**Response to Full Fact**
This and other additional material related to Fraiman et al. (2022) are on the Zenodo repository

**Preamble (written September 18, 2022)**

The preprint of our article was the subject of three fact checks:

- Lead Stories (July 1, 2022 archived version here)
- Health Feedback (July 1, 2022 archived version here)
- Full Fact (July 28, 2022 archived version here)

Of these three, Full Fact was the only organization that contacted us before publishing.

Full Fact contacted us on June 29 and asked for a response by the end of the day. We requested 24 hours (and subsequently another 24 hours), within which we returned our reply. Full Fact posted their fact check 4 weeks later (July 28).

Full Fact did not flag our preprint on social media, unlike Lead Stories. After inquiring, Full Fact informed us, "We only rated the Expose article link, so our check will only affect that article in terms of how it spreads on social media."

Our complete response to Full Fact is appended below.

Our response addresses concerns critics have raised regarding:

- **our comparison of harms vs. benefit: see pp 2-3**; our published paper in *Vaccine* also expanded the discussion of this topic
- **analyzing the number of SAEs vs. number of people with any SAE: see page 4**
- **p-hacking** and **cherry picking: see pp 5-6.** Also see two supplemental files we have made available on the Zenodo repository:
    - "Examples describing the different types of decisions that were made in the process of matching SAE types to AESI terms" (see Examples of AESI matching decision process.pdf)
    - "Sensitivity analysis excluding SAEs requiring clinical judgments" (see Sensitivity analysis excluding SAEs requiring clinical judgments.pdf)

Sarah Turnnidge
Full Fact

July 1, 2022

Dear Sarah,

Thank you for taking the time to seek our response.

We think there are two distinct concerns here:

1. Concerns about our preprint itself, such as the three points you raised in your email.
2. Concerns about how others have reported on our paper, such as the Exposé story you linked to in your original email, and critiques that accused us of transgressions which (as explained in our paper) we took precautions to avoid.

We think it is vital that the two are separated. We are clearly responsible for our preprint, and respond to each of your three points below. Nonetheless, we cannot take responsibility for what has been written about it, especially when that writing indicates the author paid no careful attention to the actual text of our paper and its supplements.

We hope your fact checking separates the above two issues, as errors in coverage of our paper should not be treated as actual errors in our paper. We are of course especially concerned about attempts to cast doubt on our integrity by portrayals of errors (whether real or not) as intentional.

With respect to the 3 points you listed in your email, our full responses are below. The experts you quoted raised points that are a mixture of valid and invalid concerns. For example, it is correct that the original trials could not definitively evaluate long-term effects such as vaccine protection against COVID-19 hospitalization (due to early unblinding and subsequent vaccinations of placebo controls); on the other hand, charges of P-hacking and cherry-picking are simply erroneous and ignore the measures we took to prevent those problems.

With respect to other errors in coverage, we share concerns about erroneous claims made about our article, such as that it was a World Health Organization study, or that our results are definitive in any way. In a future revision of the paper we will more thoroughly emphasize what we are not addressing, did not do, did not show, and do not claim, using a list of popular misinterpretations as our guide.

To be clear, we concluded that the data that has been made public showed “an excess risk of serious AESIs greater than the reduction in COVID-19 hospitalizations in both

Pfizer and Moderna trials." This is very concerning, as it suggests that harms might outweigh benefits for those at low risk of COVID-19 hospitalization. Our preprint lists the limitations of our analysis, some of which were forced by the unwillingness of Pfizer and Moderna to release anonymized individual participant level datasets. Regardless of those limitations, we believe our results reinforce calls for detailed quantitative analyses of the balance between vaccine harms and benefits using the most up-to-date data available, with special attention to individual patient characteristics (see the Discussion sections of the abstract and full text).

We address your queries in detail below. As a brief summary: Again, some criticisms of our study (especially charges of P-hacking) appear to be based on very careless reading of our study and its supplementary materials, and are outright misrepresentations of what we actually did. Others are more reasonable, but addressing them with the data available to us does not alter our conclusions. Addressing them in complete detail would require access to the individual participant level datasets, something neither Pfizer nor Moderna will do until their trials complete (which has not occurred). Our conclusions thus remain as stated in our abstract: "The excess risk of serious adverse events found in our study points to the need for formal harm-benefit analyses, particularly in individuals at low risk of COVID-19 hospitalization or death." To this we would now add: Our study points to the need for public release of the individual participant level datasets from the trials. Others have expressed similar concerns throughout the vaccination programs.

Turning to your numbered queries:

> 1. The study only considers Covid hospitalization during the vaccine trials themselves, which covered only around two months at a time when Covid-19 rates were low. The benefit of the vaccines extends beyond this two month period - one expert said "the harm/benefit comparison used in this study seems entirely inappropriate". Furthermore, the trials weren't primarily designed to look at hospitalization rates, (though they did report some data on hospitalizations), and to get accurate measures on hospitalization the trials would have had to have been longer and included more patients.

Our paper analyzes the benefits and harms data that supported the Emergency Use Authorization in December 2020. This is the same time point used by the FDA and other regulators worldwide prior to mass vaccination, and is used for most of the published information from the trials.

As we explain in the preprint, another reason we used the EUA time point is that participants began to be unblinded following EUA, and placebo recipients were vaccinated. This renders the post-EUA dataset less reliable due to breaking of randomization (see lines 96-102). That said, there are hypothetical reasons such as those given by the experts you spoke with (e.g. a longer trial, or a trial in higher risk populations) that could alter the harm- benefit balance as they apply to populations and

time periods different from those studied in the trial. But it is critical to realize that the hypotheticals do not all point in one direction. Some hypotheticals improve the anticipated harm-benefit, while others worsen it. A key message from our paper is that it is unlikely there is a single harm-benefit ratio for all populations at all time periods.

Here is a list of various hypotheticals, each of which change the predicted harm-benefit balance, for the better or for the worse:

- Had the trial continued in a blinded fashion for a longer period of time, the absolute magnitude of benefits against COVID-19 hospitalizations likely would have been larger, improving the balance of benefits against harms. Whether this difference would be large enough to outweigh the harms is unknown.
- Had the trials enrolled populations at high risk of COVID-19 hospitalization, the balance of benefits against harms would likely have been better. We concur with the expert who said the trials were not designed to study the vaccines' potential benefit against hospitalizations (one of us pointed this out in October 2020, before results were known, and advocated changing the trials' primary endpoint).
- If similarly sized trials (numbering in the tens of thousands) had been performed in those at lower risk of COVID-19 hospitalizations (such as young adults and children), the harm-benefit ratio would likely have been worse than in the actual trials.
- Changes in the circulating virus as seen today include higher infectivity, less virulence, and less effectiveness of the original (and still-current) vaccines. The net impact of these changes on the harm-benefit ratio would have to be examined with current data, although the reduced viral virulence and reduced vaccine effectiveness should increase the harm-benefit ratio since the vaccine side effects have not been reduced.
- Finally, given the novelty of large-scale use of mRNA vaccines and the lack of a very large randomized trial designed to study long-term SAEs, it remains a possibility (however remote) and an expressed public concern that the vaccines may cause delayed serious adverse events. We hope that ongoing observational studies will provide definitive assurances against this possibility.

We further hope that larger clinical trials are performed with proper patient-oriented clinical outcomes such as hospitalization as the primary endpoint, as the experts you spoke with have also deemed necessary to perform a proper harm-benefit analysis

> 2. The analysis in the study looks at the number of adverse events, rather than the number of people who experienced adverse events. Another expert told us: "It is far more conventional in adverse event analysis to analyse the number of patients with events, as it is not always possible to know whether multiple events in the same patient are really separate events or just a single pathological process with repeated flare-ups."

Regarding analyzing the number of SAEs rather than the number of people with any SAE, there is merit to both approaches, as they answer related but different questions.

Contrary to the above quote, our analysis does not include repeated flare ups of the same SAE, as this would have been reported as a single event. Lacking the individual participant level datasets, we were unable to examine this concern in our analysis. It is thus possible that our SAE rates are *underestimated*.

If one participant experienced two different SAEs, this *was* reported, and our analysis did take this into account. For example, if one patient experienced a heart attack after the first vaccine dose, and then a stroke after the second vaccine dose, our analysis would have accounted for both events, while counting participants who experienced any SAE would count the heart attack and the stroke (two SAEs) as one.

It is plausible that individuals susceptible to one vaccine induced SAEs may be susceptible to multiple SAEs. We did identify a signal that this *may* - we stress *may* - be occurring with the mRNA vaccines; as stated in the paper "approximately twice as many individuals in the vaccine group experienced multiple SAEs than the placebo group." Arguably, it is worse for two people to experience an SAE than for one person to experience two SAEs; nonetheless, it is also the case that two SAEs in one person is worse than one SAE in the same person. This is the rationale for counting *both* the number of events and the number of participants experiencing an event.

Most importantly, however, we could not complete the analysis of the number of participants with any SAE because individual participant level datasets are not available. If we gain access to these data, we will run both analyses. But the present lack of individual patient data forced us to perform the analysis only on the number of serious adverse events of special interest (AESI), rather than on participants. Our paper draws attention to the limitations of lacking these datasets, and one of our authors publicly called on the vaccine manufacturers to provide this information so that independent analysts can validate their work.

> 3. Questions have also been raised about the list of Serious Adverse Events used in the analysis, with accusations of P-Hacking due to apparent inconsistencies in the inclusion/exclusion of reported SAEs such as diarrhoea being included, but vomiting excluded. An expert told us: "If you look at serious adverse events overall, there is no significant difference between the vaccine and placebo groups. The results were not statistically robust to small variations in analysis: the difference between vaccine and placebo groups was no longer statistically significant in a sensitivity analysis in which they restricted their "adverse events of special interest" to a pre-specified list (the Brighton Collaboration's SPEAC list) and the events that the authors had added themselves were not considered. If other adverse events had been included the results might have looked quite different.

No P-hacking or “cherry-picking” occurred in our study. Such problems correspond to selection of classifications and results based on the observed P-values or interval estimates. We did not do that. We did not report P-values or make any reference to statistical significance, nor did we select classifications on the basis of the statistical outcomes. Thus accusations of P-hacking and cherry-picking can only reflect a failure to read and describe our paper carefully.

In reality, we used two independent clinician reviewers who were blinded to the results while judging whether to include or exclude various AESIs from the list of SAEs, thus refuting charges of P-hacking or cherry-picking. Furthermore, as stated in the paper “agreement between the two independent clinician reviewers was 86% (281/325); 40 of the 44 disagreements were resolved through consensus, and only four disagreements necessitated a third clinician reviewer.”

The standard approach to avoiding problems like P-hacking and cherry picking is to prospectively declare the analysis strategy based only on what is known prior to the study. That is exactly what we did. We based the analyses on the Brighton Collaboration list which used data collected prior to results from the trials; thus it is an *a priori* list. Again, this refutes claims that we engaged in P-hacking and cherry-picking, because those require highlighting selected outcomes derived from consideration of observed statistical results.

Furthermore, our analyses made statistical adjustments that *widened* the confidence intervals, making our results less likely to achieve statistical significance. Claims that our adjustments were inadequate are based on failure to recognize that more accurate adjustments could have resulted in less widening and hence *more* statistical significance; such adjustments would however require the individual participant data, which is being withheld from public scrutiny by Pfizer, Moderna, and the FDA.

We have adhered to the highest scientific standards in being transparent about what we’ve done and why. Our preprint includes links that allow readers to replicate our work, including public access to our full dataset. We included a sensitivity analysis on AESI lists (Supplemental Table 2); the results were statistically compatible with our main results, rather than qualitatively different as the expert you spoke with claimed. There are now dozens of articles and books explaining why changes in “statistical significance” is a fallacious basis for claiming results conflict; as one example we direct you to the following article in *The American Statistician* (an official publication of The American Statistical Association), which we attach for your convenience:

> Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.C., Poole, C., Goodman, S.N., Altman, D.G. (2016). Statistical tests, confidence intervals, and power: A guide to misinterpretations. *The American Statistician*, 70, online supplement 1 at

https://amstat.tandfonline.com/doi/suppl/10.1080/00031305.2016.1154108/suppl_file/utas_a_1154108_sm5368.pdf [reprinted in the *European Journal of Epidemiology*, 31, 337-350].

The claim that there were “apparent inconsistencies in the inclusion/exclusion of reported SAEs” is also refuted by close examination of the data. For example, including diarrhea but excluding vomiting is actually a consistency rather than an inconsistency, as diarrhea was one of the 29 clinical diagnoses on the Brighton AESI list, while vomiting was not. A legitimate concern is that it is a list of clinical diagnoses which “were known to have been reported but not in sufficient numbers to merit inclusion on the AESI list”, and its entries were not re-evaluated for potential inclusion on the official list used in the trials. Nonetheless, including additional types of SAEs not associated with vaccination should only increase noise, *reducing* statistical differences between vaccine and placebo group; yet the opposite occurred in both Pfizer and Moderna trials, showing increased risk with more precise confidence intervals.

We also presented results from looking at total SAEs. In the Pfizer trial there was a 36% increase in the total number of SAEs. Counting the number of SAEs in each group, as preferred by some critics, indicated an increase in the vaccine group. We have not found this increase in the number of SAEs in the vaccine group mentioned in public reports from the FDA, or why this result has received so little attention by those critics.

In conclusion, we are most concerned that, a year and a half after mass vaccination programs commenced, manufacturers have yet to publicly release essential trial data. Those who volunteered for the trials did so in the good faith that the data would be used for the greatest public good, and it is becoming clear that this good requires data release. Furthermore, there are well-established means for anonymizing the records so that the effort at individual identification would far exceed any value to any party in doing so. Covid vaccines are among the most widely disseminated medicines in the history of the world and they are paid for using government funds. We believe the public has a legitimate see exactly what happened in the trials. Placing analysis datasets in the public domain is an important step in this direction, and could help dispel some of the concerns that have arisen regarding vaccine safety.