Response to comments on Wiliamson et al. (OpenSAFELY)

We thank Westreich et al [1], and Tennant and Murray [2] for their thoughtful comments on our paper.

We entirely agree that readers should avoid interpreting our findings in causal terms. Following feedback on the initial preprint version (demonstrating that we did engage with such feedback), we made it clear in the paper that we had never set out to estimate causal effects, as both sets of correspondents acknowledge. In response to early feedback we also changed our terminology to avoid the ambiguous phrase “risk factors”.

While many epidemiological studies can be classified into causal or predictive, there is a third category which we believe is equally important – those with a descriptive aim. Our analysis was intended to map out and quantify, with unprecedented power, characteristics associated with people in the general population dying from COVID-19 [3]. Multivariable descriptions like this are, in our view, hugely useful in building a knowledge base about a novel disease. Key understandings about COVID-19 have already arisen from multivariable descriptive analyses, such as the growing realization that thrombosis is a key pathology in COVID-19. Additionally, descriptive studies help inform the planning of subsequent predictive and causal studies.

As the pandemic evolves, risk prediction models will be needed to help guide policy decisions. Our paper does not present a risk prediction tool; however, our findings will help researchers to prioritise which data should be acquired to develop such a tool, and to determine sample sizes needed.

Our results can also help formulate causal hypotheses and design causal studies. For example, our identification of higher risks among BAME groups highlights this as a priority area for causal study and has already led to hypotheses about the relative importance of exposure to infection versus genetic factors. Our results can also help epidemiologists to assess whether their posited causal diagrams and theories are consistent with observation. Causal analyses of observational data require extreme care, need focus on one factor at a time, and need to take into account the totality of the evidence ranging from molecular biology through to global health policy in their interpretation. And even with the best available causal methods, one must accept that we will likely never know the extent to which all confounding has been overcome.

Both sets of correspondents query our comments about the implications of our results for important causal questions, particularly regarding ethnic disparities in COVID-19 death. Our results identify several priority areas - including this - for subsequent causal study. In the meantime, policy decisions must be as informed as possible. Informed conjecture about the implications of our results for such questions, alongside calls for properly conducted causal analyses and cautions regarding study limitations, seems entirely appropriate in a discussion section, given the current and urgent relevance of these issues.

Our paper presents the most powerful assessment to date of factors associated with Covid-19 death for a wide range of factors. It was delivered swiftly, very early in the pandemic, just 7 weeks after global daily deaths breached 1,000. Such descriptive analyses are not the end of the COVID-19 story; they are rather the beginning. We look forward to research groups, including ourselves, using our results to inform the development of validated risk prediction tools and specific causal studies.

REFERENCES

[1] Westreich, D., van Smeden, M., Edwards, J.K. Comment on Williamson et al. (OpenSAFELY). Epidemiology (2020).

[2] Tennant, P.W.G., Murray, E.J. The quest for timely insights into COVID-19 should not come at the cost of scientific rigour. Epidemiology (2020).

[3] Williamson, E.J., Walker, A.J., Bhaskaran, K. *et al.* OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2521-4>