

Appendix E: Pilot study

We have conducted a pilot study employing the same manipulation in different contexts. A pre-registered pilot study ($N = 1,225$) was conducted in Singapore to examine the influence of pseudoscientific content and accusations of censorship on misinformed beliefs about COVID-19 vaccine side effects, with a specific focus on claims about excess mortality rates. Employing a 2 (pseudoscientific content vs. no pseudoscientific content) * 2 (censorship: presence vs. absence) * 2 (platform: health blog vs. messaging application) between-subjects factorial design, the study did not include a neutral control group, relying instead on comparative effects between misinformation versions.

Results indicated that the mere inclusion of pseudo-statistical details (pseudoscience)—such as correlation coefficients and confidence intervals—did not significantly alter belief in misinformation or sharing intentions. Conversely, accusations of censorship produced a “backfire effect,” significantly reducing respondents’ intentions to share the information on social media compared to messages without such claims.

The study also revealed a critical boundary condition for these main effects: a significant interaction between pseudoscientific content and censorship claims. When pseudoscientific content was combined with censorship accusations, participants reported greater belief in the accuracy of the misinformation than in conditions where censorship claims appeared in isolation. Still, the overall effect sizes were small.

It is also important to note that the nature of the manipulation in this pilot study was somewhat different: it did not include a table of statistical results, and only statistical findings were reported in parentheses within the text. We determined it had a weak dosage manipulation, and some people may not have paid attention to the few statistical details. Data were also collected at a peak time (higher infection rates) during the pandemic. See further discussion below in the manipulation checks section.

Manipulation checks results in the pilot study: We included two manipulation check questions. The first task asked participants to identify the content they had read, offering one correct and three incorrect response options. The second task asked them to select all applicable content features from a list of 10 items (5 correct, 5 incorrect). The first manipulation checks generally performed well. However, the second question yielded mixed results: pseudoscientific content manipulation: Of 606 respondents exposed to messages with pseudo-statistics, only 171 (28.2%) correctly indicated having seen statistics.

Censorship manipulation: Of 597 respondents exposed to censorship claims, only 141 (23.6%) correctly identified this element.

Platform manipulation: Of 606 respondents shown a Substack-like blog format, 367 (60.6%) correctly identified it. Of 619 respondents shown a WhatsApp-like messaging application format, only 110 (17.8%) correctly identified it, while 169 (27.3%) incorrectly reported seeing a Substack-like blog post.

The pilot study provided valuable insights but also revealed methodological limitations that we addressed in the current study. Specifically:

First, the unsatisfactory performance of the second manipulation check question raised concerns about whether participants adequately processed the pseudoscientific content and censorship elements. To address this, we designed multiple more targeted manipulation-check questions in the current studies to ensure that our treatments were properly perceived and processed (see Appendix A for validation results).

Second, the pilot study's focus on COVID-19 vaccine misinformation and its use of presentation formats (health blog vs. messaging app) as experimental stimuli may have lacked ecological validity. In practice, detailed misinformation containing statistics appears more commonly in certain contexts than others. Therefore, in the current study, we deliberately selected two distinct topics to examine whether the interaction between pseudoscientific content and censorship claims generalizes across domains (as a conceptual replication) or represents a topic-specific phenomenon.