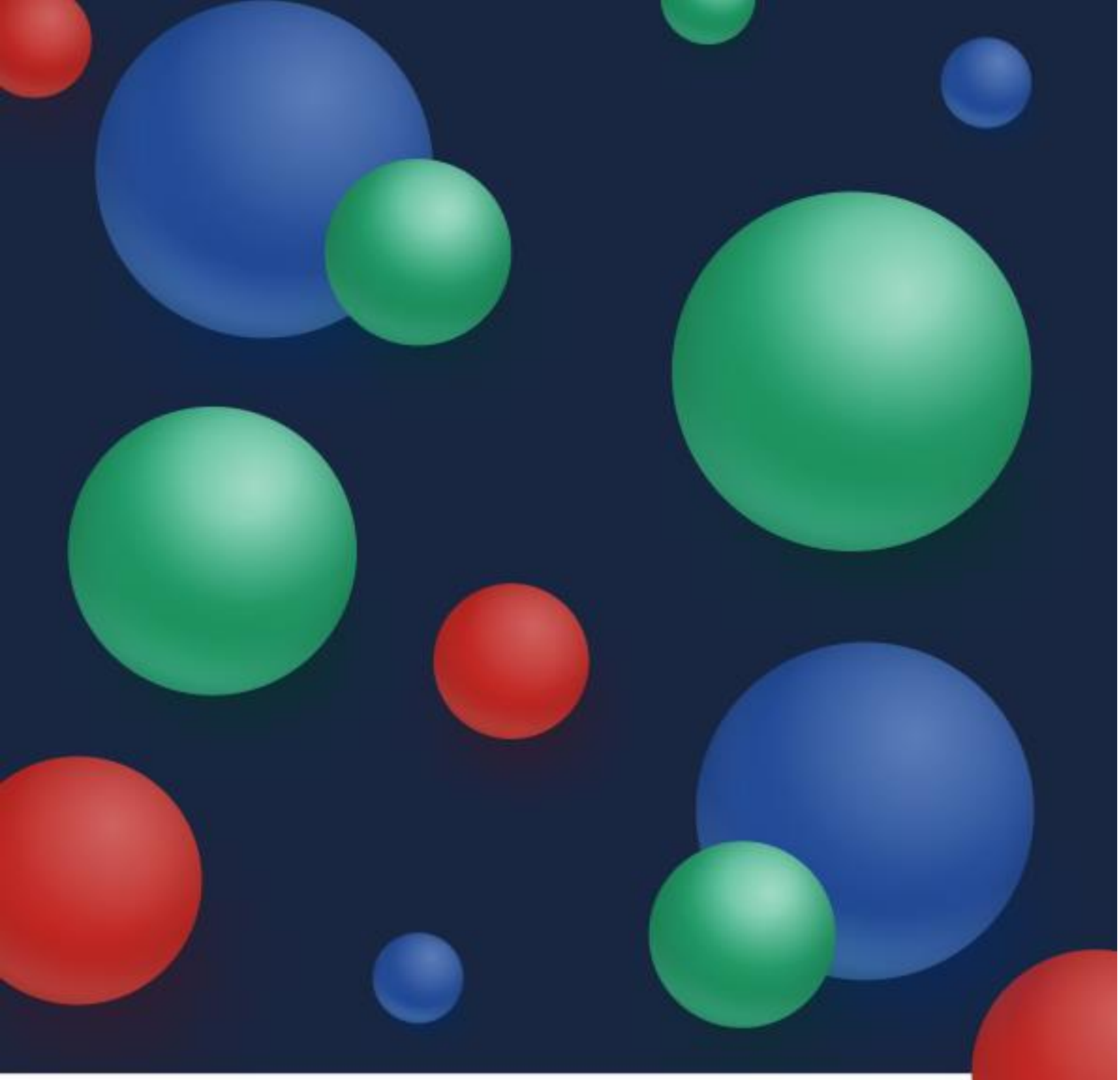


IPIE
International Panel on the Information Environment

Platform Responses to Misinformation

A Meta-Analysis of Data

Synthesis Report 2023.2

Platform Responses to Misinformation

A Meta-Analysis of Data

Synthesis Report 2023.2

SYNOPSIS

Digital misinformation is a critical issue affecting the global information environment. Countering misinformation and its effects is a major objective of governments, international organizations, concerned citizens, and technology firms.

This *Synthesis Report* examines the effectiveness of countermeasures against misinformation on social media platforms, focusing on the two most examined remedies: content labeling and corrective information interventions. A meta-analysis of 43 studies from 18 peer-reviewed manuscripts selected from a comprehensive database of 4,798 publications. A meta-analysis is a research process for synthesizing and aggregating the findings of many, independent studies, using statistical methods to calculate overall effects from multiple data sources.

First, there is an emerging scientific consensus that content labels and corrective information help people evaluate misinformation on social media. Other mitigation strategies may be viable, but there is less consensus about their effectiveness.

Second, understanding the global information environment requires: (i) more research from countries around the world, (ii) about user experiences in languages other than English, (iii) with genuine access to social media data from firms, (iv) that allows scientists to standardize measures and definitions for robustly reporting the results of independent research.

Different publications interpret countermeasures differently, which, along with the variation in the design of the studies we analyzed, makes it difficult to estimate the exact effects of many proposed countermeasures.

The scholarship needs to develop and adopt more standardized measures and definitions, focus on misinformation outside Western contexts, and adopt more robust reporting when testing countermeasures. This will help to ensure confidence when evaluating policy and design interventions that have the potential to improve the global information environment.

CONTENTS

SECTION 1. INTRODUCTION

The reliability of the information that people find on the internet continues to be a significant issue in the global information environment. It is widely known that online platforms can enable the distribution of false or misleading content [1], [2]. This content is often described as "misleading information," "disinformation," "fake news," "rumors," or "computational propaganda" [3], [4], and here these phenomena are referred to under an umbrella term "misinformation." To deter the spread of such digital falsehoods, it is essential to examine how platforms respond to misinformation, and to develop the effective remedies.

In this *Synthesis Report* (2023.2), we present the results of a meta-analysis of publications relying on randomized controlled trials and quasi-experimental design published in peer-reviewed academic journals between November 1, 2006, and December 31, 2022. These publications tested the effectiveness of key measures for combatting the spread of misinformation on social media platforms, that is, countermeasures. This meta-analysis is based on 43 studies published in 18 empirically informed manuscripts that appeared in peer-reviewed journals included in two major academic literature databases. We filtered the manuscripts through an analysis of one of the largest and most comprehensive samples of scientific literature on misinformation spreading on platforms. The samples contains 4,798 peer-reviewed publications, 588 of which satisfy our publication inclusion criteria designed following the recommendations of the rigorous Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 [5].

There is no universally agreed-upon definition of misinformation. As a result, we do not distinguish between various kinds of incorrect, false, or misleading information in this report. Instead, we examine the literature that focuses on "misinformation" and synonymous concepts, such as "propaganda," "disinformation," or "fake news" [1], [3], [4].

We focus on examining the average effect estimates of the most commonly proposed and empirically analyzed countermeasures that aim to mitigate misinformation spreading in digital platforms: content labeling and corrective information. These two

countermeasures are the most recommended remedies according to our systematic review of 588 peer-reviewed publications [6]. Other interventions, unfortunately, are not yet endorsed by a substantial number of scientific publications to be analyzed extensively in this meta-analysis. Conducting this evaluation helps identify a direction for future research about the effectiveness of information & media literacy content, which appears to be even less conclusive than evidence about content labeling and corrective information. Hence, this meta-analysis demonstrates that content labeling and corrective information promise some improvement in people's ability to evaluate misinformation on digital platforms more accurately.

Toolbox of Countermeasures

Scientists and practitioners have offered many tools and remedies for reducing the spread of misinformation online. Table 1 summarizes the countermeasures that are most commonly endorsed and displays the number of publications which passed our quality screens. It reveals that the two countermeasures that this report focuses on are the two most often endorsed. However, we could not include many of the publications that endorsed a countermeasure in our meta-analysis due to methodological or study design differences and incompleteness. Please see *Synthesis Report* (2023.1) for the discussion of how we evaluated the scientific endorsements [6]. In this review, we focus on the studies that evaluate the effectiveness of these common countermeasures.

SECTION 2. PUBLICATION SELECTION

Our report followed the established Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 recommendations [5]. These guidelines are an industry standard for systematic review and meta-analysis studies, are used across disciplines, and are recommended by Cochrane, a scholarly network focused on designing protocols that

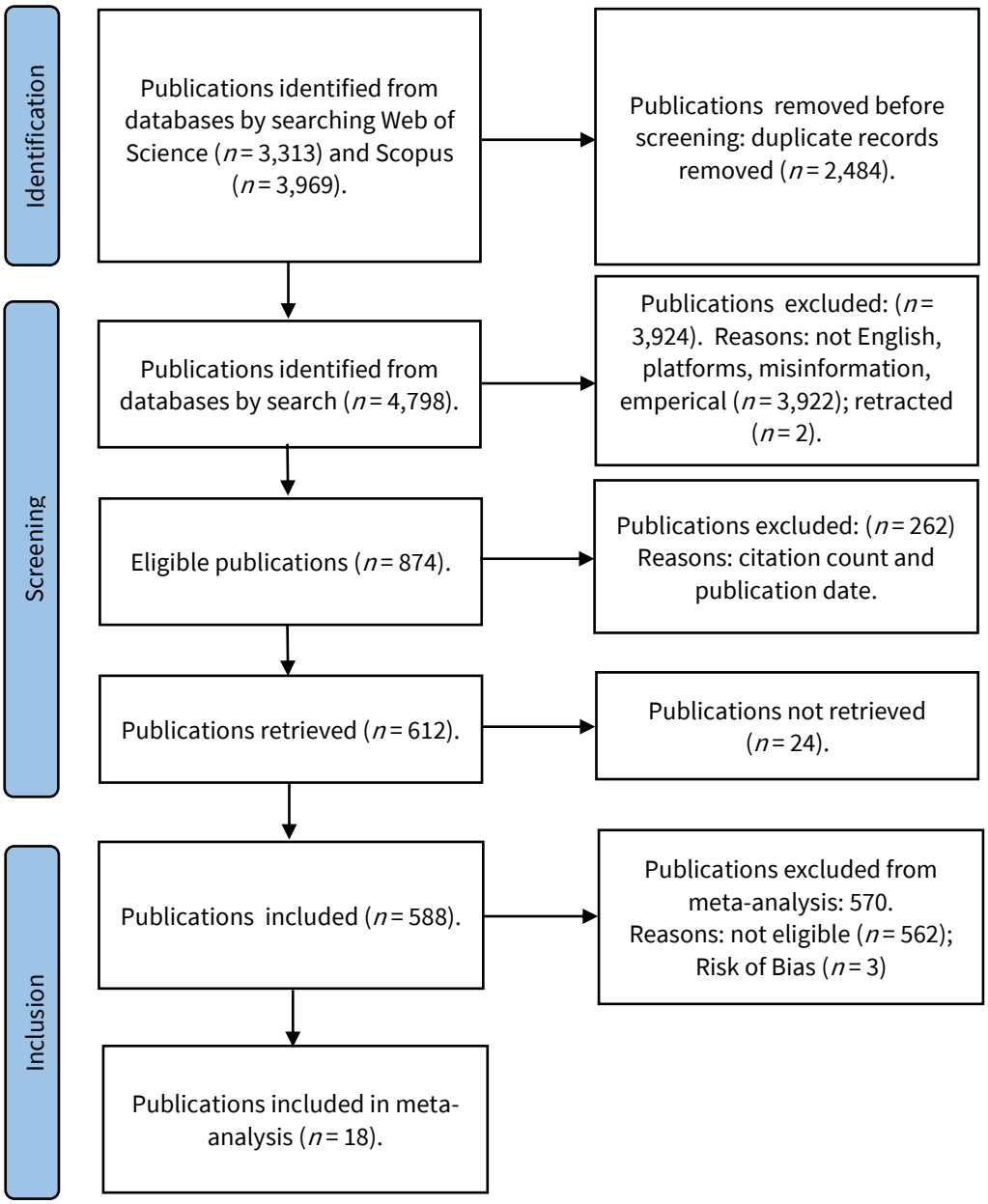
Endorsed Countermeasures for Mitigating Misinformation

Countermeasures	Examples	Number of publications where countermeasure effect reported
Content labeling–Labeling posts, accounts, and stories with tags about fact-checking, funding, or advertising, or any other forms of tagging or flagging, including providing further context without the user having to click through to receive the additional information.	A platform adds a "disputed" label to a user post, or a platform labels posts by state media with a "warning" sign.	41
Content or account moderation–Taking down or marking content; using human or algorithmic moderation to suspend and block accounts.	YouTube downranks content, or Twitter reduces interactions with accounts that users don't follow.	27
Information & media literacy–Educating users to identify misinformation by giving them tips or suggestions or by training them.	Facebook offers Tips to Spot False News, including "be skeptical of headlines," "look closely at the URL," and "investigate the source."	22
Publishing corrective information materials–An organization, platform, or individual provides accurate information without regard to whether users have preconceptions about it.	Governments or private enterprises publicly debunk a rumor on social media in a separate, unlinked piece of content, or user–generated content debunks a conspiracy.	17

Source: [6].

are aimed at enabling high-quality information to be gathered from health research. To complete the meta-analysis, the first step is to find and select as many studies as possible that covered the themes and topics worthy of examination. Therefore, a systematic literature review was conducted before starting the meta-analysis [6]. Figure 1 summarizes the process of finding and selecting the publications for meta-analysis. First, peer-reviewed studies in two academic databases, Web of Science and Scopus, were identified using the selection criteria described in Section S1 of Supplementary Information. Second, after removing duplicated publications resulting from this search, 4,798 publication abstracts were screened for four basic criteria for eligibility: English, misinformation, platforms, and empirical data. This process resulted in 874 publications being moved to the next analysis stage. Third, as explained in Supplementary Information, some publications that did not satisfy the criteria for publication date and citation count were removed. Fourth, the process resulted in identifying 588 empirical publications about misinformation circulating on digital platforms, and all of these were reviewed. Fifth, from these publications, only 21 reported on countermeasures effects that could be

Figure 1. Sampling Stages of Systematic Review.



pooled together for our report. However, three of these publications had to be excluded due to a high risk of bias, as discussed in Section S2 of Supplementary Information.

Note: Flow of publications is presented based on a standard design suggested by the PRISMA recommendations [5].
Source: IPIE calculations based on data collected.

SECTION 3. RESULTS

Methodological Approaches to Evaluating Countermeasures

It is important to rigorously evaluate the effectiveness of the most prominent countermeasures for mitigating misinformation that are being proposed. Content labeling and corrective information are the two most commonly validated countermeasures in the peer-reviewed research on misinformation analyzed here (Table 1). Data on the sample, effect sizes, and other relevant statistics were extracted from individual studies. The dependent variable was defined as forms of information perception. If the key statistics were provided within the studies, allowing for calculations of pooled effect sizes using a random-effects model [7], those studies were included in the meta-analysis.

Before presenting the results of the meta-analysis, it is important to briefly discuss various empirical approaches to assessing the impact of countermeasures for addressing misinformation on social media platforms.

The IPIE *Systematic Review* (SR2023.1) showed that scientists use diverse methods to study countermeasures for misinformation on digital platforms. Content analysis and surveys are the most common methods used among the 588 studies we assessed. These methods are straightforward and can provide important information. However, these methods are likely to give biased estimates if they aim to demonstrate the impact of a certain countermeasure by relying on raw data. There are many factors that might explain differences in misinformation exposure between any groups of users exposed to mitigating countermeasures, and these need to be considered before drawing reliable conclusions about the impact of countermeasures.

Therefore, to collect and compare the most reliable evidence across the studies, two types of studies were used here: (1) randomized control trials (RCT) and (2) non-randomized quasi-experimental studies that test specific countermeasures. Such studies aim to identify the impact of countermeasures for addressing misinformation on exposure levels after controlling for a range of other factors and often compare the impact of countermeasures between two or more groups of people. One of these groups is a control group—those people who were not treated with a countermeasure. All the studies

included in our meta-analysis sample use one of the two types of high-quality experimental or quasi-experimental design. Hence, out of 4,798 publications that we included in our sample initially, 43 used experimental or quasi-experimental designs to quantitatively measure the impact of content labeling and corrective information and satisfied the validity assessment protocols as outlined in Supplementary Information. Other methods, including qualitative approaches, can also produce valuable results, but these should be compared in a different way. See Section S4 of Supplementary Information for the full list of publications reviewed, including those excluded from the sample of content labeling and corrective information studies as the result of bias evaluation, as well as studies reviewed in a supplementary evaluation for the effects of information & media literacy content and information sharing behavior.

Content Labeling

Publications that examine the effects of content labeling primarily analyze how flagging social media posts, such as tweets and news headlines on Facebook, influences the "believability" or "perceived accuracy" of false information. The publications included in our sample conducted experiments with participants, mostly focused on the USA context, with the exception of one study (in addition to another study that did not report on this). RCT studies comprised nearly a quarter of the sample. Table 2 summarizes the types of independent variables that these studies found to be effective: the types of "tags" or "flags" attached to the pieces of content in different domains. The table demonstrates that we have more evidence about "disputed" labels—challenged by fact-checkers, experts, or the media—than about flags based on content rating and source rating in the health information domain. That said, one of the studies that examined the latter flag type demonstrated the largest effect size for improvement in user information perception across the sample (Figure 2).

Table 2. Publications that Test the Effect of Content Labeling on Information Perception.

Domain	Type of content labels	Number of Studies	Publications
Health	"Disputed" flag without correction	5	Kirchner J and Reuter C, 2020 Lee J, Kim JW, and Lee HY, 2022 Yang F and Overton H, 2022
	Content rating	2	Kim, A, Moravec, PL, and Dennis, AR, 2019
	Source rating	2	Kim and Dennis, 2019 Kim, A, Moravec, PL, and Dennis, AR, 2019
Political or social	"Disputed" flag without correction	9	Kirchner J and Reuter C, 2020 Moravec P, Kim A, and Dennis A, 2020 Pennycook G, Cannon T, and Rand D, 2018
	Labeling content as misinformation	5	Chung M and Kim N, 2021 Clayton et al., 2019 Ecker U, O'Reilly Z, Reid J, and Chang E, 2020
	Content rating	1	Clayton et al., 2019

Notes: n = 18 studies from 10 publications; some studies were included in several categories because they covered multiple types of content labeling.
Source: IPIE calculations based on data collected.

Examples of popular labels that researchers successfully tested include caution signs that warn of misinformation and invite to "learn more" [9], "false-tag" retraction flags [10], and labels that give a rating to the source of information. One study distinguishes labels based on the cognition process they triggered (automatic or deliberate cognition) and tests both types of labels for their effect on the believability of information [11]. In addition, a study that nudges social media users to consider the source of an article and then experimentally tests whether a source rating given by users influenced the believability of the content of the article [12].

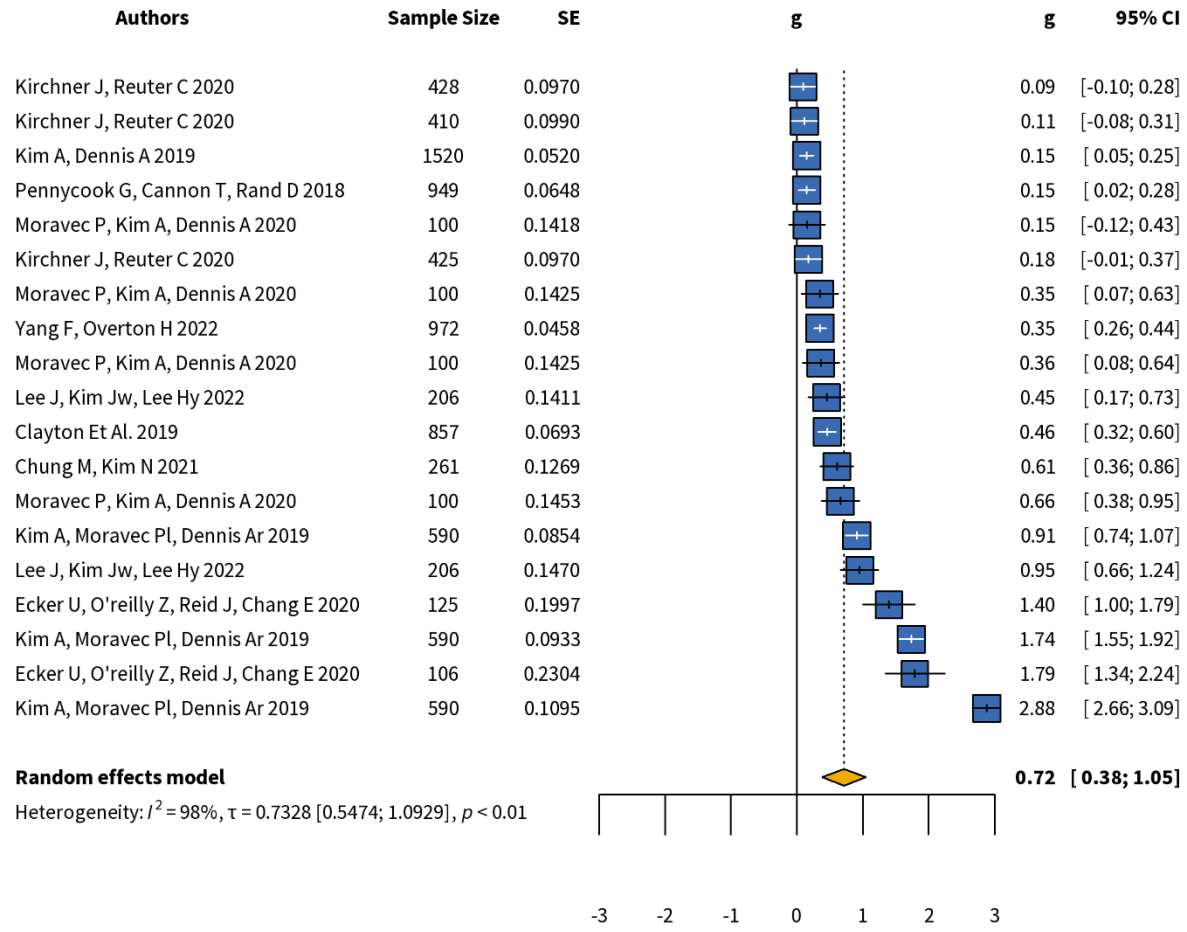
The results of the meta-analysis indicate the effectiveness of content labeling countermeasures. Figure 2 reveals that the effect of adding a label to a misleading statement ranged from 0.09 to 2.88, with the average effect of Hedges' g = 0.72. The 95% confidence interval ranged from 0.38 to 1.05, with some individual prediction outcomes including the null. The effect was significant (p <0.0001). We discuss the interpretation of these results in Section 4.

Nevertheless, I^2 and τ levels, which describe the proportion of observed variability that can be attributed to among-study heterogeneity, are very high ("When I^2 is near zero, the observed variability is mostly down to sampling error; when I^2 is near 100, most of the observed variability reflects differences in population effect sizes"; see [13], [14]). The I^2

measure has disadvantages. Hence it is appropriate to focus on discussing τ levels. A psychology-focused review found that close replication studies reported an average τ or 0.09, but studies that pooled more diverse experiments reported an average τ between 0.31 and 0.35 [14]. In the domain of misinformation research, high heterogeneity is a common issue [15].

In our sample, heterogeneity was substantially higher in many of the subgroups of studies we reported in Figures S1a-S7b. However, when we grouped studies by research design, misinformation domain topic, or type of labels attached to content, heterogeneity often dropped beyond average levels in psychological science, indicating increased certainty about labels' effectiveness in certain knowledge domains or formats. We found that the effects of using labels become less heterogeneous when misinformation is unrelated to health ($\tau = 0.15$ for political misinformation; Figure S6a), or a label conveys factual elaboration rather than simple rebuttal ($\tau = 0.09$; Figure S4a), or a warning sign ($\tau = 0.15$; Figure S5a). Heterogeneity is also reduced substantially when a study adopts an RCT design ($\tau = 0.2$; Figure S1a). This suggests that the source of heterogeneity in these studies is linked to the type of flag attached to content and the source of information displayed in this flag. Additional discussion heterogeneity appears below.

Figure 2. Meta-Analysis of the Effect of Content Labeling on Information Perception.



Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' *g*. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

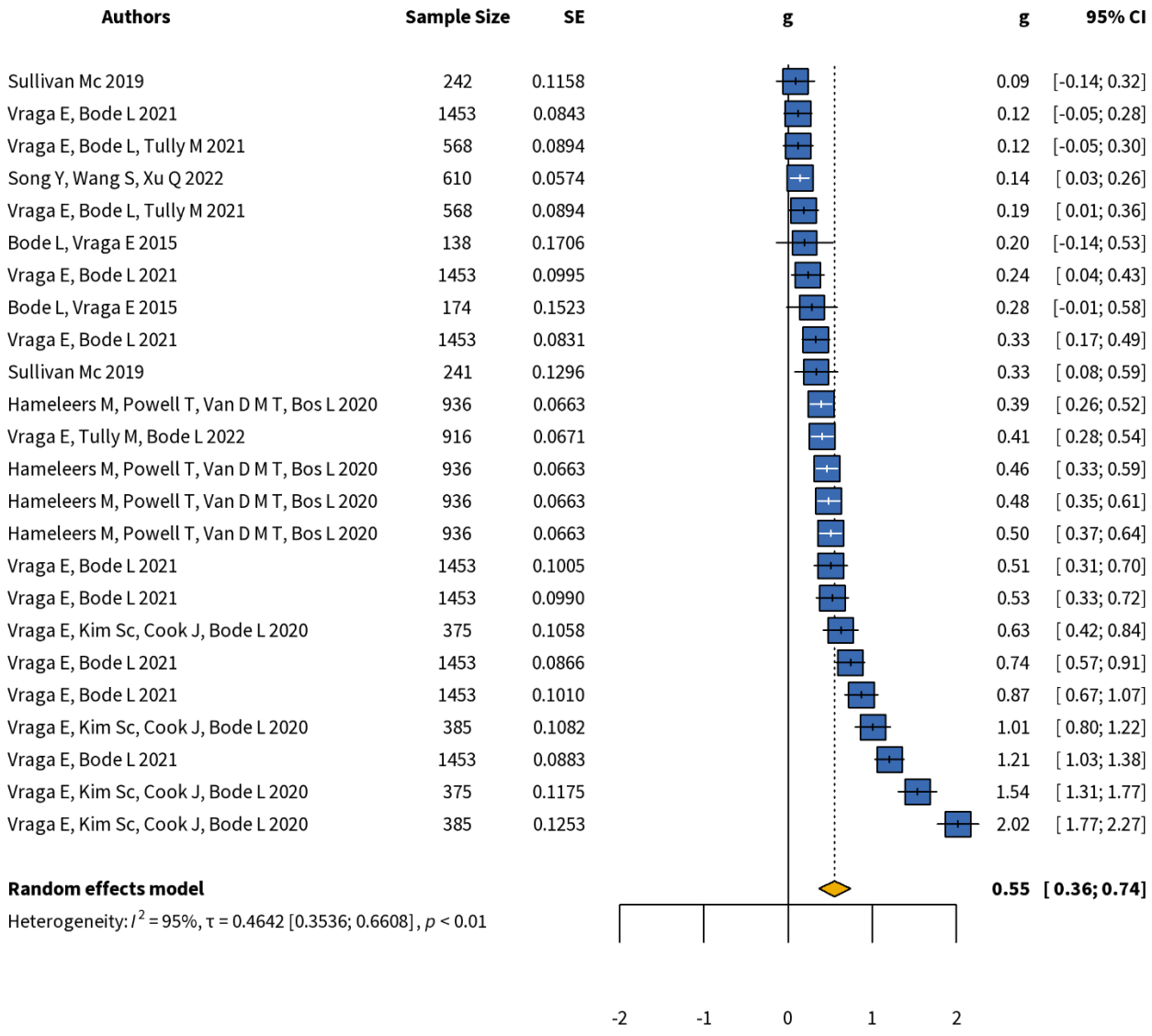
Indeed, the analysis of subgroups of studies shows a lower but more consistent effect when misinformation is linked to a political issue (g = 0.35 [0.19; 0.50], Figure S6a) and when a label takes the form of text accompanied by a warning sign (g = 0.22 [0.08; 0.36], Figure S5a). Although text without any warning signs or flags indicates a large improvement in the ability of users to recognize misinformation, the effects were more heterogeneous. This reveals more certainty that content labels presented in the form of text with a warning sign "authored" by a platform company and applied to misinformation are likely to moderately improve users' perception of information credibility, trustworthiness, or accuracy.

Corrective Information

Studies examining the influence of corrective information materials usually test participants' perceptions before and after seeing fact-based corrective messages about topical issues. These are commonly measured with Likert-scale surveys. The publications included in this sample conducted experiments with participants, mostly focus on the USA context, with the exception of one study that focuses on Hong Kong (in addition to another study that did not report on this). RCT studies comprised four-fifths of this sample. Most common corrections originate from third-party organizations, like NGOs or fact-checkers, or other platform users. One such study examines the change in attitudes toward false statements about genetically modified organisms and another investigates whether logic- or fact-focused corrective materials better tackle misinformation about climate change on Instagram [16].

The results of the meta-analysis on corrective information are shown in Figure 3. Corrective information generally decreases misperception levels and improves the ability of users to perceive the credibility, trustworthiness, or accuracy of information. In other words, following the exposure to corrective information, social media users are less likely to perceive misinformation as credible, believable, or accurate. The pooled effect was $g = 0.55$, with the 95% confidence interval ranging from 0.36 to 0.74. The effect was significant ($p < 0.0001$).

Figure 3. Meta-Analysis of Studies Analyzing the Effect of Corrective Information on Information Misperception.



Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' *g*. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Studies were grouped by the format of the corrective material, and it is observed that more elaborated information materials that focused on fact-supported corrections by relying on text and images rather than on a simple rebuttal using text-only content can produce a stronger effect on people's ability to distinguish "fake news" from other types of information. Figures S4b and S5b illustrate this finding. At the same time, there is a higher confidence level about the average effects of corrective information in the health domain

because most of the analyzed studies focused on health-related misperceptions. Figure S6b, in addition, shows that the largest effects in the studies we analyzed occurred in those that debunked non-health and non-political misinformation, such as climate change or misperceptions about plants. This suggests that the issue in question and the context remain crucially important for the success of corrective information.

Because most studies focus on Facebook and Twitter, we found that the countermeasures varied less when a study discussed these platforms. Only one publication in our sample simulated misinformation on Instagram (with four experiments), but it reported the largest effect on misperceptions [16]. This particular effect, though, came from a design less common in our sample: a post containing a purely scientific statement about climate change was debunked during an experiment, and the perception of participants was measured before and after debunking. We also noted that effect sizes across quasi-experimental studies are generally smaller than those for RCTs (Figure S1b), with the former showing the average effect of information materials at the level of 0.17; the confidence interval crossed the null in most of the reviewed studies in this category, though.

Just as there are many ways to label content, there are many ways to correct misinformation, so again, we observed high levels of heterogeneity ($\tau = 0.46$). However, heterogeneity was reduced if we focused on studies reporting about publishing corrections on Facebook ($\tau = 0.31$; Figure S2b), covering health misinformation ($\tau = 0.3$; Figure S6b), or having a large sample ($\tau = 0.28$; Figure S7b). However, heterogeneity was not reduced if studies are grouped by randomization type or images versus text-only format (Figures S1b, S5b). Finally, we should note that all the analyzed information materials and content labeling studies that reported their geographic context focused only on the USA, and the overall sample was constrained to publications in English.

Information & Media Literacy and Misinformation Sharing

Two separate smaller samples were also reviewed to provide preliminary evidence for a direction for future research. One included four studies that tested whether information & media literacy content can improve users' ability to detect misinformation (Figure S8).

Another sample contained two studies asking whether content labeling can affect users' plans to share misinformation (Figure S9). First, information & media literacy content can have a marginally positive effect on information perception. However, the confidence interval for all but one study in the sample reached or crossed the null. Second, the two pooled studies found that attaching labels can help to prevent sharing of misinformation on social media. The test of heterogeneity was statistically significant, though the overall random-effects model was not ($p = 0.12$). This evidence from these two samples is insufficient to draw any conclusions, as only a few studies from our large sample are able to offer any relevant evidence.

Heterogeneity

An inevitable limitation of our broad and ambitious approach to literature sampling is the high levels of heterogeneity in the scholarship, as revealed by our meta-analysis. Given that a relatively small number of studies available for the full examination and that the quality of the evidence from these studies was sometimes less than optimal, it was hard to tease out the exact causal sources of heterogeneity. Heterogeneity is an omnipresent problem in social science research due to the wide variation of methods used and the characteristics of the independent variables chosen. For example, a recent meta-analysis of the effectiveness of fact-checking for correcting political misinformation also reported the high heterogeneity of the pooled studies (86%) [17]. Another meta-analysis of the factors underlying messages that are effective in countering misinformation found a 99% level of heterogeneity [18].

To address the high levels of heterogeneity and the likelihood of publication bias in our sample, the established practices and tested the outcomes for subgroups were followed [19]. The subgroups were defined theoretically, and the details for their coding can be found in Table S1. A key source of heterogeneity is in study design or the study's risk of bias of the individual publications we analyzed. In particular, there were differences in the definitions. For instance, some publications defined content labels as simple signs, while others provided elaborate definitions. In this meta-analysis, studies were grouped by more narrow definitions of content labels, so heterogeneity was reduced. In the case of content labeling, some groups of studies demonstrated levels of heterogeneity that suggested

higher levels of confidence in our results. At the same time, grouping corrective information studies decreased heterogeneity only in a few cases. This provides some support for our definition discrepancy hypothesis: corrective materials can come in multiple forms and types, while content label formats and presentations are potentially less diverse. Still, the high heterogeneity levels make any generalization from our broad sample less consistent.

Heterogeneity can also arise from the prominence of quasi-experimental studies in our sample. An emerging discussion on the standards for reviews summarizing non-medical research, such as Education Research, recommends applying strong criteria for study selection if effects are pooled from quasi-experimental studies [20]. Developing robust standards for meta-analysis research for social science should be one of the priorities of future work.

Discussion of Results

In several respects, additional research that uses diverse methods and consistently analyzes results will allow for more conclusive evaluations of all of the proposed ways of responding to misinformation. First, the literature was not conclusive regarding the size of the effects of the proposed countermeasures. Deciding whether an effect is “large” is not straightforward in any field, and this becomes even more difficult when researchers combine studies from different fields. The effect sizes for two key countermeasures we analyzed—0.72 for labeling (Figure S1a) and 0.55 for corrective information (Figure S1b)—fall between “medium (~0.5 s.d.)” and “large (~0.8 s.d.)” as per Cohen’s suggestions. Despite different fields commonly applying this interpretation of the findings, it was not intended to be absolute [21]. Cohen cautioned that a generic application of effect size interpretation to all research fields was “an operation fraught with many dangers” because of the risk of being misunderstood [21] (p. 12), while these and similar conventions are arbitrary but necessary for statistical analyses. This means that the size of effects must be judged within the context of the field and the methods used in the study.

Another way of norming effect size is to consider the size of countermeasure effects against the findings of other studies. A recent study on the effects of fact-checking in

correcting political misinformation reported a "positive and significant" ($d = 0.29$) effect of fact-checking [17]. An unpublished meta-analysis of the effects of media literacy campaigns found "the effect of media literacy interventions on fake news discernment" measured at the level of $d = 0.53$ [22].

How should we evaluate the average effect size observed? For instance, it may be interpreted as meaning that correcting materials can generate a slightly larger improvement than media literacy campaigns. Moreover, practical considerations are also important. For example, a measure might show a smaller effect across a range of studies, but it may be more cost-effective and practical to implement this measure rather than more complex solutions, particularly for influence long-term outcomes. This is particularly the case regarding content labeling compared to much more costly solutions, such as information & media literacy campaigns.

Second, this investigation proceeded with one outcome variable, the perception of information. The studies we analyzed used varying definitions: some emphasize users' trust in content or its "believability," while others straightforwardly asked participants to distinguish "fake news" from other types of information. This variation was substantial to introduce high heterogeneity that we observed in almost all the subgroups we analyzed. In addition, only two publications worked with a non-USA sample (Southeast Asia and the EU), and two more failed to specify where their participants were based.

Third, our meta-analysis may also have the small-study bias of the reviewed research, as some funnel plots depict asymmetric effect sizes. In addition, 95% prediction intervals for the pooled effects sometimes included the null. This suggests that some results might be overestimates that include no impact. Moreover, the publications often provide insufficient evidence about the relative effectiveness of all the different countermeasures being considered. For example, one publication shows an effect of a logic-focused information intervention but not a fact-focused one [38]. Even within the same publication, effect sizes vary according to the measured aspect of information perception.

Despite the exponential growth in publications about the effectiveness of countermeasures to misinformation, we were only able to include 18 manuscripts in our meta-analysis due to a lack of standardized measures and definitions, a high risk of bias, or the lack of relevant statistics reported. This small sample size means that we were

unable to robustly discuss solutions proposed by study authors. The relatively small sample size suggests that the field needs to ensure more robust reporting of research results. Another area of improvement that scientific inquiry needs to implement is developing more standardized measures and definitions. Nevertheless, previous research has demonstrated that studies based on typical social science research designs can still provide good-quality material for meta-analysis [23], [24].

SECTION 4. CONCLUSION

Online misinformation is a critically important issue affecting the global information environment. Countering misinformation and its effects is a major objective of governments, international organizations, and concerned citizens. Being successful will require a long-term approach involving various strategies and tools. In this Synthesis Report, we used meta-analysis to investigate the impact of countermeasures on mitigating misinformation that spreads on digital platforms and can be implemented by platforms. Based on previous research, several of the most recommended countermeasures were identified [6]. The available evidence allowed for a focus on two remedies: content labeling and corrective information. The empirical studies analyzed rely on diverse premises and use varying approaches, and the meta-analysis confirms several robust findings, summarized in Table 3.

First, there is some evidence supporting the role of content labeling in mitigating the effects of misinformation by moderately improving platform users' ability to perceive information credibility, trustworthiness, or accuracy. Such labels as a source rating or warning sign based on fact-checking promise improvement in users' perception of (mis)information. At the same time, content labels presented in the form of text with a warning sign and applied to political misinformation are linked to smaller but more consistent effects. However, these effects are highly heterogeneous from study to study—probably due to the differences in study design and risk of bias—which decreases the certainty of conclusions.

Second, there is some evidence that corrective information can mitigate the effects of misinformation. Corrective information remedies that used factual elaboration were more effective in reducing misperceptions than simple rebuttals. Moreover, combining approaches, such as labeling and corrective information, can be more effective than any one remedy as evidenced in a recent study [25]. However, the issue in question and the context remain crucially important for the success of corrective information. In addition, the effects of corrective information and content labeling are hard to pin down to any specific number. It is possible that some corrective strategies will have no effect at all, as

the 95% prediction intervals were consistent with negative, null, and positive effects in particular cases.

Third, it is unclear what the effects of information & media literacy are. The nascent evidence suggests that literacy content can be effective, but perhaps less effective than labeling or corrective information. Further research should test if information & media literacy campaigns have consistent, positive effect on users. Sharing behavior, one of the strongest drivers of misinformation spread online, remains under-investigated as well.

Fourth, this review points to important gaps in the understanding of the effectiveness of other possible countermeasures to combat online misinformation. Further research is also needed here, to determine what types and sources of labels and corrective information are most effective. The effectiveness of countermeasures will also vary by platform. Platforms such as TikTok, for example, rely on more visual content, while others emphasize text-based content.

Research results should always be interpreted with caution, and with awareness of study caveats and limitations. Many studies rely on convenience samples because technology firms do not share data required for rigorous analysis. Other studies lack the random assignment of a countermeasure, or have other methodological shortcomings. In addition, the vast majority of research relies on experiments conducted in the USA, which leads to skewed results as the types of individuals in USA populations may not be generalizable to other populations [26]. This complicates broad generalization about the dynamics of the global information environment.

Overall, this meta-analysis provides evidence that content labeling and corrective information are effective countermeasures for mitigating the impact of misinformation on digital platforms.

Table 3. Strategies for Improving the Global Information Environment.

Consensus		Countermeasures	Examples
Endorsed	Validated	Content labeling–Labeling posts, accounts, and stories with tags about fact-checking, funding, or advertising, or any other forms of tagging or flagging, including providing further context without the user having to click through to receive the additional information.	A platform adds a "disputed" label to a user post, or a platform labels posts by state media with a "warning" sign.
		Corrective information–An organization, platform, or individual provides accurate information without regard to whether users have preconceptions about it.	Governments, private enterprises, or users publicly debunk a rumor or conspiracy on social media in a separate, unlinked piece of content.
		Content or account moderation–Taking down or marking content; using human or algorithmic moderation to suspend and block accounts.	YouTube downranks content, or Twitter reduces interactions with accounts that users don't follow.
		Information & media literacy–Educating users to identify misinformation by giving them tips or suggestions or by training them.	Facebook offers Tips to Spot False News, including "be skeptical of headlines," "look closely at the URL," and "investigate the source."
Uncertain		Advertisement policy–Modifying the advertisement policy of the platform, which often adds a user-facing component to the advertising mechanisms.	Facebook requires the "Paid for by" label or introduces an information button for advertisements.
		Content reporting–Changing how users report potential misinformation on a platform.	TikTok introduces a "misinformation" option in the content reporting options.
		Content user sharing–Targeting the distribution of misleading content by users.	WhatsApp limits opportunities to forward a message, or Pinterest prevents pinning or saving posts.
		Disclosure–Informing a user that they have come in contact, shared, or interacted with misinformation.	Reddit tells users they have interacted with misinformation.
		Redirection–Redirecting users to additional information, accounts, or posts, usually by taking users to different content or by overlaying accurate information and alerts.	Instagram shows content from local health authorities when users search for COVID-19 information, or Facebook and Twitter introduce election hubs before the election period.
		Security or verification–Increasing or decreasing the security or verification requirements on a platform.	Twitter's protection program for political officials.
		Self-fact-checking–Providing users with an opportunity to fact-check information for themselves.	A platform offers users an opportunity to interact with fact-checkers to verify the information they consume using private messages.

Note: White cells indicate plausible interventions with uncertain effects as discussed in SR2023.1, light orange indicates the strategies endorsed by a systematic review of the research SR2023.1, dark orange indicates the strategies validated by the research and validated through this meta-analysis.

REFERENCES

[1] C. Wittenberg and A. J. Berinsky, 'Misinformation and Its Correction', in *Social Media and Democracy: The State of the Field, Prospects for Reform*, Cambridge University Press, 2020, pp. 163–198.

[2] M. Broussard, *More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. Cambridge, Massachusetts: MIT Press, 2023.

[3] A. Herasimenka, J. Bright, A. Knuutila, and P. N. Howard, 'Misinformation and Professional News on Largely Unmoderated Platforms: The Case of Telegram', *J. Inf. Technol. Polit.*, vol. 20, no. 2, pp. 1–15, 2023, doi: 10.1080/19331681.2022.2076272.

[4] P. N. Howard, *Lie machines: How to Save Democracy From Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. New Haven: London: Yale University Press, 2020.

[5] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and The PRISMA Group, 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement', *PLOS Med.*, vol. 6, no. 7, p. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.

[6] International Panel on the Information Environment, *Countermeasures for Mitigating Digital Misinformation: A Systematic Review*. in SR2023.1. Zurich, Switzerland: IPIE, 2023. [Online]. Available: www.IPIE.info

[7] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd, 2009. doi: 10.1002/9780470743386.

[8] V. Suarez-Lledo and J. Alvarez-Galvez, 'Prevalence of Health Misinformation on Social Media: Systematic Review', *J. Med. Internet Res.*, vol. 23, no. 1, p. e17187, Jan. 2021, doi: 10.2196/17187.

[9] C. Lanius, R. Weber, and W. I. MacKenzie, 'Use of Bot and Content Flags to Limit the Spread of Misinformation Among Social Networks: A Behavior and Attitude Survey', *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 32, Mar. 2021, doi: 10.1007/s13278-021-00739-x.

[10] U. K. H. Ecker, Z. O'Reilly, J. S. Reid, and E. P. Chang, 'The Effectiveness of Short-Format Refutational Fact-Checks', *Br. J. Psychol.*, vol. 111, no. 1, pp. 36–54, 2020, doi: 10.1111/bjop.12383.

[11] P. L. Moravec, A. Kim, and A. R. Dennis, 'Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media', *Inf. Syst. Res.*, vol. 31, no. 3, pp. 987–1006, Sep. 2020, doi: 10.1287/isre.2020.0927.

[12] A. Kim and A. R. Dennis, 'Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media', *MIS Q.*, vol. 43, no. 3, pp. 1025–1039, Jan. 2019, doi: 10.25300/misq/2019/15188.

[13] J. P. T. Higgins and S. G. Thompson, 'Quantifying Heterogeneity in a Meta-Analysis', *Stat. Med.*, vol. 21, no. 11, pp. 1539–1558, 2002, doi: 10.1002/sim.1186.

[14] A. H. Linden and J. Hönekopp, 'Heterogeneity of Research Results: A New Perspective From Which to Assess and Promote Progress in Psychological Science', *Perspect. Psychol. Sci.*, vol. 16, no. 2, pp. 358–376, Mar. 2021, doi: 10.1177/1745691620964193.

[15] K. Bryanov and V. Vziatysheva, 'Determinants of Individuals' Belief in Fake News: A Scoping Review Determinants of Belief in Fake News', *PLoS ONE*, vol. 16, no. 6, p. e0253717, Jun. 2021, doi: 10.1371/journal.pone.0253717.

[16] E. K. Vraga, S. C. Kim, J. Cook, and L. Bode, 'Testing the Effectiveness of Correction Placement and Type on Instagram', *Int. J. Press.*, vol. 25, no. 4, pp. 632–652, Oct. 2020, doi: 10.1177/1940161220919082.

[17] N. Walter, J. Cohen, R. L. Holbert, and Y. Morag, 'Fact-Checking: A Meta-Analysis of What Works and for Whom', *Polit. Commun.*, vol. 37, no. 3, pp. 350–375, May 2020, doi: 10.1080/10584609.2019.1668894.

[18] M. S. Chan, C. R. Jones, K. Hall Jamieson, and D. Albarracín, 'Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation', *Psychol. Sci.*, vol. 28, no. 11, pp. 1531–1546, Nov. 2017, doi: 10.1177/0956797617714579.

[19] R. Ryan and Cochrane Consumers and Communication Review Group, 'Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: Planning the Analysis at Protocol Stage', Dec. 2016. [Online]. Available: http://cccrg.cochrane.org

[20] What Works Clearinghouse, 'Procedures Handbook', 4.1, 2012. [Online]. Available: http://ies.ed.gov/ncee/wwc

[21] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Lawrence Erlbaum Associates, 1988.

[22] C. Lu, B. Hu, M.-M. Bao, C. Bi, and X.-D. Ju, 'Can Media Literacy Interventions Improve Fake News Discernment? A Meta-Analysis'. Rochester, NY, Mar. 14, 2023. doi: 10.2139/ssrn.4377372.

[23] N. W. Jager, J. Newig, E. Challies, E. Kochskämper, and H. von Wehrden, 'Case Study Meta-Analysis in The Social Sciences. Insights on Data Quality and Reliability From a Large-N Case Survey', *Res. Synth. Methods*, vol. 13, no. 1, pp. 12–27, 2022, doi: 10.1002/jrsm.1514.

[24] G. Tong and G. Guo, 'Meta-analysis in Sociological Research: Power and Heterogeneity', *Sociol. Methods Res.*, vol. 51, no. 2, pp. 566–604, May 2022, doi: 10.1177/0049124119882479.

[25] J. B. Bak-Coleman *et al.*, 'Combining interventions to reduce the spread of viral misinformation', *Nat. Hum. Behav.*, vol. 6, no. 10, Art. no. 10, Oct. 2022, doi: 10.1038/s41562-022-01388-6.

[26] J. Henrich, S. J. Heine, and A. Norenzayan, 'Most people are not WEIRD', *Nature*, vol. 466, no. 7302, Art. no. 7302, Jul. 2010, doi: 10.1038/466029a.

[27] M. Norris and C. Oppenheim, 'Comparing Alternatives to the Web of Science for Coverage of the Social Sciences' Literature', *J. Informetr.*, vol. 1, no. 2, pp. 161–169, Apr. 2007, doi: 10.1016/j.joi.2006.12.001.

[28] D. Mahl, M. S. Schäfer, and J. Zeng, 'Conspiracy Theories in Online Environments: An Interdisciplinary Literature Review and Agenda for Future Research', *New Media Soc.*, p. 14614448221075760, Feb. 2022, doi: 10.1177/14614448221075759.

[29] A. Bechmann and S. Lomborg, 'Mapping Actor Roles in Social Media: Different Perspectives on Value Creation in Theories of User Participation', *New Media Soc.*, vol. 15, no. 5, pp. 765–781, Aug. 2013, doi: 10.1177/1461444812462853.

[30] A. P. Siddaway, A. M. Wood, and L. V. Hedges, 'How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses', *Annu. Rev. Psychol.*, vol. 70, pp. 747–770, Jan. 2019, doi: 10.1146/annurev-psych-010418-102803.

[31] K. Yadav, 'Platform Interventions: How Social Media Counters Influence Operations', *Carnegie Endowment for International Peace*, Jan. 25, 2021.

[32] L. Courchesne, J. Ilhardt, and J. N. Shapiro, 'Review of Social Science Research on the Impact of Countermeasures Against Influence Operations', *Harv. Kennedy Sch. Misinformation Rev.*, Sep. 2021, doi: 10.37016/mr-2020-79.

[33] J. P. T. Higgins and S. Green, Eds., *Cochrane Handbook for Systematic Reviews of Interventions*, 5.1.0. The Cochrane Collaboration, 2011. Accessed: Jul. 25, 2022. [Online]. Available: https://training.cochrane.org/handbook

[34] R. Ryan and Cochrane Consumers and Communication Review Group, 'Cochrane Consumers and Communication Group reviews: Meta-analysis', Dec. 2016. [Online]. Available: http://cccrg.cochrane.org

[35] G. Knapp and J. Hartung, 'Improved Tests for a Random Effects Meta-Regression with a Single Covariate', *Stat. Med.*, vol. 22, no. 17, pp. 2693–2710, Sep. 2003, doi: 10.1002/sim.1482.

[36] J. P. T. Higgins *et al.*, 'The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials', *BMJ*, vol. 343, p. d5928, Oct. 2011, doi: 10.1136/bmj.d5928.

[37] J. A. C. Sterne *et al.*, 'RoB 2: a revised tool for assessing risk of bias in randomised trials', *BMJ*, p. l4898, Aug. 2019, doi: 10.1136/bmj.l4898.

[38] J. A. Sterne *et al.*, 'ROBINS-I: A Tool for Assessing Risk of Bias in Non-Randomised Studies of Interventions', *BMJ*, p. i4919, Oct. 2016, doi: 10.1136/bmj.i4919.

[39] L. G. Smithers, A. C. P. Sawyer, C. R. Chittleborough, N. M. Davies, G. Davey Smith, and J. W. Lynch, 'A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes', *Nat. Hum. Behav.*, vol. 2, no. 11, Art. no. 11, Nov. 2018, doi: 10.1038/s41562-018-0461-x.

[40] T. d. Stanley and H. Doucouliagos, 'Picture This: A Simple Graph That Reveals Much Ado About Research', *J. Econ. Surv.*, vol. 24, no. 1, pp. 170–191, 2010, doi: 10.1111/j.1467-6419.2009.00593.x.

[41] N. Walter, J. J. Brooks, C. J. Saucier, and S. Suresh, 'Evaluating the Impact of Attempts to Correct Health Misinformation on Social Media: A Meta-Analysis', *Health Commun.*, vol. 36, no. 13, pp. 1776–1784, Nov. 2021, doi: 10.1080/10410236.2020.1794553.

SUPPLEMENTARY INFORMATION

SECTION S1. Methodological Details

Search Strategy

We searched for peer-reviewed studies in two academic databases, Web of Science and Scopus. These databases offer a valid instrument for evaluating scholarly contributions in social science [27] and have been used in past systematic reviews and meta-analyses that asked similar questions [15], [28]. In addition, they incorporate key publications in social sciences, humanities, health, and other areas that are likely to publish relevant work, including journals that publish qualitative research. These databases also provide an interface that allows the extraction of large amounts of bibliographic information that were needed for our large-scope comprehensive study. As we focused only on studies of misinformation, we searched for the synonyms for "misinformation" in publication's abstract, title, or keywords while searching. For the publication selection, we adopted an intentionally broad definition array of misinformation (see Table S1 in Supplementary Information). We did this by including any publication that used the term "misleading information" or "disinformation" or "propaganda" or "fake news" or "rumors" to describe the main research questions of the article.

Eligibility criteria (see details in Table S1):

1. Publication date. A study published between 1st of November 2006 (after Facebook, perhaps the first modern social media platform, was made publicly available) and 31st of December 2022;

2. English. Study available in English;

3. Misinformation. Study discusses any aspects of misinformation, disinformation, fake news, propaganda, or similar concepts;

4. Digital platforms. Mention of any social media platform as a study object; a digital platform was defined as a website that is dependent on user-generated content that facilitates two-way interaction [29]. However, we did not focus on studies offering solutions exclusively in relation to misinformation spreading on non-social media websites, such as those of mainstream media or political websites;

5. Empirical. Study is based on empirical evidence—information acquired by observation or experimentation that is analyzed in a scientific publication. This includes qualitative and quantitative research designs.

We used eight terms for "misinformation" and three terms for "social media" to search for publications. See Table S3 for the full Boolean search string. We tested several search term combinations, but the one we chose (Table S3) ensured the highest recall compared to other search term combinations. In addition, we deliberately refrained from including specific prominent platform names, including Twitter and YouTube (although we tested this approach), as this would have skewed the search results by systematically overlooking lesser-known platforms. We decided not to consider gray literature, such as non-peer-reviewed conference abstracts or presentations, because articles that have been through peer review are more likely to provide rigorous findings, complete methodology information, more refined analysis, and transportable statistics on effects.

Screening and Coding

Once selections from each database were finalized, we merged the results from the databases (N_{Scopus} = 3,313, N_{WOS} = 3,969) as shown in Figure 1. After eliminating duplicates, we were left with a total of 4,798 publications. We developed a codebook to assess the eligibility of the collected publications (Table S1). These codes covered eligibility criteria, methods, proposed countermeasures, and their measurement. Five researchers ran a pilot test in pairs (four graduate students and a postdoctoral researcher) to achieve an optimum level of reliability in the proposed coding template. Intercoder reliability showed high agreement based on a random sample of 163 publications for eligibility criteria (see Table S2 for Krippendorff's α measures). Disagreements were discussed in the group and resolved [30].

The five coders read the titles, abstracts, and keywords of the publications to affirm the eligibility of a study based on the criteria listed above. Some publications contained all of the search terms but did not consider misinformation as an object of study, were not empirical, or did not focus on digital platforms. Eventually, 874 publications met our eligibility criteria. Due to the size of the resulting publication dataset and resource constraints, we prioritized and reviewed all publications that appeared online before July 2022 and had been cited more than 11 times since their publication, which was verified using scholarly databases mentioned above. To account for newer studies that may have reached this citation threshold, we also reviewed all articles published between January 2020 and July 2022 without verifying the citation count. allowed us to focus on the most prominent and recent scientific research (the average age of a publication in our final dataset was 2.6 years as of March 2023) and on those studies that might have reached large numbers of citations due to their recent publication. This mixed approach to study sampling is an accepted strategy in systematic analysis [30]. Furthermore, 24 publications could not be retrieved, either because the library of University of Oxford that we used for access did not have a subscription to the relevant publisher or because a URL provided by a database was not active. The process resulted in the identification of 588 empirical publications about misinformation circulating on digital platforms, and we reviewed all of them.

Synthesis and Analysis

After coding for eligibility, we switched to three sets of codes that summarized the literature: methods, countermeasures, and their measurement. We relied on existing research to develop pilot coding and then finalized it as a typology to classify studies according to the types of countermeasures proposed [31], [32]. Following the pilot coding exercise, we compared our coding experiences and adopted the final coding template. Intercoder reliability measures were calculated based on 50 randomly selected publications (see Table S2 for Krippendorff's α measures and agreement). The publications were randomly divided among the five researchers. Pairs of researchers compared their article coding and resolved any discrepancies through conversations with the broader research team.

Meta-Analysis

We used data collected by following the steps described above to select studies for meta-analysis. We focused on two types of countermeasures that were frequently empirically tested in the studies we analyzed in *Synthesis Report* (2023.1): content labeling and corrective information interventions. We then selected only RCT studies and non-randomized (quasi-experimental designs) studies. RCT studies conduct experiments that randomly assign an individual to receive a countermeasure or a control measure and put restrictions in place, so that certain variables do not impact the outcome. Quasi-experimental studies lack random assignment, and groups are assigned to a countermeasure using self-selection or administrator selection, or both. All the studies we considered at this stage focused on the effect of corrective information or content labeling on the perception of information. Although the definitions of the dependent variable differ slightly between studies, they all aim to operationalize the perception of information before and after the experimental process as such a variable, with a focus on aspects of information perception such as accuracy, believability, credibility, and trustworthiness. Nevertheless, these slight differences in variable operationalization increased the external validity of the meta-analysis in terms of thematic application. We also reviewed other types of countermeasures and dependent variables, such as news literacy materials and content-sharing behavior; however, they produced insufficient evidence.

To empirically examine whether the estimated effects reported in the selected studies were simply random errors due to study differences or consistent and systematic patterns across different studies, a pair of researchers read the identified publications in full and extracted relevant data on the sample and effect sizes reported. One of the paired researchers extracted data from publications, and the other researcher verified 15% of the data extraction. Where possible, we extracted an effect size or calculated it when the relevant information was available. When standard errors were not available, we calculated the standard error from an exact *P* value. In a few cases, when the *P* value was reported as $P < p$, we assumed that $P = p$. For example, when an exact *p* value was available (e.g., $p = 0.015$), we used the exact value (e.g., 0.015) to calculate the standard error. In some cases, publications used the same samples for different experiments that used the same modeling approach but reported different effect sizes. We chose the study

that conveyed the more conservative effect size in these cases. Most studies we analyzed expected a positive effect after introducing content labels and corrective information. We also coded a few additional variables to analyze the sample characteristics, intervention context, and other potentially relevant differences between the studies (Table S1). To aid the reader's interpretation of the results, we have converted all effects for content labeling and corrective information studies to be in the same direction, so that positive effects indicate better information perception outcomes.

We were unable to include those studies that did not report effect sizes and standard errors in line with recommended practices [33]. These included 21 labeling studies that did not provide enough data to calculate effect size. In addition, we removed 23 content labeling and ten corrective information studies in which the *P* value was reported as $P > p$.

This approach to meta-analysis sampling resulted in 19 labeling studies and 24 corrective information studies analyzed from 18 publications in total. Cochrane guidelines suggest that these numbers of studies are sufficient for meta-analysis. In fact, two studies are enough if their results can be meaningfully pulled together [34]. We calculated the pooled effect using the random-effects size model as this includes consideration of heterogeneity in the effect estimate [19]. We also used Knapp-Hartung adjustments to calculate the confidence interval around the pooled effect [35].

SECTION S2. Validity Assessment: Risk of Bias Analysis

We screened the studies included in the meta-analysis for risk of bias by confirming the internal validity of the publications we pooled for meta-analysis—the extent to which they are free from bias [36]—using two protocols recommended by Cochrane: the Risk of Bias 2 (RoB 2) protocol for RCTs [37] and the Risk of Bias In Non-randomized Studies of Interventions (ROBINS-I) [38]. The same researchers who performed other coding exercises in our study assessed each eligible publication and analyzed evidence related to risk of bias and validity arising from the randomization process: deviations from the intended interventions (effect of adhering to interventions), missing data, diverse measurement, and selective reporting, as indicated in the RoB 2 protocol. For quasi-experiments, the researchers assessed risk of bias that might have arisen due to confounding variables, participant selection, missing data, diverse measurement, and selective reporting, as indicated in the ROBINS-I protocol.

The results of the RoB 2 assessment showed that most studies in our sample contained a moderate risk of bias or issues with internal validity. Those studies which raised a high number of concerns were removed from the analysis. Applying the ROBINS-I protocol did not reveal serious concerns, though some moderate concerns were recorded. A “low” overall score for ROBINS-I means the study is comparable to an RCT. Typically, it is hard for a study to receive a low overall bias score due to the participants in the study not being truly randomized. One recent systematic review of social science literature observed a risk of bias in almost every RCT study it analyzed [39].

The studies in our sample performed relatively well, scoring “low” to “moderate” for bias in most domains. Areas of our concern were consistent with similar assessments. For example, it was reported that missing data and measurement are the two most challenging areas of RoB assessment [36]. These areas raised most concerns in out assessment as well. One publication in our sample contained a high risk of bias because its design deviated from the intended interventions. Another did have critical bias and was highly likely to have uncontrolled confounds. Overall, we removed three publication due to a critical or high risk of bias, using approaches previously applied in similar conditions [39]. The results of the ROBINS-I and RoB 2 assessments for content labeling and

corrective information studies are presented in Tables 4 and 5. These assessment tools are developed in the context of health research and might not always be optimal for social science publications, and they explicitly incorporate assessors' judgment [36]

Table 4. Risk of Bias (Using ROBINS-I) of Quasi-Experimental Publication.

Study	Confoun-ding	Selection	Classificati on	Deviations from the Intended Interventio ns	Missing data	Measurem ent	Selective Reporting	Overall
Bode and Vraga, 2015	Low	Low	Low	Low	Moderate	Low	Moderate	Moderate
Ecker, O'Reilly, Reid, and Chang, 2020	Moderate	Low	Low	Low	Moderate	Moderate	Low	Moderate
Kim and Dennis, 2019	Moderate	Low	Low	Low	Low	Low	Low	Moderate
Kim, Moravec, and Dennis, 2019	Low	Low	Moderate	Low	Low	Moderate	Moderate	Moderate
Lanius, Weber, and Mackenzie, 2021	Critical	Low	Moderate	Low	Moderate	Moderate	Low	Critical
Lee, Kim, and Lee, 2022	Low	Low	Low	Low	Low	Low	Low	Low
Moravec, Kim, and Dennis, 2020	Moderate	Low	Low	Low	Low	Moderate	Low	Moderate
Pennycook, Cannon, and Rand, 2018	Low	Low	Low	Low	Moderate	Low	Low	Moderate
Song, Wang and Xu, 2022	Moderate	Low	Low	Moderate	Low	Low	Low	Moderate
Sullivan, 2019	Moderate	Low	Low	Moderate	Low	Moderate	Low	Moderate
Yang and Overton, 2022	Low	Low	Low	Moderate	Low	Moderate	Low	Moderate

Note: Content labeling and corrective information studies are grouped based on the type of an experiment they conducted; scale is Low / Moderate / Serious / Critical.
Source: IPIE calculations based on data collected.

Table 5. Risk of Bias (Using RoB 2) of Publications Employing Randomized Control Trials.

Study	Randomization	Deviations from the Intended Interventions	Missing Data	Measurement	Selective Reporting	Overall
Chung and Kim, 2021	Low	Some concerns	Low	Some concerns	Low	Some concerns
Clayton et al., 2019	Low	Low	Low	Low	Low	Low
Hameleers, Powell, Van, and Bos, 2020	Low	Some concerns	Low	Low	Some concerns	Some concerns
Kirchner and Reuter, 2020	Low	Some concerns	Low	Low	Low	Some concerns
Vraga and Bode, 2017	Low	Low	Some concerns	Low	High	High
Vraga and Bode, 2021	Low	Low	Low	Some concerns	Low	Some concerns
Vraga, Bode, and Tully, 2020	Low	Low	Low	High	Some concerns	High
Vraga, Bode, and Tully, 2021	Low	Some concerns	Low	Low	Some concerns	Some concerns
Vraga, Kim, Cook, and Bode, 2020	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Vraga, Tully and Bode, 2022	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns

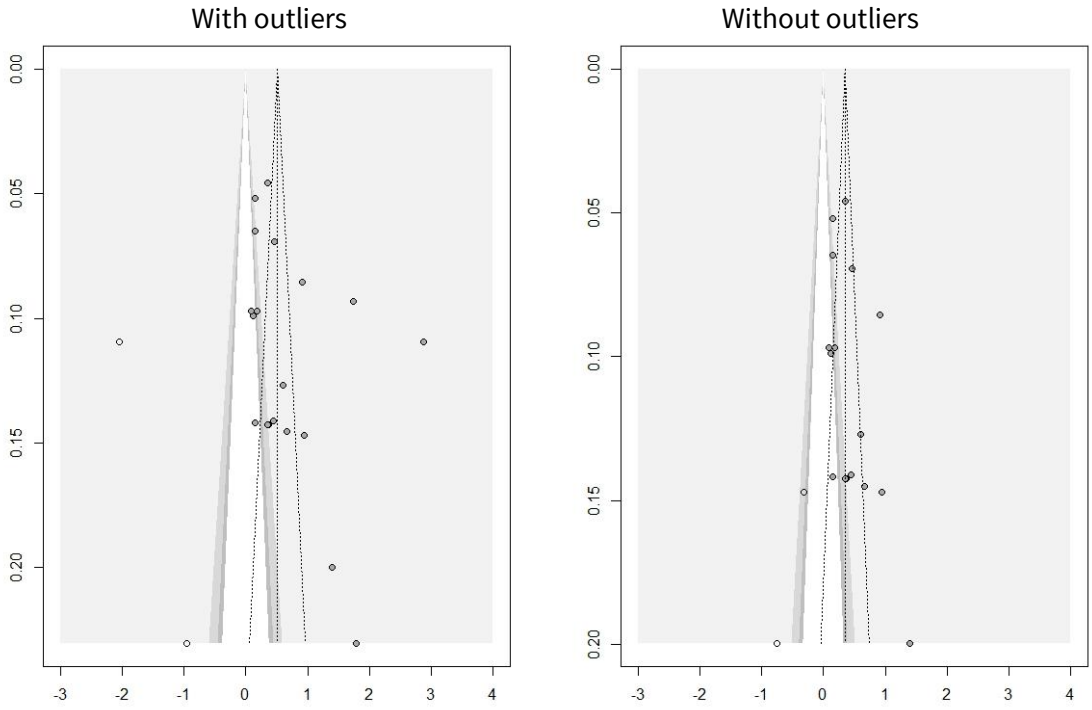
Note: Content labeling and corrective information studies are grouped based on the type of an experiment they conducted; scale is Low / Some concerns / High.
Source: IPIE calculations based on data collected.

In addition to the RoB 2 and ROBINS-I protocols, we built funnel plots for each meta-analysis. These plots show the distribution of reported estimates illustrating the extent of variation in these estimates and help to assess publication selection bias [40]. Unfortunately, social sciences publications often treat statistically significant results more favorably. As a result, larger, more significant effects are often overrepresented in peer-reviewed literature [40]. Publication selection bias may imply funnel plot asymmetry: its shape will be different from the approximate shape of an inverted funnel (Figures 4 and 5).

The plots for content labeling show a larger variation in the standard error (Y-axis). This is likely due to the wide range of the sample sizes in the publications. Egger's test for asymmetry showed the absence of significant asymmetry ($p = 0.11$), which suggests that plots are closer to symmetrical and there is no significant evidence of publication bias or small-study effects from this particular test, though the original funnel plot was skewed to the right. The trim-and-fill technique added two new measures, signaling that the pooled effect in our meta-analysis of content labeling could have been overestimated due to small-study effects.

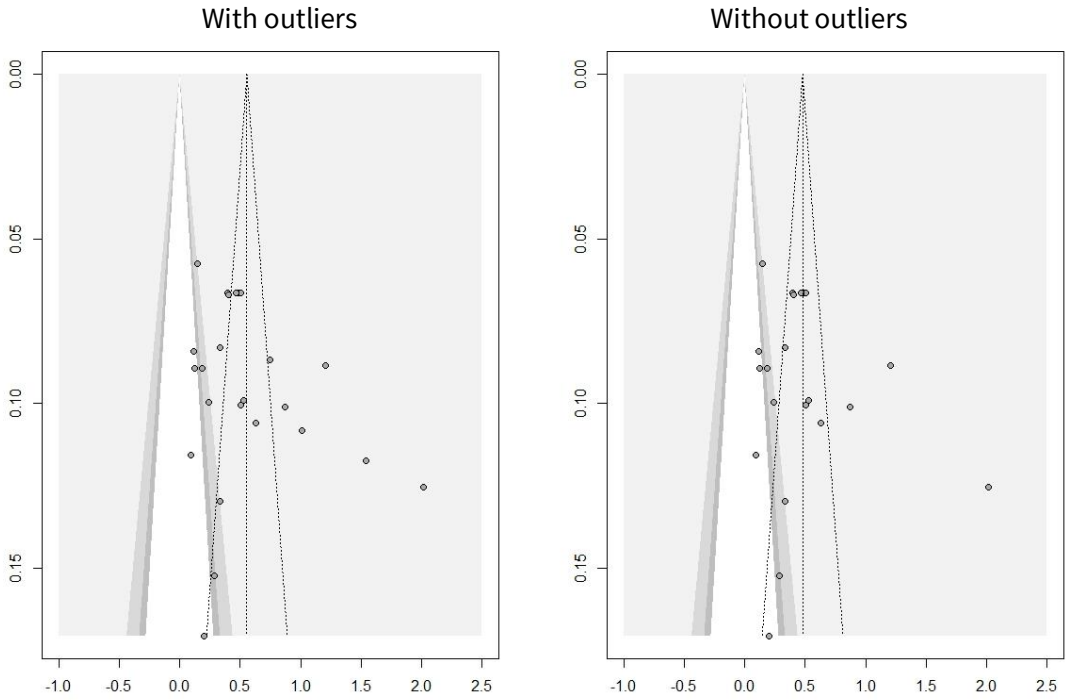
The funnel plots for the studies focusing on correction information materials show a potential bias in the distribution along the X-axis, with the funnel skewed toward the left. However, the trim-and-fill technique did not detect any bias, while Egger's test for asymmetry ($p = 0.12$) for these studies was not significant. Both results indicate that there may not be publication bias detected through this specific technique. However, we observed a clustering on the Y-axis, and a few studies in this category were very close to the significance threshold, suggesting a potential underrepresentation of studies with stronger effects, another indication pointing to potential publication bias. Hence, we found no definitive evidence of publication bias in either of the two cases.

Figure 4. Funnel Plots to Assess Publication Bias, Content Labeling Studies.



Note: Figures represent Duval and Tweedie's trim-and-fill method funnel plots. In shaded regions, darker areas represent $p < 0.1$, and lighter areas represent $p < 0.01$. The X-axis refers to the standardized mean difference and the Y-axis refers to the standard error.
Source: IPIE calculations based on data collected.

Figure 5. Funnel Plots to Assess Publication Bias, Corrective Information Studies.



Note: Figures represent Duval and Tweedie's trim-and-fill method funnel plots. In shaded regions, darker areas represent $p < 0.1$, and lighter areas represent $p < 0.01$. The X-axis refers to the standardized mean difference and the Y-axis refers to the standard error.
Source: IPIE calculations based on data collected.

SECTION S3. Additional Limitations

We narrowed the sample to publications that satisfied the exacting criteria for inclusion in the meta-analysis sample. It is possible that some qualifying studies were excluded from the sample even though we undertook an extensive search that included two large social science databases and numerous search terms. Other more recently available databases, like Microsoft Academic and Dimensions, could provide additional sources for even more comprehensive analysis. We could have considered additional sources and so-called gray literature by using databases like Google Scholar to add more literature and, if resources allowed, to reduce the chance of file-drawer bias. The update to this work should include publications in other formats, such as peer-reviewed reports and books.

Reviews of this type are limited by language, especially as the major databases primarily contain research published in English. This limits the scope of the evidence and skews it away from research published in places where English is not the main language of academic inquiry. Despite our aim that the reviewed publications should be from diverse countries where their authors are based, our review remains Western-centric, with USA-originating scholarship dominating the dataset. Focusing on the literature in other languages is a natural next step in our work. Because the literature is rapidly expanding, we hope we will be able to test a greater variety of countermeasures when more publications provide empirical evidence about their effectiveness.

SECTION S4. List of Publications Reviewed with Meta-Analysis

[1] L. Bode and E. K. Vraga, 'In Related News, That was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media', *J. Commun.*, vol. 65, no. 4, pp. 619–638, Aug. 2015, doi: 10.1111/jcom.12166.

[2] M. Chung and N. Kim, 'When I Learn the News is False: How Fact-Checking Information Stems the Spread of Fake News Via Third-Person Perception', *Hum. Commun. Res.*, vol. 47, no. 1, pp. 1–24, Jan. 2021, doi: 10.1093/hcr/hqaa010.

[3] K. Clayton *et al.*, 'Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media', *Polit. Behav.*, vol. 42, no. 4, pp. 1073–1095, Dec. 2020, doi: 10.1007/s11109-019-09533-0.

[4] U. K. H. Ecker, Z. O'Reilly, J. S. Reid, and E. P. Chang, 'The Effectiveness of Short-Format Refutational Fact-Checks', *Br. J. Psychol.*, vol. 111, no. 1, pp. 36–54, 2020, doi: 10.1111/bjop.12383.

[5] M. Hameleers, T. E. Powell, T. G. L. A. V. D. Meer, and L. Bos, 'A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media', *Polit. Commun.*, vol. 37, no. 2, pp. 281–301, Mar. 2020, doi: 10.1080/10584609.2019.1674979.

[6] A. Kim, P. L. Moravec, and A. R. Dennis, 'Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings', *J. Manag. Inf. Syst.*, vol. 36, no. 3, pp. 931–968, Jul. 2019, doi: 10.1080/07421222.2019.1628921.

[7] A. Kim and A. R. Dennis, 'Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media', *MIS Q.*, vol. 43, no. 3, pp. 1025–1039, Jan. 2019, doi: 10.25300/misq/2019/15188.

[8] J. Kirchner and C. Reuter, 'Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness', *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW2, pp. 1–27, Oct. 2020, doi: 10.1145/3415211.

[9] C. Lanius, R. Weber, and W. I. MacKenzie, 'Use of Bot and Content Flags to Limit the Spread of Misinformation Among Social Networks: A Behavior and Attitude Survey', *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 32, Mar. 2021, doi: 10.1007/s13278-021-00739-x.

[10] J. Lee, J. W. Kim, and H. Yun Lee, 'Unlocking Conspiracy Belief Systems: How Fact-Checking Label on Twitter Counters Conspiratorial MMR Vaccine Misinformation', *Health Commun.*, vol. 0, no. 0, pp. 1–13, Jan. 2022, doi: 10.1080/10410236.2022.2031452.

[11] L. Lutzke, C. Drummond, P. Slovic, and J. Árvai, 'Priming Critical Thinking: Simple Interventions Limit the Influence of Fake News About Climate Change on Facebook', *Glob. Environ. Change*, vol. 58, p. 101964, Sep. 2019, doi: 10.1016/j.gloenvcha.2019.101964.

[12] P. L. Moravec, A. Kim, and A. R. Dennis, 'Appealing to Sense and Sensibility: System 1 and System 2 Interventions for Fake News on Social Media', *Inf. Syst. Res.*, vol. 31, no. 3, pp. 987–1006, Sep. 2020, doi: 10.1287/isre.2020.0927.

[13] E. Nekmat, 'Nudge Effect of Fact-Check Alerts: Source Influence and Media Skepticism on Sharing of News Misinformation in Social Media', *Soc. Media Mathplus Soc.*, vol. 6, no. 1, p. 205630511989732, Jan. 2020, doi: 10.1177/2056305119897322.

[14] G. Pennycook, T. D. Cannon, and D. G. Rand, 'Prior exposure increases perceived accuracy of fake news.', *J. Exp. Psychol. Gen.*, vol. 147, no. 12, pp. 1865–1880, Dec. 2018, doi: 10.1037/xge0000465.

[15] L. Soetekouw and S. Angelopoulos, 'Digital Resilience Through Training Protocols: Learning To Identify Fake News On Social Media', *Inf. Syst. Front.*, Jan. 2022, doi: 10.1007/s10796-021-10240-7.

[16] Y. Song, S. Wang, and Q. Xu, 'Fighting Misinformation on Social Media: Effects of Evidence Type and Presentation Mode', *Health Educ. Res.*, vol. 37, no. 3, pp. 185–198, Jun. 2022, doi: 10.1093/her/cyac011.

[17] M. C. Sullivan, 'Leveraging Library Trust to Combat Misinformation on Social Media', *Libr. Inf. Sci. Res.*, vol. 41, no. 1, pp. 2–10, Jan. 2019, doi: 10.1016/j.lisr.2019.02.004.

[18] M. Tully, E. K. Vraga, and L. Bode, 'Designing and Testing News Literacy Messages for Social Media', *Mass Commun. Soc.*, vol. 23, no. 1, pp. 22–46, May 2019, doi: 10.1080/15205436.2019.1604970.

[19] E. Vraga, M. Tully, and L. Bode, 'Assessing the Relative Merits of News Literacy and Corrections in Responding to Misinformation on Twitter', *New Media Soc.*, vol. 24, no. 10, pp. 2354–2371, Oct. 2022, doi: 10.1177/1461444821998691.

[20] E. K. Vraga, L. Bode, and M. Tully, 'Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter', *Commun. Res.*, p. 0093650219898094, Jan. 2020, doi: 10.1177/0093650219898094.

[21] E. K. Vraga, S. C. Kim, J. Cook, and L. Bode, 'Testing the Effectiveness of Correction Placement and Type on Instagram', *Int. J. Press.*, vol. 25, no. 4, pp. 632–652, Oct. 2020, doi: 10.1177/1940161220919082.

[22] E. K. Vraga, L. Bode, and M. Tully, 'The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Related to Sunscreen and Skin Cancer', *Health Commun.*, pp. 1–9, Apr. 2021, doi: 10.1080/10410236.2021.1910165.

[23] E. K. Vraga and L. Bode, 'Addressing COVID-19 Misinformation on Social Media Preemptively and Responsively', *Emerg. Infect. Dis.*, vol. 27, no. 2, pp. 396–403, Feb. 2021, doi: 10.3201/eid2702.203139.

[24] E. K. Vraga and L. Bode, 'Using Expert Sources to Correct Health Misinformation in Social Media', *Sci. Commun.*, vol. 39, no. 5, pp. 621–645, Oct. 2017, doi: 10.1177/1075547017731776.

[25] F. Yang, 'What If Unmotivated Is More Dangerous? The Motivation-Contingent Effectiveness of Misinformation Correction on Social Media', 2022.

SECTION S5. Supplementary Tables

Table S1. Codes Used to Organize and Synthesize Research Evidence.

Code	Description of the code	Variables
Eligibility		
English	Publication is available in English.	0 = no 1 = yes
Misinformation	Publication discusses any aspects of misinformation, disinformation, fake news, propaganda, rumors, "credibility" of information or digital/automated manipulation as a key object of study. If there are several objects, misinformation should be at least one out of three of them.	0 = no 1 = yes
Digital platforms	Social media platforms (websites that are dependent primarily on user-generated content that facilitates two-way interaction) constitute a significant focus of the research. The study names any social media platform as a major study object or its key background, field, or context. Forums are not considered as social media.	0 = no 1 = yes
Empirical	A study is based on empirical evidence that is analyzed in the publication, which should be clear from Abstract.	0 = no 1 = yes
Full paper coding		
Methods	Methods used to collect and analyze data. Multiple choice.	0 = none 1 = survey 2 = interviews 3 = focus groups 4 = ethnography 5 = experiment 6 = content analysis (manual) 7 = content analysis (automated) involving social media data extraction 8 = network analysis 9 = agent-based modeling, simulation 10 = process tracing and (or) case study 100 = other

Measures proposed	Countermeasures mitigating the impact of misinformation that are related to the functioning of platforms proposed in the study. The proposed measures should mitigate the impact of misinformation that spreads on digital platforms, according to this study. Multiple choice.	NA = nothing proposed 0 = broad 1 = advertisement policy 2 = labeling 3 = content or account moderation 4 = content reporting 5 = content distribution & sharing 6 = corrective information materials 7 = disinformation disclosure 8 = information literacy 9 = redirection 10 = security or verification 11 = other 12 = fact-checking infrastructure
Measures proposed specify	Shortly describe specific countermeasures proposed to address the problem of misinformation.	[open coding]
Link between proposed countermeasure and study object(s)	Is there a link between the proposed countermeasures and the object of the study? The countermeasures should be linked to and derived from the research presented in the paper.	0 = no 1 = yes
Proposed countermeasures target any human misinformation creators	Do the proposed countermeasures actively target any human creators/sources of misinformation content?	0 = no 1 = yes
Meta-analysis codes		
Study type	What type of study design?	0 = RCT 1 = quasi-experimental design
Platform	What social media platform does this study design focus on?	0 = general (overall exposure to/no specific platform) 1 = Facebook 2 = Twitter 3 = blog 4 = YouTube 5 = other
Sample type	Who was recruited for the study?	1 = college students 2 = patients 3 = general adult sample 4 = mixed community and college sample 5 = other: children, adolescents, etc.

Geographic context	Where data collection was conducted	1 = USA (national) 2 = Canada 3 = Europe 4 = Israel 5 = East Asia (e.g., China) 6 = South Asia (e.g., India) 7 = Oceania 8 = Africa 9 = Central/South America 10 = MTurk unspecified 11 = other
Label/correction source	Who/what was the source of the correction/label?	0 = no Source Provided 1 = peer (friends, other social media users, independent users etc.) 2 = experts (health professions, government agency or intergovernmental bodies, university, related nonprofit) 3 = news agency (CNN, BBC, Reuters) 4 = for-profit commercial company (platform itself, social media tag) 5 = other
Label/correction format	How was the correction/label formatted?	0 = no particular media context 1 = text (blog, post) 2 = text and image 3 = audiovisual passive (television, movie, YouTube clip) 4 = other
Participants' issue involvement	How involved were the participants in the issue/ were participants explicitly recruited due to their relationship to the issue?	1 = low (participants are not recruited due to their relationship to the issue) 2 = high (participants are explicitly recruited due to their relationship to the issue)
Misinformation format	How was the misinformation formatted?	0 = no particular media context 1 = text (blog, post) 2 = text and image 3 = audiovisual passive (television, movie, YouTube clip) 4 = other
Issue type	What type of issue was the misinformation?	1 = health 2 = political 3 = social (non-political) 4 = other

Note: Codes are ordered as they appeared in the codebook provided to coders. For replication purposes, a version with examples is available upon request. Table 1 presents the summary of definitions and examples used to code the Measures proposed variable. Codes for meta-analysis studies were partially adapted from [41]
Source: IPIE.

Table S2. Inter-coder Reliability Assessment (Krippendorff's α and Percentage Agreement).

Category	α	%
Eligibility		
Digital platforms	0.92	98.00
Empirical	0.88	93.90
English	1.00	100.00
Misinformation	0.76	87.80
Methods and countermeasures		
Link between proposed countermeasures and study object(s)	0.72	81.80
Methods	0.74	77.30
Proposed countermeasures target human misinformation creators	0.83	97.70

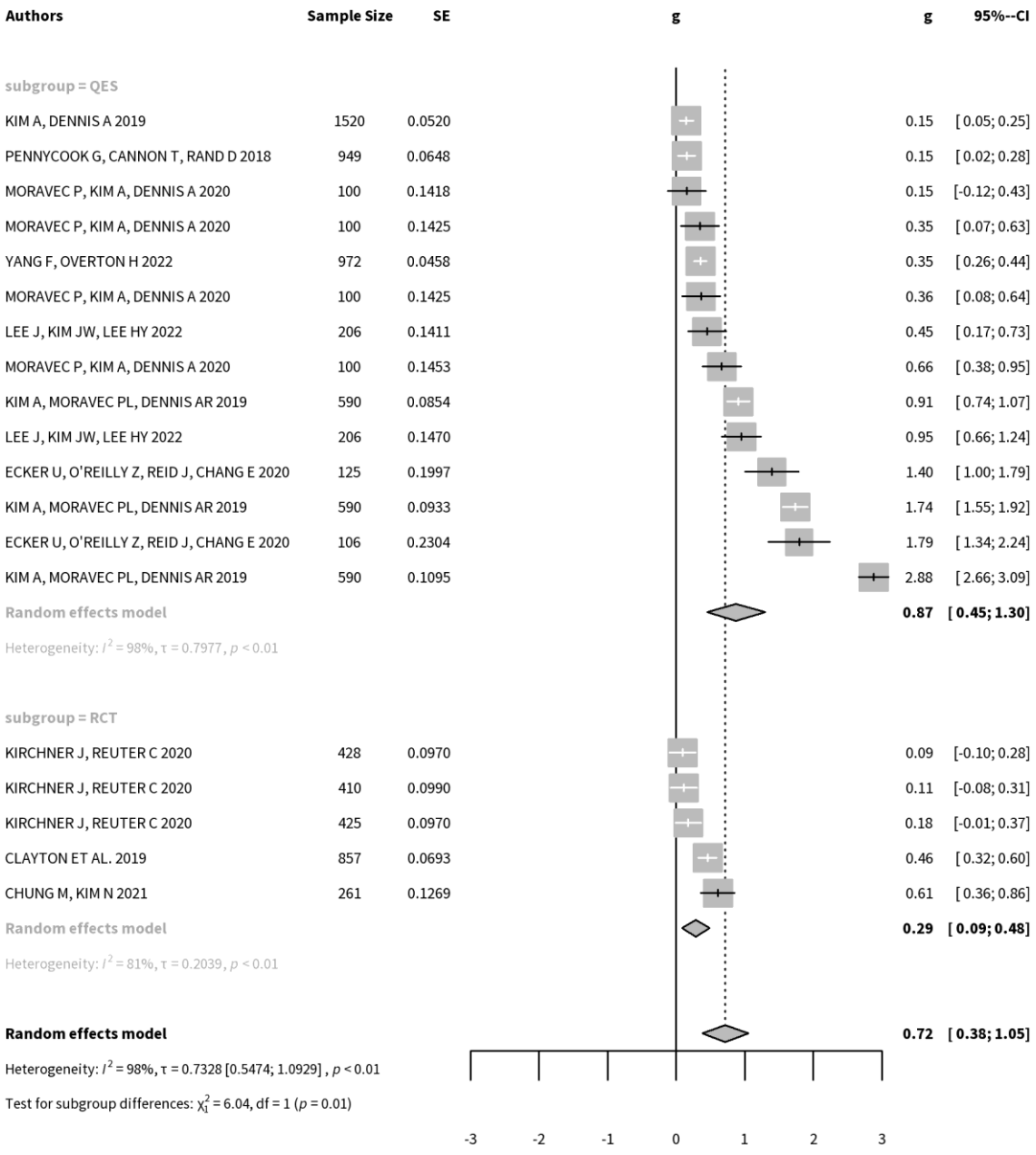
Source: IPIE calculations based on data collected.

Table S3. Boolean Search Term in the Web of Science 2021 Syntax.

AB=(disinformation OR misinformation OR propaganda OR "fake news" OR rumo OR "misleading information" OR "false information" OR "computational propaganda") OR TI=(disinformation OR misinformation OR propaganda OR "fake news" OR rumo* OR "misleading information" OR "false information" OR "computational propaganda") OR KP=(disinformation OR misinformation OR propaganda OR "fake news" OR rumo* OR "misleading information" OR "false information" OR "computational propaganda")*
AND
(AB=("social media" OR "social networking site" OR "digital platform") OR TI=("social media" OR "social networking site" OR "digital platform*") OR KP=("social media" OR "social networking site" OR "digital platform*")*
AND
(LA==("ENGLISH") AND DT==("ARTICLE" OR "EARLY ACCESS"))

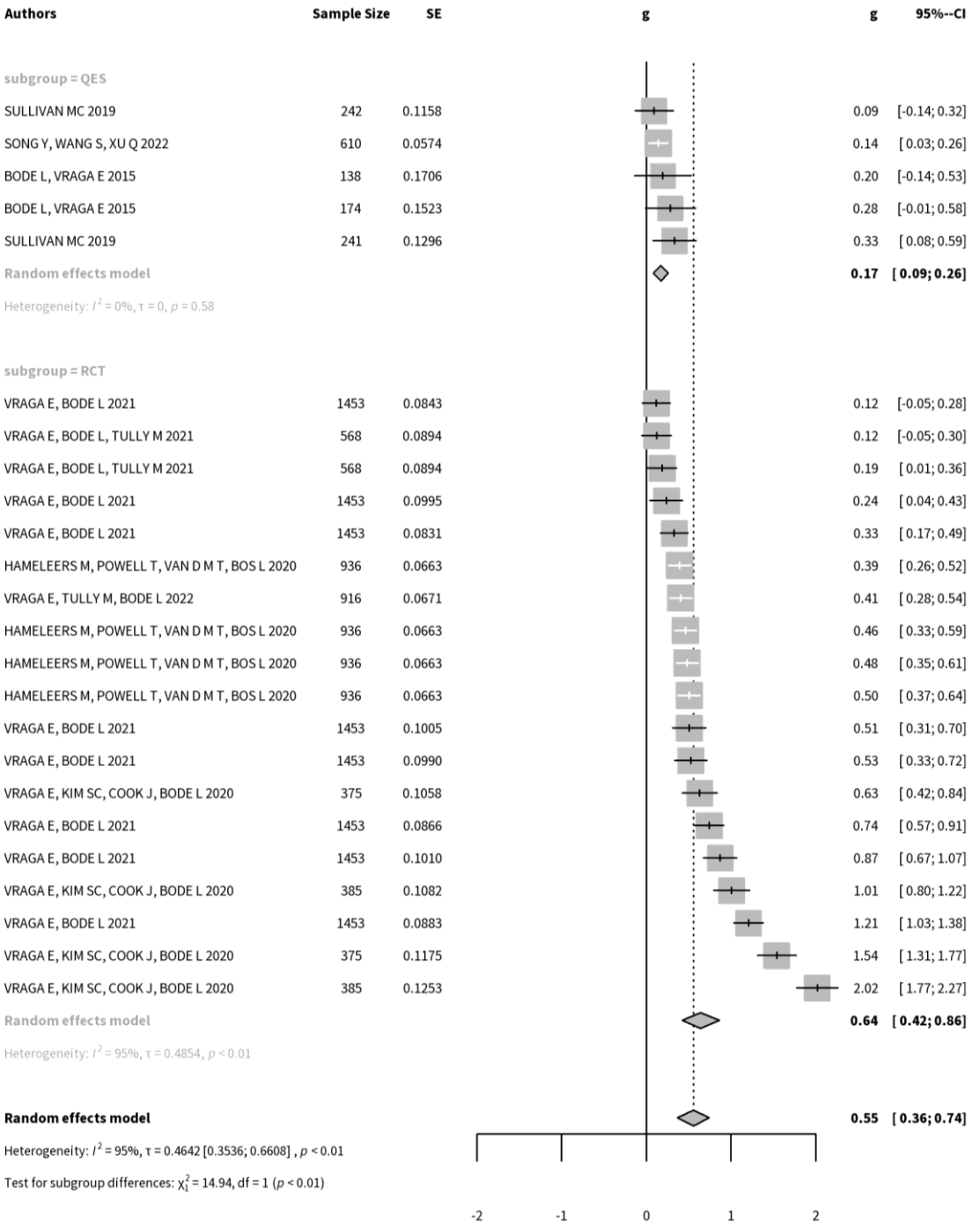
SECTION S6. Supplementary Figures

Figure S1a. Meta-Analysis of Content Labeling Studies Grouped by Randomization Type (RCT Or Quasi-Experimental Studies (QES)).



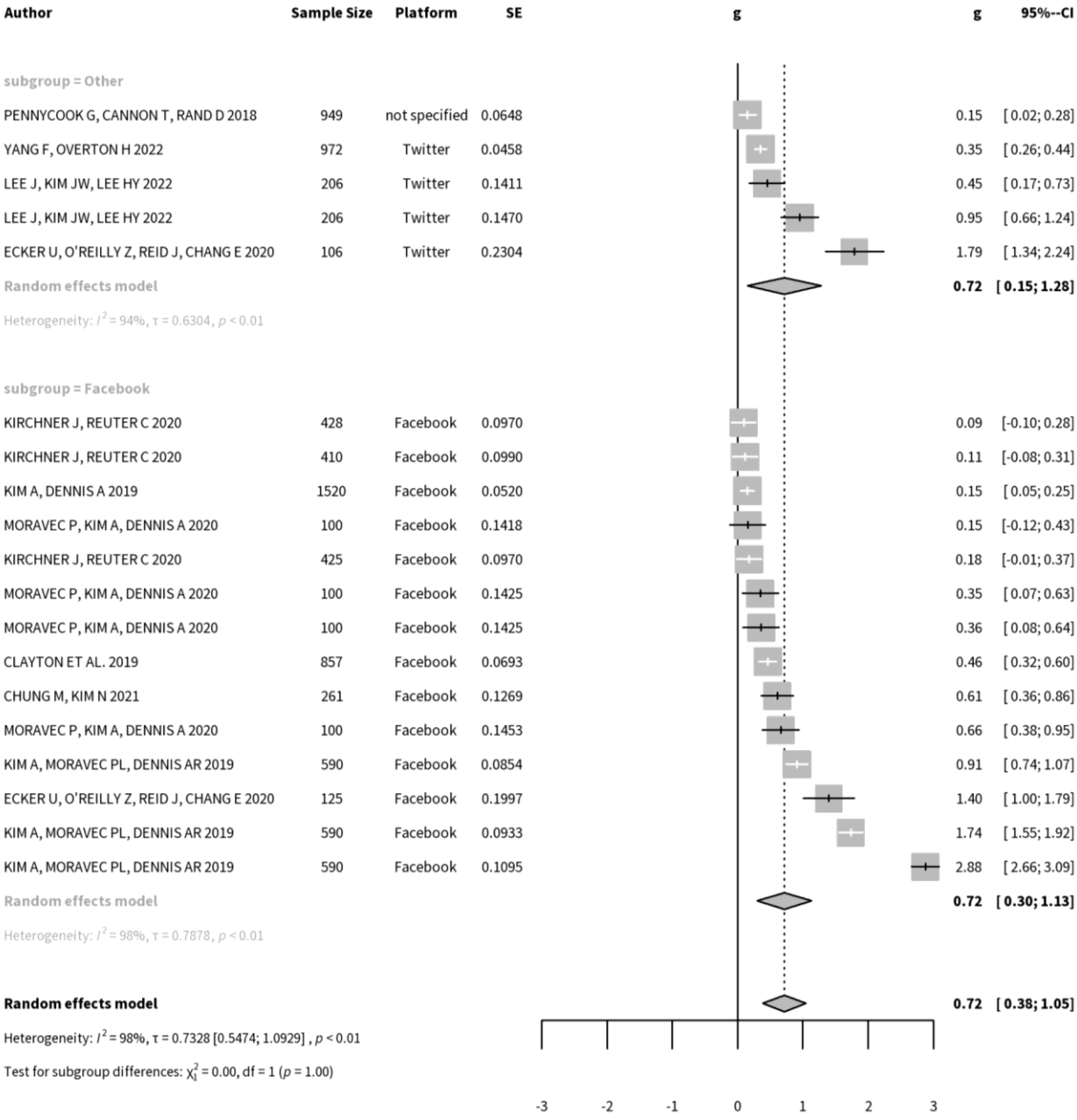
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S1b. Meta-Analysis of Corrective Information Studies Grouped by Randomization Type (RCT Or Quasi-Experimental Studies (QES)).



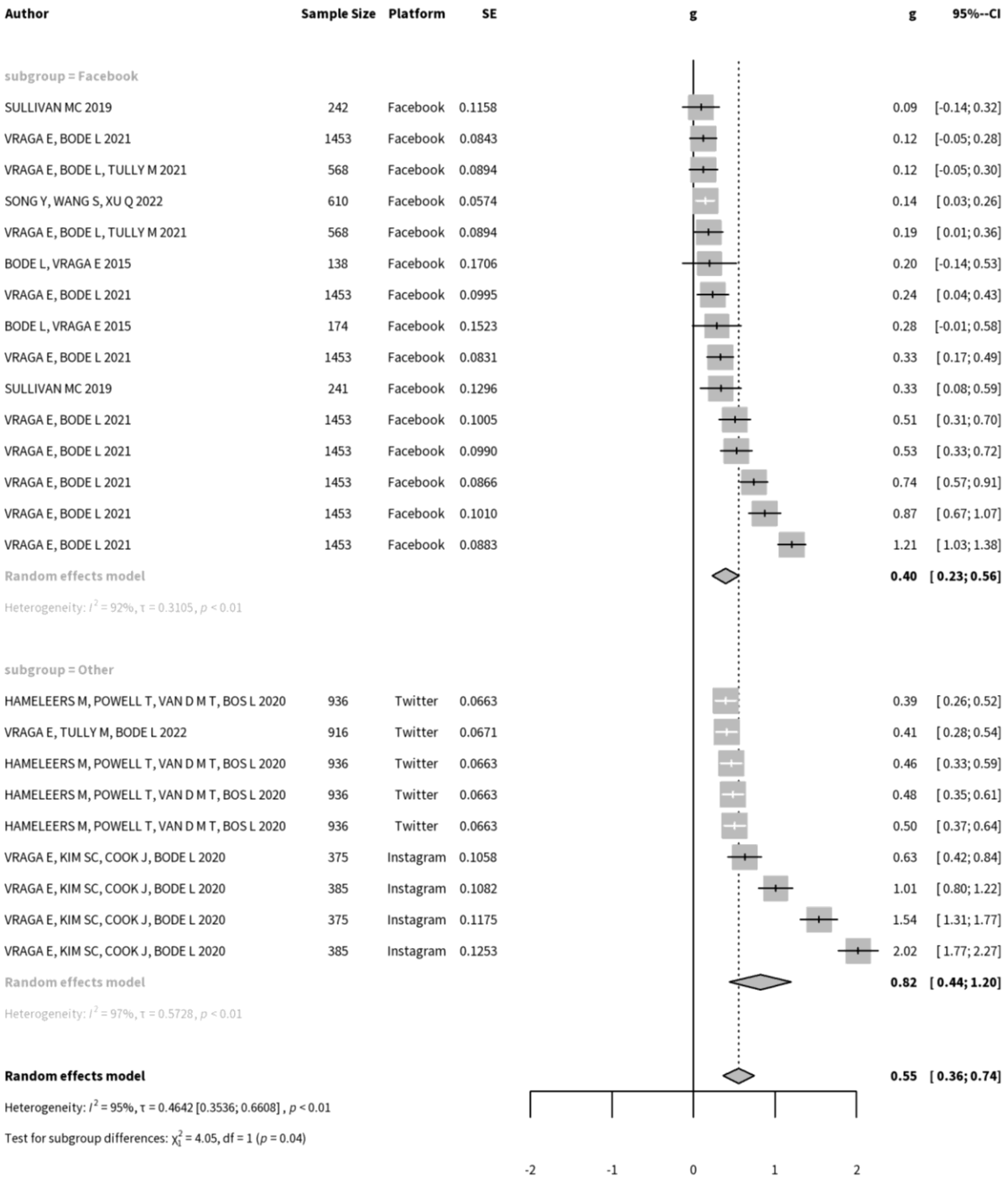
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S2a. Meta-Analysis of Content Labeling Studies by Platform.



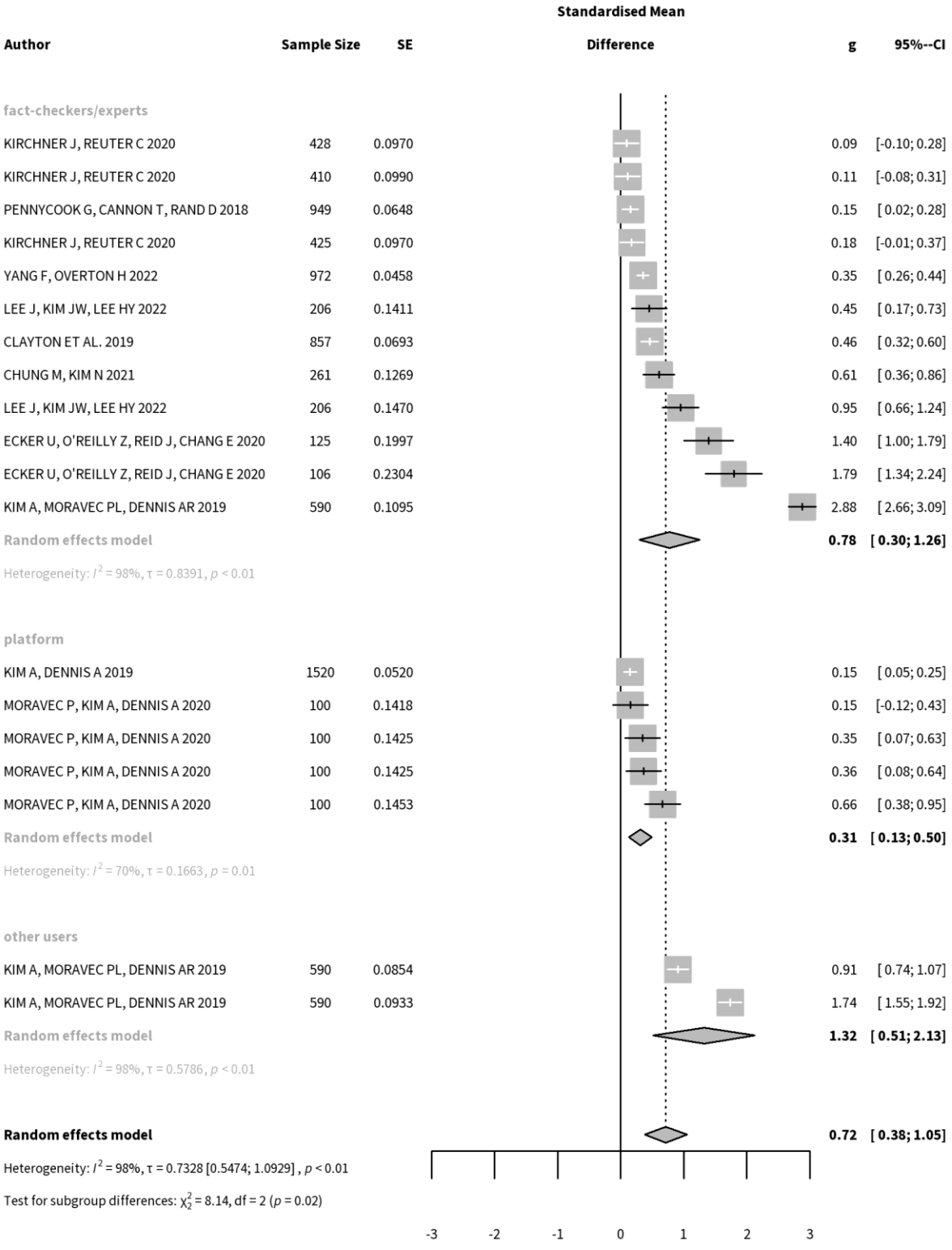
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S2b. Meta-Analysis of Corrective Information Studies by Platform.



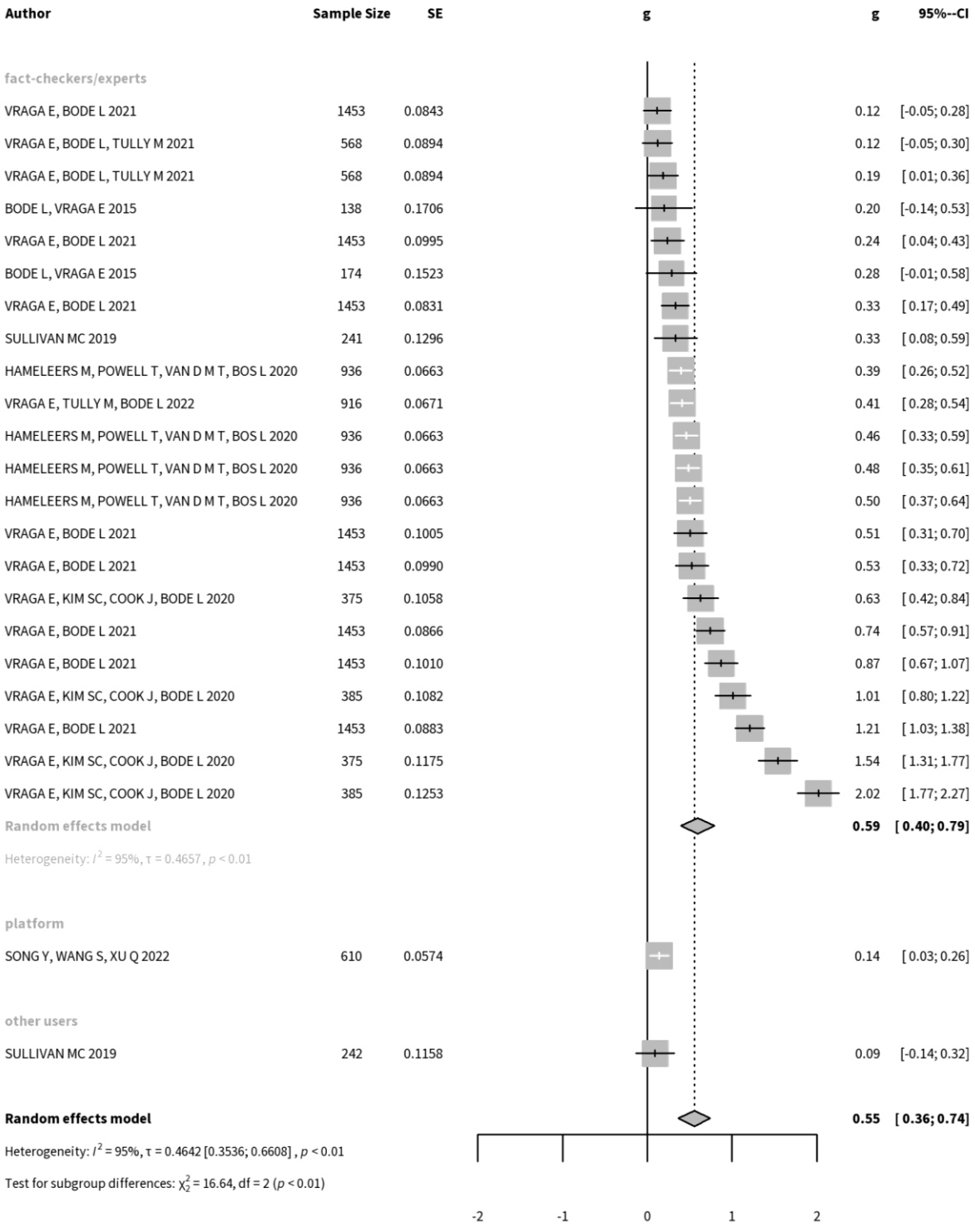
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S3a. Meta-Analysis of Content Labeling Studies by Label Source.



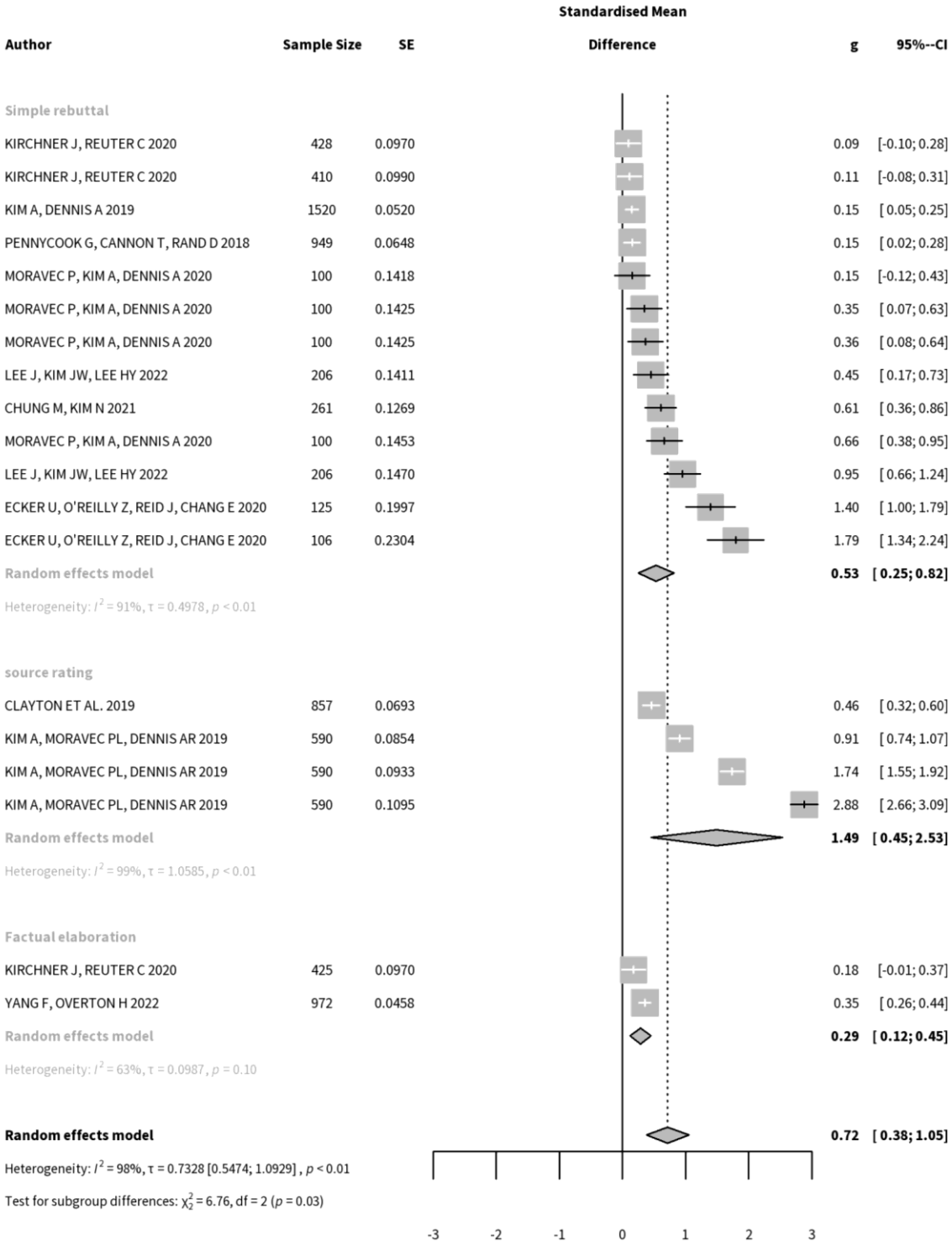
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S3b. Meta-Analysis of Corrective Information Studies by Correction Source.



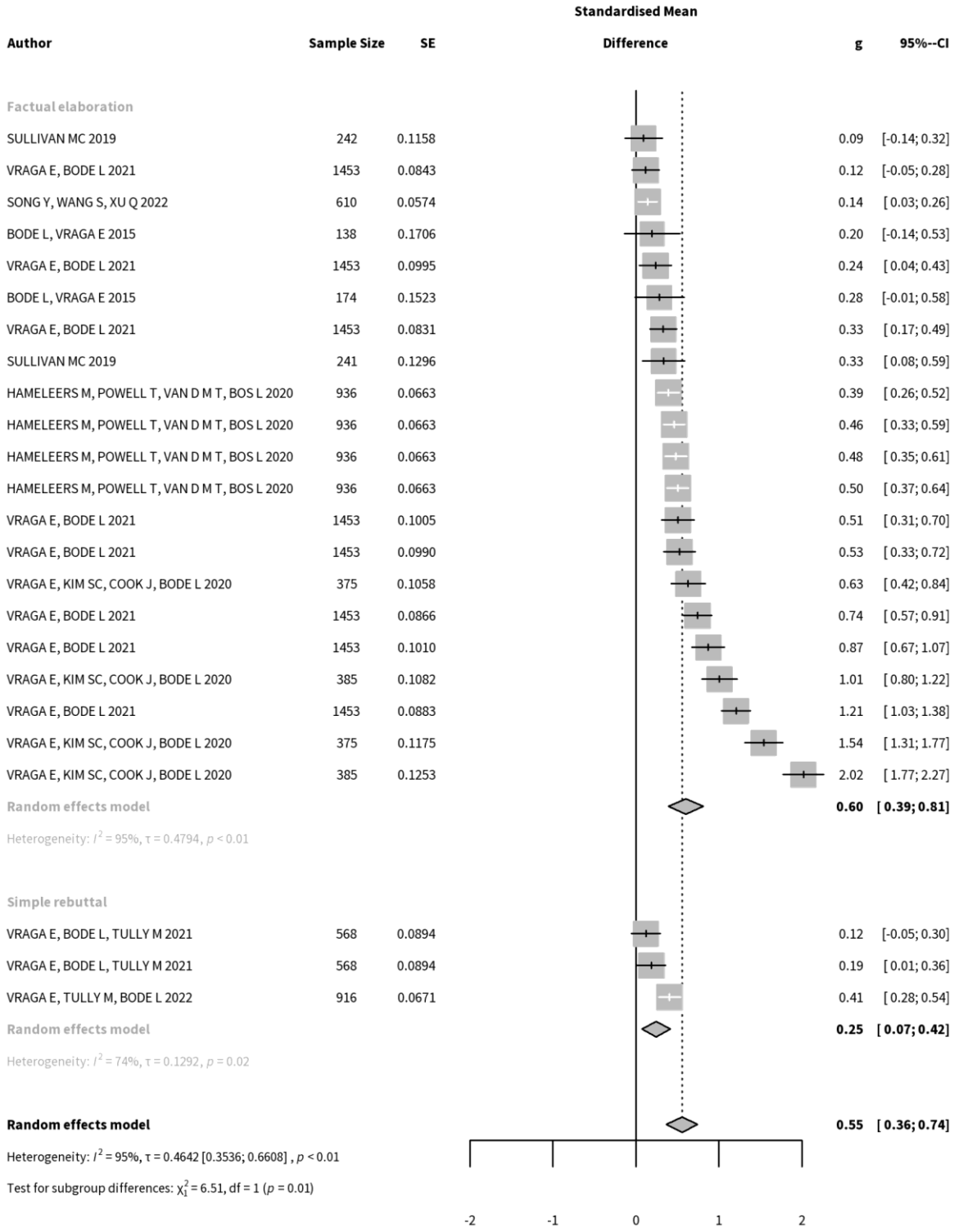
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S4a. Meta-Analysis of Content Labeling Studies by Label Elaboration Level.



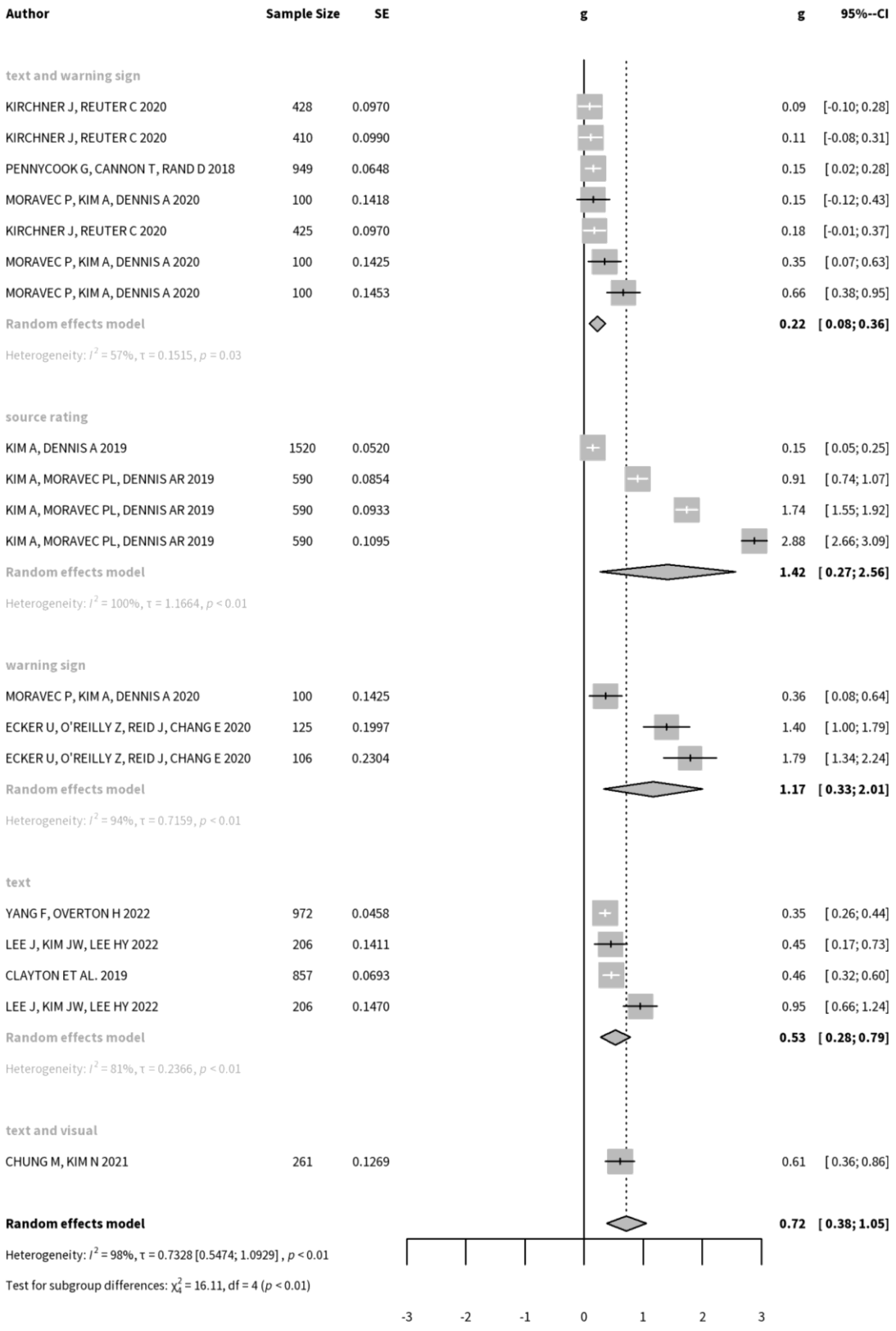
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S4b. Meta-Analysis of Corrective Information Studies by Information Material Elaboration Level.



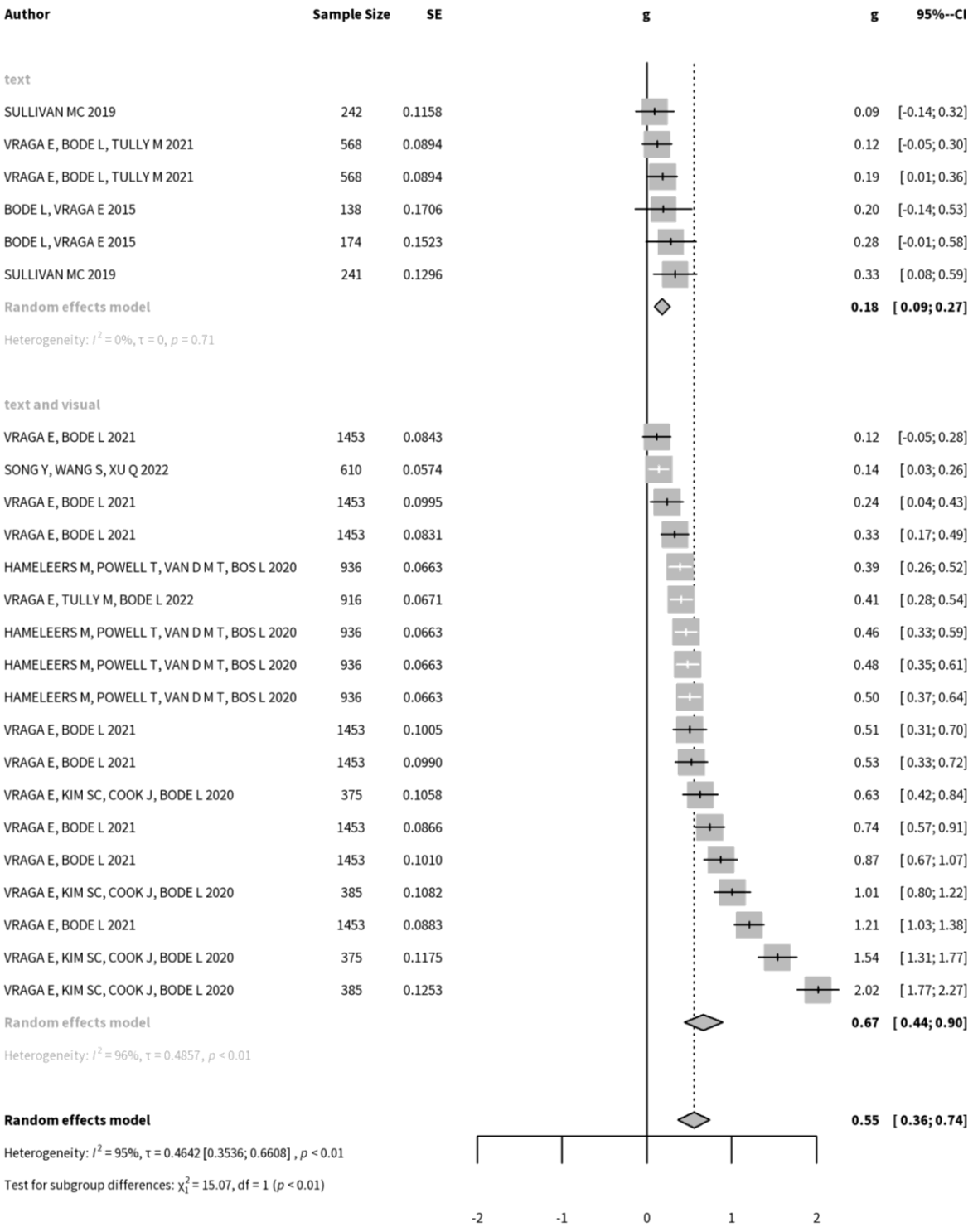
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S5a. Meta-Analysis of Content Labeling Studies by Label Format.



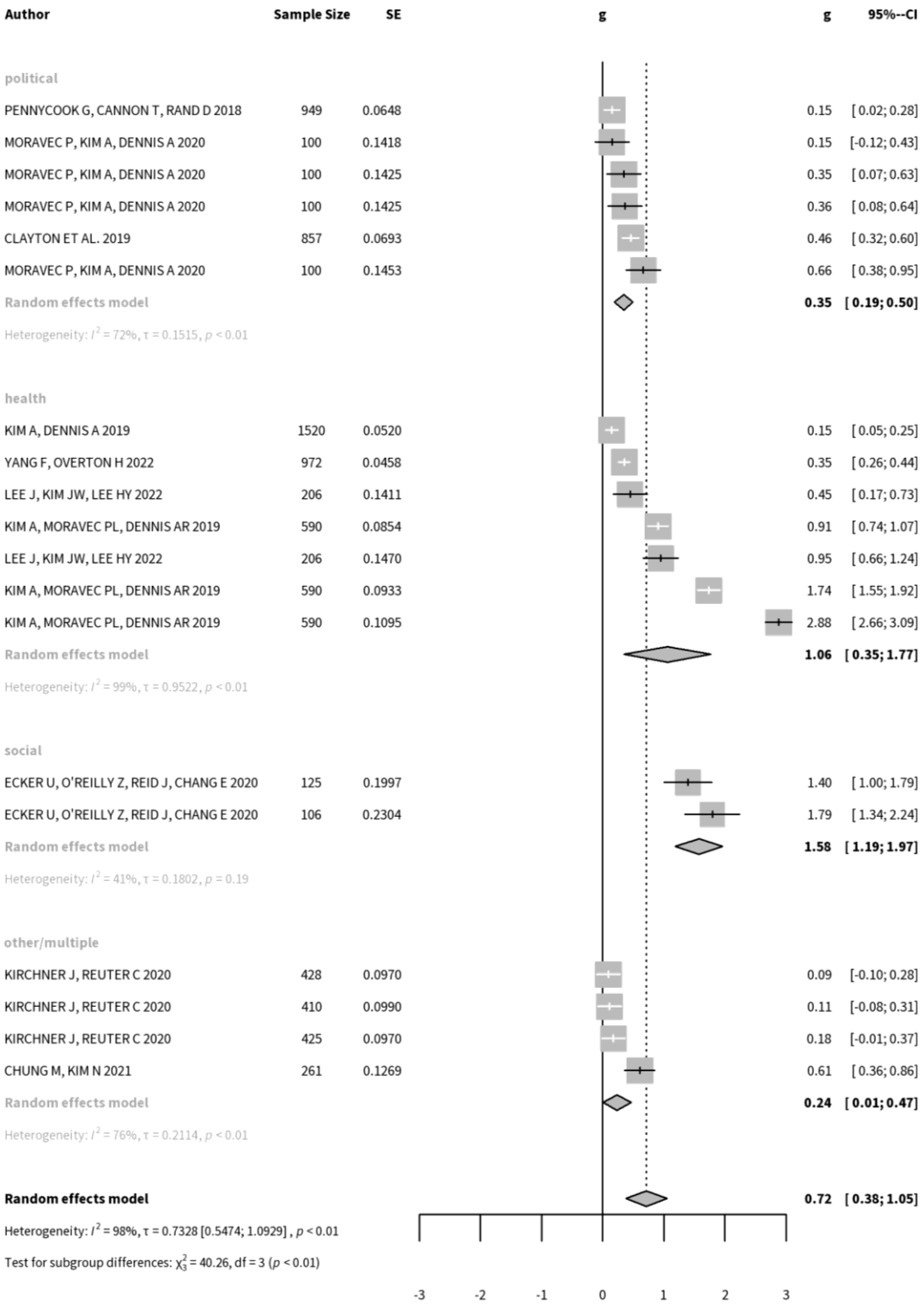
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S5b. Meta-Analysis of Corrective Information Studies by Corrective Material Format (Text or Other).



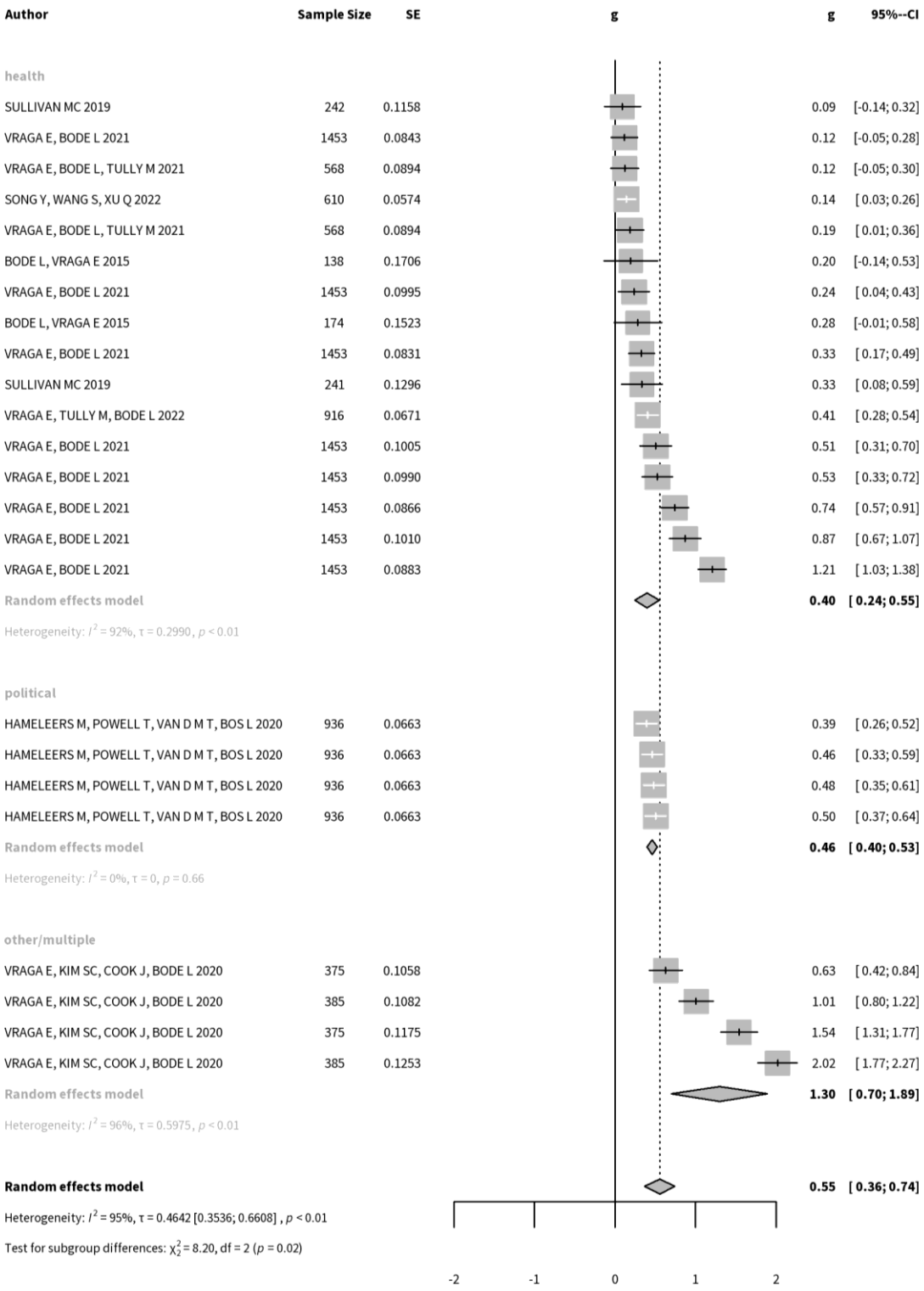
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S6a. Meta-Analysis of Content Labeling Studies for Misinformation Issue.



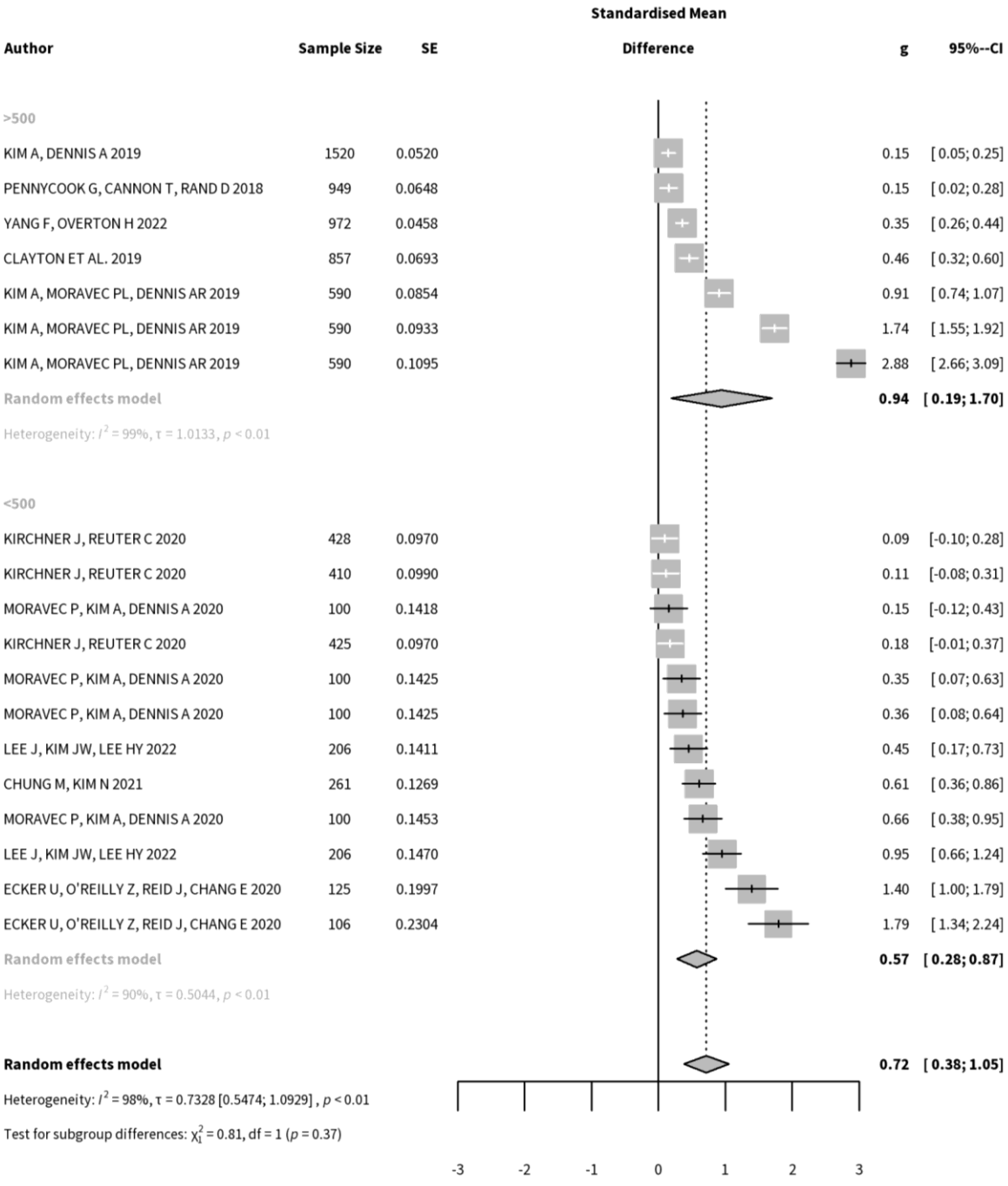
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S6b. Meta-Analysis of Corrective Information Studies by Misinformation Issue.



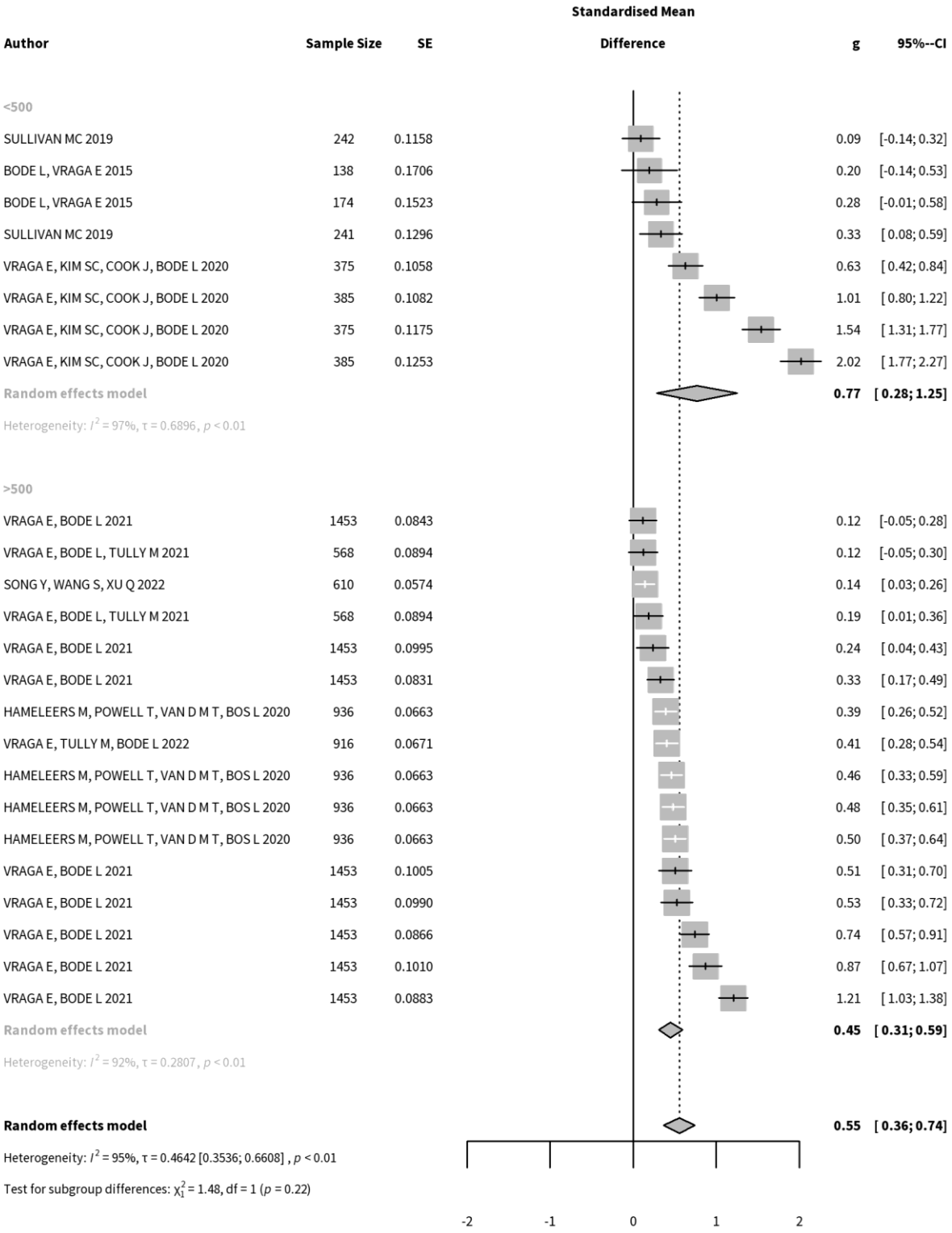
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S7a. Meta-Analysis of Corrective Information Studies by Sample Size.



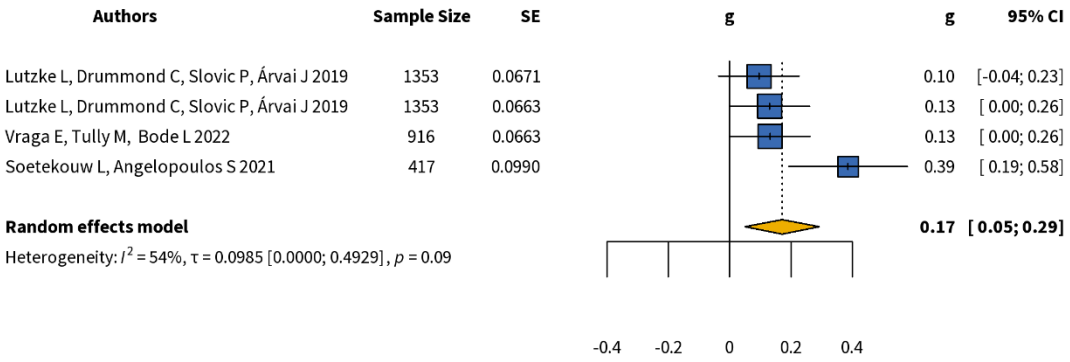
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S7b. Meta-Analysis of Corrective Information Studies by Sample Size.



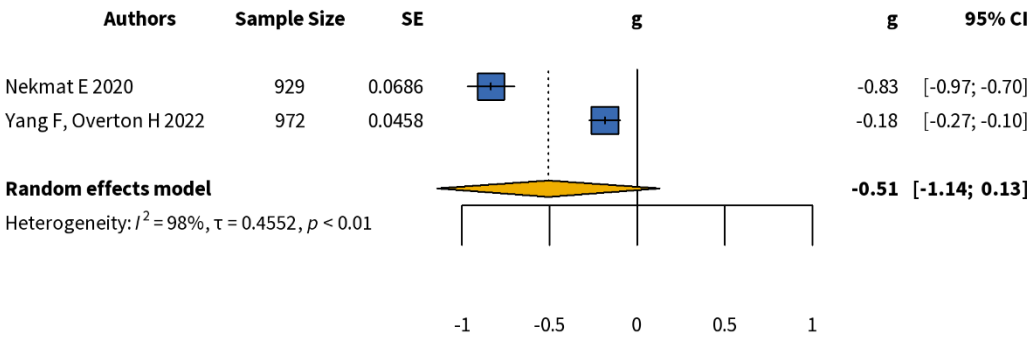
Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' *g*. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S8. Meta-Analysis of Studies Analyzing the Effect of Information & Media Literacy on Information Misperception.



Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' *g*. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

Figure S9. Meta-Analysis of Studies Analyzing the Effect of Content Labeling on Misinformation Sharing Behavior.



Note: Effect sizes were calculated from random-effects meta-analysis using Hedges' g. τ was calculated using the Paule-Mandel estimator.
Source: IPIE calculations based on data collected.

ACKNOWLEDGMENTS

Contributors

Drafting Authors: Aliaksandr Herasimenka (Consulting Scientist, Belarus/UK), Sebastián Valenzuela (Chair of the Methodology Panel, Chile), Wendy Hui Kyong Chun (Vice Chair of the Methodology Panel, Canada), Young Mie Kim (Vice Chair of the Methodology Panel, South Korea/USA), Philip Howard (IPIE Chair, Canada/UK). Independent methodology reviews: Shelley Boulianne (Canada/France), Simon Munzert (Germany). Independent general reviews: Larry Vernon Hedges (USA), Jake Shapiro (USA). Code stress test: Valentina Proust (Chile/USA). Research assistance: Anna George (USA), Adrienn Lawson (Hungary), Alexandra Pavliuc (Canada), and Xianlingchen Wang (China). Design, Formatting, and Layout: Domenico Di Donna, Raffaele Fulgente. Copyediting: Beverley Sykes. Independent fact-checking: Heidi Schultz. We gratefully acknowledge support from the IPIE Secretariat: Sheldon Himelfarb, Egerton Neto, Donna Seymour, and Alex Young.

Funders

The International Panel on the Information Environment (IPIE) gratefully acknowledges the support of the Oxford Martin School, Ford Foundation, Conrad N. Hilton Foundation, Rockefeller Brothers Fund, Alfred P. Sloan Foundation, and Skoll Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the IPIE and do not necessarily reflect the views of the funders.

Declaration of Interests

IPIE reports are developed and reviewed by a global network of research affiliates and consulting scientists who constitute focused Scientific Panels and contributor teams. All contributors and reviewers complete declarations of interests, which are reviewed by the IPIE at the appropriate stages of work.

Preferred Citation

An IPIE *Summary for Policy Makers* provides a high-level precis of the state of knowledge and is written for a broad audience. An IPIE *Synthesis Report* makes use of scientific meta-analysis techniques, systematic review, and other tools for evidence aggregation, knowledge generalization, and scientific consensus building, and is written for an expert audience. An IPIE *Technical Report* addresses particular questions of methodology, or provides a policy analysis on a focused regulatory problem. All reports are available on the IPIE website (www.IPIE.info). This document should be cited as:

International Panel on the Information Environment, 2023. *Platform Responses to Misinformation: A Meta-Analysis of Data*. SR2023.2. Zurich, Switzerland: IPIE.

Copyright Information

ABOUT THE IPIE

The International Panel on the Information Environment (IPIE) is an independent and global science organization committed to providing the most actionable scientific knowledge about threats to the world's information environment. Based in Switzerland, the mission of the IPIE is to provide policymakers, industry, and civil society with independent scientific assessments on the global information environment by organizing, evaluating, and elevating research, with the broad aim of improving the global information environment. Hundreds of researchers from around the world contribute to the IPIE's reports.

For more information, please contact the International Panel on the Information Environment (IPIE), secretariat@IPIE.info. Seefeldstrasse 123, P.O. Box, 8034 Zurich, Switzerland.