Supplementary Material

A revised behavioral analysis of the late 2020 anti-vaccination infodemic on Twitter

Response to issues listed in the retraction notice

Issue 1: "The sampling reported in the study is inadequate; the sample size is too small, and the study does not report adequate information to assess whether the selected sample is sufficiently representative. In addition, it appears no attempt was made to match profiles by profile characteristics such as number of followers, how long the profile has been active, or the number of tweets."

We analyzed, as reported at the time of our original submission, and as accepted following peer review, 50 different profiles for each group (control, anti-vaccination, and pro-vaccination), retrieving them randomly through a search on Twitter using the hashtags defined in the materials and methods section of the paper. While this point was not criticised when the paper was scrutinised by PLOS ONE peer reviewers during the submission process, the ad hoc post-publication review carried out by PLOS ONE found that the sample size is too small. The analysis was conducted manually, and required reading, analyzing, assessing, and categorizing, several thousand tweets. Any automated detection method, to our knowledge, would have failed at providing a similar depth or accuracy of the analysis – thus a manual approach, which limited the sample size, was preferred. Of note, this concern should have been raised at the time of submission, after the initial editorial assessment, or at least during the peer review phase. As for the "profile characteristics", they were not provided since: a) demographic characteristics are generally not made public by Twitter users; b) it would require publishing personal information beyond the scope of the paper, doing no practical service to science, without enhancing or corroborating the findings of our research. Similarly, we considered that knowing for how long a profile was active on Twitter was not relevant for the scope of our research. This, and any other subjective concerns about our work should have been raised, if any, during the editorial evaluation phase or during the peer review phase. That said, we agree that a sample of 50 may be limited and not necessarily representative of the anti- or pro-vaccination communities at large, but the sample was likely sufficiently representative of the pro- and anti-vaccination communities on Twitter, at the time of the analysis.

Issue 2: “The anti-vaccination and pro-vaccination search terms used in the study may not have been balanced appropriately and the study does not report on the justification for the choice of hashtags used. Similarly, the use of a random word generator to create a random hashtag to use as control is inappropriate and suggests that the study did not include an appropriate characterization of underlying Twitter behavior.“

As for issue n.1, also the choice of hashtags could have been questioned when most appropriate: during the editorial evaluation phase or during the peer review phase. We made the choice of using the hashtags described in the materials and methods section of the paper, because they are the most representative and most used by the pro- and anti-vaccination profiles on Twitter, at the time of the analysis. Other hashtags, less used, exist, of course. The most comprehensive way to catch anti-vaccine discourse would have required us to detect any circulating tweet on Twitter matching to pro- or anti-vaccine discourse, a type of analysis that is not aligned with the scope our study, as it would not have allowed for a detailed categorization of each tweet under scrutiny. The use of different hashtags, we agree, would have led to different results (please note this is an obvious statement), yet, we believe, it would have led to very similar underlying interpretations and conclusions, and value of our study. Choosing different hashtags without proper reasoning behind the choice would have negatively impacted the broader validity of our results.

Issue 3: “The study does not provide an adequate definition of “emotional language”, and the related results reporting on the use of emotional language include an outlier data-point in the pro-vaccine group, which could drive the effect significantly in a study with a small sample size. “

Concerning the definition of “emotional language”, we rely on a broad definition, i.e., the use of specific words or combination of words to describe or evoke an emotional reaction. Any concerns regarding the adopted definition of emotional language should have been raised prior to publication, or at least we should be given the opportunity to clarify this with a corrected version of the manuscript. As for the outliers, the handling of outliers has been clearly stated in the materials and methods section of the paper and in the figure legends in the main section of the manuscript. These outliers were removed since they were “behavioural outliers” – as an explanatory example, an outlier is a profile tweeting hundreds of tweets per day, when the average is a few tweets per profile per day in that specific group under analysis. These types of behaviours indicate that the Twitter profile could be a bot, or even if a real person is handling the profile, it would still indicate a behaviour outside the norm, which does not correspond to the behaviour of most users in the group. We thus stand by our choice to remove outliers as indicated.

Issue 4: "The network analysis includes only a small number of profiles with an unbalanced number of neighbors. In addition, the clustering coefficient is inappropriate and the absence of confidence intervals in Fig 5C is problematic. As currently presented, these results are not sufficient to draw meaningful conclusions."

The unbalanced number of neighbors in the pro-vaccination versus the anti-vaccination web is a result, i.e., a finding of our study, which we purposefully highlighted. To compose the webs, as explained throughout the paper, for each Twitter profile we retrieved the 10 most retweeted profiles – i.e., for profile "X", we retrieved Profile 1 to 10, whose tweets were retweeted the most by profile "X". We also included those large influencers (with at least 5 connections in our analysis – i.e., at least 5 of the analyzed profiles had them in the list of the 10 most retweeted profiles) and checked for and added connections with other profiles in our web when these were in the list of their 10 most retweeted profiles. We then connected profiles – those in our starting analysis as well as those retrieved with the above-mentioned analysis – and obtained the webs. The unmatched number of neighbors, thus, indicates that the pro- and anti-vaccination webs, considering the same initial pool of analyzed profiles, are having different characteristics – i.e., profiles in the anti-vaccination group are well connected with each other, and especially are connected to a web of strong and common influencers. As for the "missing confidence interval in Figure 5C", there is no confidence interval since the analysis is based on a snapshot of the situation at the time of the analysis (and thus relies on a single datapoint, which does not allow for statistical analysis of this kind). We believe the independent subject expert recruited by PLOS ONE may refer to the fact that for each node – i.e., profile in our web – we could have calculated the number of neighbors, and thus we could have represented the variation of the average number of neighbors with a confidence interval. Despite we agree with this, in our specific case, given that the "boarders" of our webs are defined – i.e., we removed individual clusters, as defined in the materials and methods section of the paper, which were not connected to other clusters, and the fact that the analysis does not represent the overall view of the webs, but a simplified web resulting from the limited pool of analyzed profiles, makes the request to add a confidence interval irrelevant.

Issue 5: “The reported conclusion “Our data demonstrate that Donald Trump, before his profile was suspended, was the main driver of vaccine misinformation on Twitter.” is not supported by the research reported in this study. Although the reported results suggests that people who tweet anti-vaccine content are likely to be in Trump’s network, the reported results are not sufficient to support the claim that Trump himself is driving vaccine misinformation.”

We agree that the sentence in the abstract, “Our data demonstrate that Donald Trump, before his profile was suspended, was the main *driver* of vaccine misinformation on Twitter", if taken alone and out of context, is not supported by the data. In this version of the manuscript, it has been changed for added clarity, as explained in the disclaimer. We recognize that several media outlets have misunderstood the results of our paper, likely because of this sentence in the abstract. Had this issue been spotted during the peer review process, we would have taken care to adjust this sentence (as in this version of the manuscript), either through the publication a small comment or through a correction of the paper – but we were not given such opportunity. That said, the discussion paragraph clarifies Trump’s role based on our result, and thus the sentence has been taken out of context by the editors requesting the retraction. In fact, based on our data we can conclude that Trump is the main *influencer* in the anti-vaccination community. This means that Trump, regardless of his opinions or positions about vaccines, had, at the time of the analysis, the capacity to influence the anti-vaccination community on Twitter.