individuals’ intention and behavior scores are strongly positively correlated, messages’ relative standing on intention outcomes could nevertheless be the opposite of those messages’ standing on behavioral outcomes.

The point here is not that non-behavioral measures are inevitably a poor substitute for behavioral measures in studies of messages’ relative persuasiveness. The point here is that the use of non-behavioral measures to address questions of messages’ relative persuasiveness with respect to behavioral outcomes cannot be justified by pointing to positive individual-level correlations between non-behavioral and behavioral measures. To see whether messages’ relative standing on behavioral outcomes matches their relative standing on non-behavioral outcomes, one needs to undertake message-level analyses. For an example of a message-level analysis, see O’Keefe (2021).

Justifying using perceived effectiveness measures in message pretesting

In message pretesting research aimed at identifying the most persuasive of several candidate messages, researchers have often used measures of perceived message effectiveness (PME) rather than measures of actual message effectiveness (AME, e.g., attitude, intention, or behavior). PME measures might be thought more convenient or easier to administer, so researchers try to identify the relatively more effective message by comparing messages’ scores on PME, even though the eventual conclusions of interest concern messages’ differences on AME.

The practice of using PME measures has often been justified by pointing to positive correlations between individuals' PME scores and their AME scores (e.g., J. P. Dillard et al., 2007). As examples: In validating the Alcohol Message Perceived Effectiveness Scale (AMPES), Jongenelis et al. (2023) pointed to positive correlations between individuals' AMPES scores and "enactment of protective behavioral strategies;" their conclusion was that "the AMPES appears to be an appropriate tool to . . . assist in the development of public health messages designed to reduce alcohol consumption." Kikut and Trzebiński (2023) used a measure of PME as a persuasive outcome, reasoning that "PME has been validated as a strong predictor of behavioral intention and behavior" (p. 6). Simchon et al. (2024) used perceived persuasiveness ratings to assess political messages, writing that "a meta-analysis revealed a reliable link between self-reported persuasion and measures of attitudes or intentions . . . confirming the appropriateness of perceived persuasion ratings for our research" (p. 3). Hackworth et al. (2023) assessed respondents' PME for various cigarette pack inserts, noting that "PME is a predictor of quit intentions and smoking cessation behaviors" (p. 3).

But the reasoning behind such justifications is defective. It could simultaneously be true that (a) *individuals'* PME and AME scores are positively correlated and (b) *messages'* PME and AME scores are negatively correlated. So, for example, even if individuals' PME and AME scores are strongly positively correlated, messages' relative standing on PME could nevertheless be the opposite of messages' standing on AME.

The point here is not that PME measures are a poor substitute for AME measures in research aimed at pretesting messages' relative persuasiveness. The point here is that the use of PME measures to address questions of messages' relative AME standing cannot be justified by pointing to positive individual-level correlations between PME and AME measures. To see whether messages' relative standing on PME matches their relative standing on AME, one needs to undertake message-level analyses. For an example of a message-level analysis, see O'Keefe (2018; also see O'Keefe, 2020).

Justifying using neuropsychological assessments as outcome measures

In recent years, researchers have begun exploring neuropsychological measures as useful additions to the methodological arsenal. Such measures offer the prospect of advancing our understanding of brain-related underpinnings of message effects. But sometimes the justification for the use of such measures takes the form of individual-level correlations with more traditional outcomes.

For example, Ntoumanis et al. (2023) exposed participants to a message about the risks of sugar consumption, reporting that electroencephalographic (EEG) responses were correlated with decreased appeal of sugar products (as assessed by willingness-to-pay). The conclusion was that "EEG is a powerful tool to design and assess health-related advertisements before they are released to the public" (p. 1), but that inference is not well founded. Even if individuals' EEG responses are correlated with attitudes, that does not imply that messages producing stronger EEG responses will be more effective than messages producing weaker EEG responses.

The point here is not that neuropsychological assessments are inevitably a poor substitute for attitude measures in studies of messages' relative persuasiveness. The point here is that the use of neuropsychological measures to address questions of messages' relative persuasiveness cannot be justified by pointing to positive individual-level correlations between neuropsychological measures and attitude measures. To see whether messages' relative standing on neuropsychological measures matches their relative standing on attitude outcomes, one needs to undertake message-level analyses. For examples of message-level analyses, see Imhof et al. (2017) and Wang et al. (2013); sensitivity to the distinction between individual-level and message-level analyses is explicitly present in Burns et al. (2019, p. E5) and Falk et al. (2015, pp. 35–36).

Summary: justifying proxy variables

In a variety of research applications, researchers have justified the use of a given proxy variable by pointing to positive individual-level correlations between the proxy variable and the variable of interest. But for underwriting claims about messages, such justifications are defective. Even if individuals' scores on two variables are positively correlated, that does not show that messages' scores on those two variables are positively correlated. To see whether a given proxy variable is appropriately substituted for some other variable in research advancing claims about messages, what's required is message-level analyses, that is, analyses that speak to the question of whether messages' relative standing on the proxy variable generally matches messages' relative standing on the other variable.

Conclusion

Researchers often invoke individual-level correlations as a basis for claims about message-level correlations. But this is a mistake. When the unit of analysis changes from individual to message, the size and direction of the correlation can change as well. Underwriting claims about message-level relationships thus requires message-level data analyses.

Mediation

Some readers have wondered how statistical mediation analyses fit into this picture. Briefly: There are two questions to be distinguished. One question is whether a message variation produces an effect on an outcome – that is, whether the two message forms being compared (e.g., gain-framed vs. loss-framed, narrative vs. non-narrative, etc.) yield a difference in mean scores on the outcome. The other question is: Assuming there *is* such an effect, what explains it? Mediation analyses can play a role in addressing that latter question, by clarifying the role of intervening psychological states.⁵

But the focus of the present discussion is the first question, and more specifically the focus is: What counts as evidence that a given message variation is systematically related to differences on some other variable (e.g., an outcome)? And the key point being made here is that individual-level relationships (e.g., individual-level correlations) do not provide good evidence; the only suitable evidence comes from message-level analyses.⁶

Avoiding the mistake

There are no easy ways to avoid this reasoning mistake. Researchers naturally discuss whether, or how strongly, two variables are correlated – but without specifying the unit of analysis. “Transportation and persuasion are positively correlated,” “intentions and behaviors are positively correlated,” “perceived website interactivity and perceived website credibility are positively correlated,” and so on. And yet, as seen above, the question of how two variables are correlated when individual is the unit of analysis is different from the question of how those variables are correlated when message is the unit of analysis.

Perhaps the best one can hope for is that in the future researchers might pay closer attention to the unit of analysis when discussing correlations. To say, for example, “transportation and persuasion are positively correlated” is an invitation to misunderstanding, precisely because the unit of analysis is not

⁵The challenges of mediation analyses should not be underestimated. See, e.g., Bullock and Green (2021), Chan et al. (2022), Fiedler et al. (2018), Loh and Ren (2022), Meule (2019), Rohrer et al. (2022), and Xu et al. (2023).

⁶If individual-level correlations are not good evidence for message-level claims, what sort of evidence *is* needed to underwrite message-level claims? Extensive discussion of that question is already in hand. Consider, for example, extant work on the weaknesses of “single-message” designs (in which a single message represents an abstract message category of interest) as a basis for supporting generalizations about message types (e.g., Jackson and Jacobs, 1983; Jackson et al., 1988, 1989). For some relevant subsequent work, see Brashers and Jackson (1999), Clifford et al. (in press), Clifford and Rainey (2024), and Reeves et al. (2016).

specified. “Individuals’ scores on transportation and persuasion are positively correlated” might at least sensitize one to how the unit of analysis is relevant.

Increasing familiarity with multilevel modeling might also alert researchers to this issue. Multilevel modeling is a family of statistical techniques for data analysis when lower levels of units are contained within higher levels, as when students are within classrooms, which are within schools, which are within school districts (for useful overviews, see Park et al., 2008; Robson & Pevalin, 2016). Because multilevel modeling involves attending to multiple levels of analysis, it is common to see an emphasis on ensuring that one’s claims be backed by an appropriate level of analysis – individual-level analyses, for example, for individual-level claims (see, e.g., Diez-Roux, 1998, pp. 218–219).⁷

Summary

Communication research is different from psychology. Claims about messages are different from claims about individuals. And so, for example, the claim that “messages evoking more transportation are generally more persuasive than messages evoking less transportation” is different from the claim that “individuals who experience greater transportation are generally more persuaded than individuals who experience less transportation.”

Correspondingly, what is required to support claims about messages is different from what is required to support claims about individuals. Supporting claims about messages needs message-level analyses, not individual-level analyses. Attending to that difference is crucial to aligning evidence and claims in communication research.

Disclosure statement

I have no known conflict of interest to disclose. Thanks to John Brooks, Hans Hoeken, and Bruce Lambert for helpful discussions.

Notes on contributor

Daniel J. O’Keefe is the Owen L. Coon Professor Emeritus in the Department of Communication Studies at Northwestern University. His research focuses on research synthesis in persuasion and on conceptual problems of data analysis and research design.

ORCID

Daniel J. O’Keefe  <http://orcid.org/0000-0003-1594-8892>

References

- Anderson, S., Frerichs, L., Kaysin, A., Wheeler, S. B., Halpern, C. T., & Lich, K. H. (2019). Effects of two educational posters on contraceptive knowledge and intentions: A randomized controlled trial. *Obstetrics and Gynecology*, 133(1), 53–62. <https://doi.org/10.1097/AOG.0000000000003012>
- Atomistic Fallacy. (2008). In M. Porta (Ed.), *A dictionary of epidemiology*. Oxford University Press. <https://doi.org/10.1093/acref/9780195314496.001.0001>
- Bozinoff, L., & Ghingold, M. (1983). Evaluating guilt arousing marketing communications. *Journal of Business Research*, 11(2), 243–255. [https://doi.org/10.1016/0148-2963\(83\)90031-0](https://doi.org/10.1016/0148-2963(83)90031-0)
- Brashers, D. E., & Jackson, S. (1999). Changing conceptions of “message effects”: A 24-year overview. *Human Communication Research*, 25(4), 457–477. <https://doi.org/10.1111/j.1468-2958.1999.tb00456.x>

⁷Multilevel modeling is not a way of avoiding the reasoning mistake under discussion. Multilevel modeling aims to account for both higher-level and lower-level effects concurrently (so that, e.g., the effect of classroom on student performance is addressed simultaneously with the effects of the individual students). This does not intrinsically prevent researchers from drawing inappropriate conclusions from individual-level analyses.

- Bullock, J. G., & Green, D. P. (2021). The failings of conventional mediation analysis and a design-based alternative. *Advances in Methods and Practices in Psychological Science*, 4(4), 1–18. <https://doi.org/10.1177/25152459211047227>
- Burns, S. M., Barnes, L. N., McCulloh, I. A., Dagher, M. M., Falk, E. B., Storey, J. D., & Lieberman, M. D. (2019). Making social neuroscience less WEIRD: Using fNIRS to measure neural signatures of persuasive influence in a Middle East participant sample. *Journal of Personality and Social Psychology*, 116(3), E1–E11. <https://doi.org/10.1037/pspa0000144>
- Chan, M., Hu, P., & Mak, M. K. F. (2022). Mediation analysis and warranted inferences in media and communication research: Examining research design in communication journals from 1996 to 2017. *Journalism & Mass Communication Quarterly*, 99(2), 463–486. <https://doi.org/10.1177/1077699020961519>
- Choi, W., Kim, S.-Y., & Luo, M. (2022). Design matters in web credibility assessment: Interactive design as a social validation tool for online health information seekers. *Asian Communication Research*, 19(3), 119–138. <https://doi.org/10.20879/acr.2022.19.3.119>
- Clifford, S., Leeper, T. J., & Rainey, C. (in press). Generalizing survey experiments using topic sampling: An application to party cues. *Political Behavior*. <https://doi.org/10.1007/s11109-023-09870-1>
- Clifford, S., & Rainey, C. (2024, February 19). The limits (and strengths) of single-topic experiments. *SocArxiv Manuscript*. <https://doi.org/10.31235/osf.io/zaykd>
- Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health*, 88(2), 216–222. <https://doi.org/10.2105/ajph.88.2.216>
- Dillard, A. J., Ferrer, R. A., & Welch, J. D. (2018). Associations between narrative transportation, risk perception and behaviour intentions following narrative messages about skin cancer. *Psychology & Health*, 33(5), 573–593. <https://doi.org/10.1080/08870446.2017.1380811>
- Dillard, J. P., Weber, K. M., & Vail, R. G. (2007). The relationship between the perceived and actual effectiveness of persuasive messages: A meta-analysis with implications for formative campaign research. *Journal of Communication*, 57(4), 613–631. <https://doi.org/10.1111/j.1460-2466.2007.00360.x>
- Falk, E. B., Cascio, C. N., & Coronel, J. C. (2015). Neural prediction of communication-relevant outcomes. *Communication Methods and Measures*, 9(1–2), 30–54. <https://doi.org/10.1080/19312458.2014.999750>
- Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests: An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, 75, 95–102. <https://doi.org/10.1016/j.jesp.2017.11.008>
- Firebaugh, G. (2015). *Ecological fallacy, statistics of*. *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed. Vol. 6, pp. 865–867). <https://doi.org/10.1016/B978-0-08-097086-8.44017-1>
- Gong, H., Huang, M., & Liu, X. (2022). Message persuasion in the pandemic: U.S. and Chinese respondents' reactions to mediating mechanisms of efficacy. *International Journal of Communication*, 16, 840–863. <https://ijoc.org/index.php/ijoc/article/view/17839>
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. <https://doi.org/10.1037/0022-3514.79.5.701>
- Green, M. C., & Clark, J. L. (2013). Transportation into narrative worlds: Implications for entertainment media influences on tobacco use. *Addiction*, 108(3), 477–484. <https://doi.org/10.1111/j.1360-0443.2012.04088.x>
- Hackworth, E. E., Budiangan, J. R., Lambert, V. C., Kim, M., Ferguson, S. G., Niederdeppe, J., Hardin, J., & Thrasher, J. F. (2023). A mixed-method study of perceptions of cigarette pack inserts among adult smokers from New York and South Carolina exposed as part of a randomized controlled trial. *Health Education Research*, 38(6), 548–562. <https://doi.org/10.1093/her/cyad030>
- Imhof, M. A., Schmalzle, R., Renner, B., & Schupp, H. T. (2017). How real-life health messages engage our brains: Shared processing of effective anti-alcohol videos. *Social Cognitive and Affective Neuroscience*, 12(7), 1188–1196. <https://doi.org/10.1093/scan/nsx044>
- Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research*, 9(2), 169–181. <https://doi.org/10.1111/j.1468-2958.1983.tb00691.x>
- Jackson, S., O'Keefe, D. J., & Jacobs, S. (1988). The search for reliable generalizations about messages: A comparison of research strategies. *Human Communication Research*, 15(1), 127–142. <https://doi.org/10.1111/j.1468-2958.1988.tb00174.x>
- Jackson, S., O'Keefe, D. J., Jacobs, S., & Brashers, D. E. (1989). Messages as replications: Toward a message-centered design strategy. *Communication Monographs*, 56(4), 364–384. <https://doi.org/10.1080/03637758909390270>
- Jongenelis, M. I., Drane, C., Hasking, P., Chikritzhs, T., Miller, P., Hastings, G., & Pettigrew, S. (2023). Development and validation of the alcohol message perceived effectiveness scale. *Scientific Reports*, 13(1), 997. <https://doi.org/10.1038/s41598-023-28141-x>
- Kievit, R. A., Frankenhuys, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, 4, 513. <https://doi.org/10.3389/fpsyg.2013.00513>
- Kikut, A. I., & Trzebiński, W. (2023). The doctor knows or the evidence shows: An online survey experiment testing the effects of source trust, pro-vaccine evidence, and dual-processing in expert messages recommending child COVID-19 vaccination to parents. *Public Library of Science ONE*, 18(7), e0288272. <https://doi.org/10.1371/journal.pone.0288272>

- Koranteng, F. N., Ham, J., Wiafe, I., & Matzat, U. (2022). The role of usability, aesthetics, usefulness and primary task support in predicting the perceived credibility of academic social networking sites. *Behaviour & Information Technology*, 41(16), 3617–3632. <https://doi.org/10.1080/0144929X.2021.2009570>
- Krisher, H. P., Darley, S. A., & Darley, J. M. (1973). Fear-provoking recommendations, intentions to take preventive actions, and actual preventive actions. *Journal of Personality and Social Psychology*, 26(2), 301–308. <https://doi.org/10.1037/h0034465>
- Kwon, E. S., King, K. W., Nyilasy, G., & Reid, L. N. (2019). Impact of media context on advertising memory: A meta-analysis of advertising effectiveness. *Journal of Advertising Research*, 59(1), 99–128. <https://doi.org/10.2501/JAR-2018-016>
- Lee, S. J., Brennan, E., Gibson, L. A., Tan, A. S. L., Kybert-Momjian, A., Liu, J., & Hornik, R. (2016). Predictive validity of an empirical approach for selecting promising message topics: A randomized-controlled study. *Journal of Communication*, 66(3), 433–453. <https://doi.org/10.1111/jcom.12227>
- Loh, W. W., & Ren, D. (2022). Improving causal inference of mediation analysis with multiple mediators using interventional indirect effects. *Social and Personality Psychology Compass*, 16(10), e12708. <https://doi.org/10.1111/spc3.12708>
- Mackenbach, J. P. (2000). Dwalingen in de methodologie. XXVI. De ecologische valkuil en zijn minder bekende tegenhanger, de atomistische valkuil [Roaming through methodology. XXVI. The ecological fallacy and its less well-known counterpart, the atomistic fallacy]. *Nederlands tijdschrift voor geneeskunde*, 144(44), 2097–2100. <https://www.ntvg.nl/artikelen/dwalingen-de-methodologie-xxvi-de-ecologische-vaalkuil-en-zijn-minder-bekende-tegenhanger>
- Malinas, G., & Bigelow, J. (2016). *Simpson's paradox*. *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/fall2016/entries/paradox-simpson>
- Malthouse, E. C., & Calder, B. J. (2010). Media placement versus advertising execution. *International Journal of Market Research*, 52(2), 217–230. <https://doi.org/10.2501/S1470785309201181>
- Malthouse, E. C., Calder, B. J., & Tamhane, A. (2007). The effects of media context experiences on advertising effectiveness. *Journal of Advertising*, 36(3), 7–18. <https://doi.org/10.2753/JOA0091-3367360301>
- Ma, Z., & Ma, R. (2022). Predicting intentions to vaccinate against COVID-19 and seasonal flu: The role of consideration of future and immediate consequences. *Health Communication*, 37(8), 952–961. <https://doi.org/10.1080/10410236.2021.1877913>
- Meule, A. (2019). Contemporary understanding of mediation testing. *Meta-Psychology*, 3, MP.2018.870. <https://doi.org/10.15626/MP.2018.870>
- Nan, X., Dahlstrom, M. F., Richards, A., & Rangarajan, S. (2015). Influence of evidence type and narrative type on HPV risk perception and intention to obtain the HPV vaccine. *Health Communication*, 30(3), 301–308. <https://doi.org/10.1080/10410236.2014.888629>
- Ntoumanis, I., Davydova, A., Sheronova, J., Panidi, K., Kosonogov, V., Shestakova, A. N., Jääskeläinen, I. P., & Klucharev, V. (2023). Neural mechanisms of expert persuasion on willingness to pay for sugar. *Frontiers in Behavioral Neuroscience*, 17, 1147140. <https://doi.org/10.3389/fnbeh.2023.1147140>
- O'Keefe, D. J. (2018). Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication*, 68(1), 120–142. <https://doi.org/10.1093/joc/jqx009>
- O'Keefe, D. J. (2020). Message pretesting using perceived persuasiveness measures: Reconsidering the correlational evidence. *Communication Methods and Measures*, 14(1), 25–37. <https://doi.org/10.1080/19312458.2019.1620711>
- O'Keefe, D. J. (2021). Persuasive message pretesting using non-behavioral outcomes: Differences in attitudinal and intention effects as diagnostic of differences in behavioral effects. *Journal of Communication*, 71(4), 623–645. <https://doi.org/10.1093/joc/jqab017>
- Panagopoulos, C. (2014). Raising hope: Hope inducement and voter turnout. *Basic and Applied Social Psychology*, 36(6), 494–501. <https://doi.org/10.1080/01973533.2014.958228>
- Park, H. S., Eveland, W. P., Jr., & Cudeck, R. (2008). Multilevel modeling: Studying people in contexts. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The SAGE sourcebook of advanced data analysis methods for communication research* (pp. 219–245). Sage.
- Ratcliff, C. L., Jensen, J. D., Scherr, C. L., Krakow, M., & Crossley, K. (2019). Loss/Gain framing, dose, and reactance: A message experiment. *Risk Analysis*, 39(12), 2640–2652. <https://doi.org/10.1111/risa.13379>
- Reeves, B., Yeykelis, L., & Cummings, J. J. (2016). The use of media in media psychology. *Media Psychology*, 19(1), 49–71. <https://doi.org/10.1080/15213269.2015.1030083>
- Roberto, A., Eden, J., Deiss, D., Savage, M., & Ramos-Salazar, L. (2017). The short-term effects of a cyberbullying prevention intervention for parents of middle school students. *International Journal of Environmental Research and Public Health*, 14(9), 1038. <https://doi.org/10.3390/ijerph14091038>
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357. <https://doi.org/10.2307/2087176>
- Robson, K., & Pevalin, D. (2016). *Multilevel modeling in plain language*. Sage.

- Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science*, 5(2). <https://doi.org/10.1177/25152459221095827>
- Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *Proceedings of the National Academy of Sciences Nexus*, 3(2), 035. <https://doi.org/10.1093/pnasnexus/pgae035>
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, 13(2), 238–241. <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>
- Tal-Or, N., & Cohen, J. (2016). Unpacking engagement: Convergence and divergence in transportation and identification. *Annals of the International Communication Association*, 40(1), 33–66. <https://doi.org/10.1080/23808985.2015.11735255>
- Wang, A. L., Ruparel, K., Loughhead, J. W., Strasser, A. A., Blady, S. J., Lynch, K. G., Romer, D., Cappella, J. N., Lerman, C., & Langleben, D. D. (2013). Content matters: Neuroimaging investigation of brain and behavioral impact of televised anti-tobacco public service announcements. *Journal of Neuroscience*, 33(17), 7420–7427. <https://doi.org/10.1523/JNEUROSCI.3840-12.2013>
- Xu, J. (2022). The impact of guilt and shame in charity advertising: The role of self-construal. *Journal of Philanthropy and Marketing*, 27(1), e1709. <https://doi.org/10.1002/nvsm.1709>
- Xu, S., Coffman, D. L., Luta, G., & Niaura, R. S. (2023). Tutorial on causal mediation analysis with binary variables: An application to health psychology research. *Health Psychology*, 42(11), 778–787. <https://doi.org/10.1037/hea0001299>
- Xu, Z., & Guo, H. (2018). A meta-analysis of the effectiveness of guilt on health-related attitudes and intentions. *Health Communication*, 33(5), 519–525. <https://doi.org/10.1080/10410236.2017.1278633>