©NIJCRHSS
Publications

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES (NIJCRHSS)

Volume 6 Issue 1 Page 53-58, 2026

https://doi.org/10.59298/NIJCRHSS/2025/61.5358

Comparative Effectiveness of Prebunking and Debunking in Misinformation Mitigation

Kakembo Aisha Annet

Faculty of Education, Kampala International University, Uganda

ABSTRACT

Misinformation poses a persistent threat to public understanding, decision-making, and democratic processes, and necessitating effective mitigation strategies. This study examines the comparative effectiveness of prebunking and debunking interventions in reducing belief in misinformation, limiting its spread, and fostering long-term resistance. Prebunking, grounded in inoculation theory, aims to prepare individuals against misinformation prior to exposure by enhancing critical awareness of manipulation techniques, while debunking focuses on correcting false claims after exposure. Drawing on experimental and real-world evidence, the analysis evaluates both short-term outcomes such as belief accuracy and sharing intentions and longer-term effects, including retention and behavioral change. Findings suggest that both strategies are effective but operate through distinct mechanisms. Prebunking demonstrates stronger potential in preventing initial susceptibility and reducing residual influence, particularly among audiences with low prior knowledge. Debunking, however, remains more effective in correcting entrenched false beliefs once misinformation has been internalized. The evidence further indicates that contextual factors such as audience characteristics, content type, and platform dynamics significantly moderate outcomes. Overall, the study highlights that neither approach is sufficient in isolation; instead, a complementary strategy integrating prebunking and debunking offers the most robust framework for mitigating misinformation and enhancing information resilience.

Keywords: Misinformation, Prebunking , Debunking, Inoculation Theory and Information Resilience.

INTRODUCTION

Misinformation continues to challenge society, raising concerns about its detrimental effects on individuals and democratic institutions [1]. Different interventions have been designed and deployed to mitigate misinformation effects, aiming to decrease belief in misinformation and increase support for counterclaims. Among these options, debunking, an intervention delivered after encountering misinformation and appears to be the most widely researched [2]. More recent projects have introduced a proactive intervention called prebunking, which is delivered before exposure to misinformation. Using a preparation-based strategy, prebunking aims to reduce the negative impact of misinformation by increasing individuals' ability to recognize manipulation tactics or by training knowledge countering specific misinformation or a general topic [3]. Initial studies suggest that prebunking is a promising approach for mitigating misinformation effects and, in some cases, can be as effective as debunking. Comparative assessments have begun to evaluate the effectiveness of prebunking against debunking interventions in terms of reducing belief in misinformation, increasing support for counterclaims, and increasing resistance against future misinformation or counterclaims [4]. To evaluate the comparative effectiveness of prebunking versus debunking, focus on studies employing the same type of material and a similar content structure for each approach [5]. Taking into consideration both short-term impacts on belief-in and intention-to-share misinformation and longer-term retention of information and motivations to counter future misinformation,

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

conclude that studies provide only limited evidence for the comparative effectiveness of prebunking and debunking, the preponderance of evidence suggests that debunking might be more effective than prebunking in addressing misinformation[5]. These findings highlight the importance of further comparative investigation to advance knowledge on how prebunking and debunking complement and compete with each other in mitigation efforts [5].

Conceptual Foundations

Prebunking and debunking are two distinct approaches to misinformation mitigation relevant across diverse contexts ranging from health, science, and politics to conspiracy theories [4]. Debunking refers to the exposure of misinformation after initial contact with a claim, whereas prebunking aims to anticipate and counteract misinformation before it is encountered [3]. Each approach serves to correct different types of misbeliefs: debunking targets inaccuracies that individuals believe to be true, whereas prebunking seeks to address claims that individuals have yet to consider exposing techniques of misinformation in advance. These definitions make the two processes conceptually similar to the distinction between exposure and inoculation [5]. Prebunking has garnered interest among researchers and practitioners, supported by theoretical models and holds promise. The comparative effectiveness of the two approaches remains empirical and exploratory [3]. Misinformation spread is facilitated by three common mechanisms. First, people detect information that matches existing beliefs. Second, they fail to establish meaningful, credible beliefs when first encountering contrasting information. Third, once individuals detect content that conflicts with memory traces of previously held beliefs, the information may be confounded and reinterpreted as confirming interaction. Misinformation correction occurs in response to these mechanisms through separate routes. Beliefs can be reassessed, facilitating revision in response to clarifying details [5]. Alternatively, a combination of correction processes emerges when individuals still accept the initial message. First, exposure to contrasting content encourages reconceptualization, triggering partial resolution of initial uncertainty; transitional representations from the first and second messages may combine, yielding a complex new belief distinct from either source [2]. Prebunking is related to either model but offers distinct advantages when applied early in the misinformation interaction cycle [1].

Definitions and Distinctions between Prebunking and Debunking

Misinformation, which includes misleading information, disinformation, and mal-information, poses a serious threat to communication and, consequently, society [5]. To combat the spread of misinformation, research has focused on the effectiveness of debunking (post-exposure correction of misinformation) versus the relative efficacy of prebunking (pre-exposure education on the techniques or content of misleading information) [1]. Prebunking interventions consist of techniques that help to increase critical reasoning towards, or awareness of, misinformation and include inoculation, warnings, counter-speech, and literacy training [2].

Theoretical Mechanisms of Misinformation Spread and Correction

Misinformation threats can be addressed pre-emptively. Misinformation has wide-ranging negative societal consequences. Techniques to reduce misinformation include prebunking and debunking [1]. Prebunking refers to pre-emptive instruction that warnings accompany content. Such up-front warnings about misleading information reduce reliance but do not eliminate it [3]. A more effective form of prebunking inoculates individuals against misinformation sources and techniques. Developing resistance is based on McGuire's inoculation theory, which suggests individuals become resistant to persuasive messages when exposed to weak versions of those messages [4]. For misinformation, the "inoculum" is information about specific techniques used to persuade (rather than the false claims or leaks). Recent studies suggest inoculation should focus on general manipulation techniques instead of particular content because misinformation propagates rapidly [4]. General techniques effective for conspiracy theories include explanations of false-balance media coverage, academic 'fake-experts', clickbait headlines, and fear appeals. Interactive online games, such as the Fake News Inoculation game, allow participants to learn about misinformation techniques. In refutational pre-emption, warnings, belief assessment, corrective feedback, and examples of misinformation accompany content to enhance resistance [5].

Methodological Approaches in Comparative Evidence

Prebunking consists of warning, inoculation, and literacy training ahead of misinformation exposure, whereas debunking pertains to correction after exposure [5]. Although theoretical perspectives suggest distinct processes governing the spread and correction of misbeliefs, the comparative nature of empirical research has received less attention. Prebunking remains both conceptually and practically connected to debunking; the latter approach often constitutes a mismatch for early misinformation encounters [4]. Satisfaction with interventions is a further consideration even a good source may evoke doubt if believability is already weak and prevention is easier than correction. Prebunking therefore constitutes a promising candidate alongside pre-bunking [3]. Consequently, a brief meta-review of evidence comparing the two methods highlights a range of experimental and real-world studies, outcome metrics, and measurement methods applied to the question of comparative effectiveness [5].

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

Without intending to perform an exhaustive survey, empirical efforts thus far are documented to illustrate emerging patterns. The review adopts an incremental perspective, beginning with the simplest short-term beliefs and intentions to establish how the combined basis for comparative evaluation has developed [4]. The studies comprise a mix of laboratory-type experiments conducted online in varied settings and of extensive pre-registered real-world efforts outside the laboratory atmosphere, where more restricted methodological options often limit the scope for experimental work [4]. Experimental evaluations of comparative effectiveness offer a strong foothold, while fully registered and openly reported real-world studies contribute additional momentum [1].

Experimental Designs and Real-World Evaluation

Misinformation is a serious threat to society and the public communication of science. Efforts to mitigate the spread and impact of misinformation include prebunking and debunking interventions [2]. Prebunking attempts to inoculate people against misinformation before its exposure through strategies that help them anticipate how and why they might be misled and recognize it when they encounter it. In contrast, debunking corrects falsehoods after they have been spread. Well-designed experiments show that prebunking interventions are effective at reducing belief in various kinds of misinformation [1]. Practical evaluations of both approaches show that both mitigate misinformation. On TikTok, just before the election in Turkey and in the context of rising tensions about the capacity of the electrical grid, a prebunking video aimed at preparing the audience for misinformation surrounding a hypothetical electricity blackout was successfully deployed [4]. However, it coincided with the systematic release of a sequence of debunking videos addressing grid stability misinformation before and after the election [3]. From an audience point of view, those videos deserved separate assessment of their individual effectiveness, but determining the possibility of undertaking both strategies simultaneously in the same campaign raised questions about which strategy could be delivered first when integrated[5].

3Outcome Metrics and Measurement Challenges

Undeniably, determining appropriate measures to evaluate the effectiveness of prebunking and debunking initiatives is a formidable task [2]. These measures encompass a wide spectrum of outcomes such as belief, perception, attitudinal intention, and behavior. Challenges emanate from the broad range of outcome metrics employed by scholars [1]. In contrast, the vastness of such a realm risks trivializing and undervaluing both approaches through their separation into distinct meta-categories of debunking and prebunking. Yet, every method still constitutes but one fix in a spectrum of conceivable strategic concoctions [3]. Instead, when defining the task more narrowly as the identification of comparative effectiveness, a distinctive emphasis emerges. Choice of method, measurement, and sample becomes paramount in differentiating modes operationally rather than categorically, thereby permitting comparison of forms from within the shared classification of analogy [4]. Hence, the ongoing exploration of past studies indicates that both interventions pivot on analogous conceptual mechanisms and endeavor to elicit comparable attitudes and behavioral intentions. The examination of comparative effectiveness focuses attention on strength and retention as key characteristics exhibiting variance across formats [4]. In establishment of criteria for determining comparative effectiveness, strength and retention subsequently assume central significance. The exploration of strength encompasses evaluation of belief in the misinformation claim, perception of the claim's veracity, plausibility evaluation, and intention of either forwarding or sharing the misinformation [5]. Retention efforts concentrate on belief alteration and attitudinal intention, naturally comprising varying aspects. Additional formats are conceivable, yet either of the outlined pairs emerges repeatedly in existing experiments. This enduring recurrence hardly testifies to the assertion that these modes represent the definitive mechanisms in the respective range of belief or comparable alternatives. Indeed, comparing metric classicism along each pairing constitutes but one nominal format within a broader search for comparative effectiveness [5]. For each of these distinct methodologies, the inquiry maintains a resolute focus on these two designated criteria.

Comparative Effectiveness: Empirical Evidence

Prior work has established that prebunking techniques are at least as effective as traditional debunking methods [1]. It is important to assess an array of comparative studies addressing firsthand empirical evidence on whether prebunking approaches are in fact superior to debunking methods [2].Prebunking and debunking interventions can promote either corrective or preemptive functions within an influence, topic, or channel framework [5]. Broad patterns indicate that prebunking is often more effective at changing general beliefs and broader social perceptions about intended and unintended misinformation on major social media platforms [3]. Followers of specific social media channels typically possess an explicit awareness of the underlying influence issue, enabling the correction of either unintended or intended misinformation during media production [2]. A more emergent pattern indicates that preemption during the media-production phase is particularly critical for types of intention media. Prebunking techniques are also found to produce lower levels of residual influence [4].

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

Short-Term Impacts on Belief and Intentions

Misinformation can exert powerful influences on individuals' beliefs and behaviour, in ways that persist long after the misinformation is seen and despite knowledge of its inaccuracy[2]. The most effective way to mitigate the impact of misinformation is not merely to debunk it after the fact, but rather to prebunk it before it is encountered. Consequently, many authorities erroneously consider the false knowledge created by misinformation to be unchangeable, which is ultimately misguided [1]. The pre-encounter debunking approach—colloquially referred to as prebunking has gained traction in communication campaigns aimed at tackling misinformation on social media, and concerns the strategic dissemination of warning information before individuals are exposed to a specific instance of misleading material[3]. Prebunking's effectiveness has been investigated at multiple levels of personal impact, such as changes in intention to share content, intent to buy advertised products after seeing misleading ads, and willingness to endorse undermined beliefs. It has also been measured via self-assessment of perceived ease of counter-arguing misleading content[4]. While satisfying the comparable measures condition stipulated earlier, attention is limited to studies that have tested both prebunking and debunking interventions on the same misinformation topic without the confound of extensively differing presentation formats or content[5]. The results suggest that, when misinformation remains salient and in the absence of alternative informative material on a similar topic, prebunking interventions exhibit greater comparative impact on belief and intention[3].

Long-Term Retention and Behavioral Change

The comparative effectiveness of prebunking and debunking in misinformation mitigation has been explored mainly with respect to short-term impact on beliefs and behavioral intentions [3]. The duration of attitude and belief change following exposure to prebunking and debunking, as well as the extent of any subsequent behavioral change, has received comparatively little attention [5]. Most of the available evidence suggests that both forms of prior warning have the potential to produce observable behavior change up to several weeks following the original exposure. In addition, a variety of moderators and boundary conditions impacting the effectiveness of each preventative strategy have been identified, enabling greater generalizability of findings across various contextual situations [5]. Research highlights a number of distinguishing characteristics of the two strategies that have implications for the nature, form, and effectiveness of communication interventions [1].

Moderators and Boundary Conditions (Audience, Content, Platform)

Evidence on moderators of the comparative effectiveness of prebunking and debunking is scarce, hindering comprehensive conclusions. Available speculations indicate how audience characteristics, content attributes, and platform settings may influence effectiveness [3]. Considerations related to the audience shape expected comparative effectiveness of prebunking and debunking interventions. Prebunking is advantageous for predisposed audiences with low prior knowledge of content. Misinformation exposure increases likelihood of accidental and intentional engagement, raising the need for interventions that build preparatory and corrective shields in what remains of the safety window [2]. Counteracting misinformation campaigns requires unobjectionable or easily ignored corroboration [3]. Omission of information that a growing segment of the audience already possesses amplifies skepticism and can lead to misinterpretation [4]. Confirmation bias hampers counter-misinformative resiliency by motivating selective exposure to congruent material, augmenting the original misconception. These insights emphasize the urgency of strategy selection informed by target audience analysis, to which the literature supplies few guidelines [5]. Content parameters similarly feature among acknowledged moderators of comparative effectiveness. Competing materials eliciting opposing propositions alongside framing elements entrench prior opinions, rendering a compelling narrative essential to advance toward resolution [3]. Compatibility of the alternative proposition with misinformation shapes corrective shields and determines ideological alignment between corrections and original statements, presenting either an opportunity or a threat. Public elaboration on recurrent, high-salience topics thwarts the emergence of vigilance and second-guessing, causing sporadic attention to the emerging discourse [2]. Each observation underlines the importance of situational and topical characteristics in the choice of platform or technique [2].

Practical Implications for Communication Interventions

When embarking on communication interventions aiming to counter misinformation, practitioners must consider when and how to engage with both prebunking and debunking strategies that promote resilience against further engagement [1]. Because exposure to misinformation can occur unpredictably, prebunking is often perceived as a more effective initial response when neither misinformation nor a corrective notice has been encountered. Further contentious issues arise with respect to the ethical implications of various communication interventions [2].

Strategic Deployment of Prebunking

Prebunking interventions are applied before encountering misinformation to prevent or reduce reliance on it. Techniques include up-front warnings, inoculation, and literacy training [2]. Up-front warnings inform people that the presented information might be misleading; research shows they reduce, but do not eliminate, reliance on

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789
misinformation [2]. Inoculation involves exposing people to examples of misleadings and general manipulation techniques, rather than specific content, due to the rapid production of misinformation [1]. Recent research emphasizes inoculating against manipulation techniques such as false balance and fake experts. Online tools like the fake news inoculation game educate users about common misinformation techniques. Refutational preemption involves presenting scenarios of misinformation with corrections, asking for belief judgments, and providing examples demonstrating debunking mechanisms [1].

Integration with Debunking Efforts and Fact-Checking

Prebunking can reduce initial impacts and mitigate long-term retention [1]. It plays a vital role in countering disinformation campaigns, particularly when falsehoods gain traction, limited by audience reach, trust, and attention [1]. Audience targeting and strategic timing can amplify strength if multiple messages stem from one pre-exposure. Messages that signal distribution are feasible in pre-exposure stages [1].

Ethical Considerations and Informed Consent

Misinformation is sometimes countered through debunking or fact-checking, that is, by attempting to rebut the specific false claims it contains [5]. In contrast, prebunking seeks to prevent the acceptance of misinformation by countering the lead-up to it, or the arguments, cues, and reasoning that are likely to accompany it [1]. Prebunking has been defined as exposure to a warning or counter-argument before misinformation in order to foster resistance to the misinformation when it is later encountered [2]. Debunking has been defined as exposure to a corrective message or a rebuttal of the target misinformation after the misinformation has already been encountered. It is therefore possible to face a decision about whether to implement a prevention- rather than a remedy-oriented communication intervention [3]. Comparative evidence on the efficacy of these two forms of intervention is limited, but it indicates that preventing the uptake of misinformation is more effective than attempting to correct it at a later stage. Both interventions are effective, but exposure to misinformation that has subsequently been corrected can reinforce the very beliefs and intentions that the correction seeks to modify. In the absence of preparatory intervention, misinformation that has already been faced acquires heightened salience and persuasive impact. The most desirable outcome is therefore a combination of prebunking inoculates against misinformation that is likely to be encountered subsequently and fact-checking on which the audience had had time to reflect [4].

Gaps in Knowledge and Future Research Directions

A systematic analysis of the existing literature reveals various gaps in knowledge regarding the comparative effectiveness of prebunking and debunking as interventions to mitigate the risk of harm associated with exposure to misinformation [1]. Understanding these gaps can inform future research efforts, which may build on, refine, and extend the existing body of work. Seven key research directions warrant further attention. First, the longitudinal retention of short-term effects is not yet clear [2]. Existing studies have demonstrated that prebunking and debunking interventions effectively reduce the likelihood that individuals will endorse, believe, or intend to share misinformation and grooming-related content or exhibit ideologically biased search intentions soon after exposure [3]. Determining how long such effects persist and whether these approaches foster more enduring protective resistance remains an important question. Although some experimental research evaluates the durability of pre-exposure inoculation, comparable insights regarding the longevity of misinformation retraction effects are lacking [5]. Second, the relative persistence of belief-change effects following different intervention types is unknown. Prebunking may disable crystallization by inoculating individuals against manipulation techniques used to construct misinformation narratives, while debunking may drive individual beliefs back toward a state of uncertainty, thereby facilitating later narrative reconstruction. Further research on this topic would clarify comparative retention times and enhance understanding of intervention mechanisms, further refining advice on when and how to combine these techniques [4]. Third, the interplay between long-term resistance and behavioral intentions remains to be explored. Short-term studies show that individuals retain a greater degree of protective resistance against manipulation techniques following prebunking, whereas comparable studies of subsequent behavior are scarce. Investigating how these constructs evolve across varying time horizons would illuminate the relationship between belief-change persistence and ongoing behavioral intent [5].

CONCLUSION

The comparative analysis of prebunking and debunking underscores that both interventions play critical but distinct roles in mitigating misinformation. Prebunking is particularly effective as a preventive strategy, equipping individuals with cognitive tools to identify and resist misleading information before exposure. Its emphasis on generalizable manipulation techniques enhances resilience across diverse contexts and reduces the likelihood of misinformation taking hold. In contrast, debunking remains indispensable for addressing already-internalized false beliefs, providing corrective information that can realign perceptions with factual accuracy. However, the evidence reveals that each approach has limitations when applied independently. Prebunking may not fully counter deeply entrenched misinformation, while debunking can sometimes reinforce false beliefs due to continued exposure

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

effects. Moreover, the effectiveness of both strategies is highly contingent on factors such as audience predispositions, message framing, and platform environments. Consequently, the most effective misinformation mitigation framework is a hybrid model that strategically integrates prebunking and debunking. Prebunking should be deployed proactively in anticipation of misinformation, particularly in high-risk information environments, while debunking should be applied responsively to correct circulating falsehoods. Future research should focus on long-term effectiveness, behavioral outcomes, and optimal sequencing of interventions to refine this combined approach. Strengthening this evidence base will be essential for designing scalable, context-sensitive communication strategies capable of addressing the evolving challenges of misinformation in the digital age.

REFERENCES

1. Siebert J, Siebert JU. Effective mitigation of the belief perseverance bias after the retraction of misinformation: awareness training and counter-speech. PLoS One. 2023;18(3):e0282202. doi:10.1371/journal.pone.0282202.
2. Bryanov K, Vziatysheva V. Determinants of individuals' belief in fake news: a scoping review determinants of belief in fake news. PLoS One. 2021;16(6):e0253717. doi:10.1371/journal.pone.0253717.
3. Xiao X. Not doomed: examining the path from misinformation exposure to verification and correction in the context of COVID-19 pandemic. Telemat Inform. 2022;74:101890. doi:10.1016/j.tele.2022.101890.
4. Sharevski F, Devine A, Pieroni E, Jachim P. Meaningful context, a red flag, or both? Preferences for enhanced misinformation warnings among US Twitter users. In: Proceedings of the 2022 European Symposium on Usable Security (EuroUSEC '22); 2022 Sep 28–29; Karlsruhe, Germany. New York (NY): Association for Computing Machinery; 2022. p. 189-201. doi:10.1145/3549015.3555671.
5. Paynter J, Luskin-Saxby S, Keen D, Fordyce K, Frost G, Imms C, et al. Evaluation of a template for countering misinformation—real-world Autism treatment myth debunking. PLoS One. 2019;14(1):e0210746. doi:10.1371/journal.pone.0210746.

CITE AS: Kakembo Aisha Annet (2026). Comparative Effectiveness of Prebunking and Debunking in Misinformation Mitigation. NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES, 6(1):53-58. https://doi.org/10.59298/NIJCRHSS/2025/61.5358