Reasons for retraction of Silva et al.

1 The main result is demonstrably spurious

The authors use a two-way fixed effects model to estimate the effect of masking on the excess mortality and on the cumulative excess mortality in the subsequent weeks. More precisely, for every week t between January 1st 2020 and January 2nd 2022, the coefficient β of their model measures the effect of the average masking on week t on the excess mortality on week $t + d$ and on the cumulative excess mortality between January 1st 2020 and week $t + d$, where the lag d ranges from 1 to 4 (see https://github.com/owid/covid-19-data/tree/master/public/data for the definition of the cumulative excess mortality variable). Point estimates of β and corresponding confidence intervals are reported in their Supplementary Table 1, which is reproduced below for convenience.

Table 1. Reproduced from Silva et al.

Lag	Weekly excess mortality β (95% CI)	Cumulative excess mortality β (95% CI)
1 week	0.0524 [−0.1008; 0.2057]	−0.0674 [−0.1270; −0.0077]
2 weeks	0.0426 [−0.1015; 0.1867]	−0.0688 [−0.1269; −0.0107]
3 weeks	0.0269 [−0.1107; 0.1644]	−0.0702 [−0.1263; −0.0140]
4 weeks	0.0134 [−0.1207; 0.1476]	−0.0713 [−0.1253; −0.0172]

If one wishes to study the effect of masking on mortality in a time dependent fashion, the natural choice of outcome variable is the weekly excess mortality a certain number of weeks after the masking, i.e., the excess mortality in the period in which the masking might have an effect. The cumulative excess mortality is a somewhat odd choice of outcome variable as it encompasses the excess mortality of a large (and varying) number of weeks before the masking. It seems that the authors understand that as, in their own words, cumulative excess mortality "*may introduce distinct biases by smoothing over heterogeneous epidemic phases*".
Note that when weekly excess mortality is used as outcome variable, the point estimates of the parameter β are positive for all lags. This is consistent with our findings that more masking is associated with more excess mortality. However, since the result is not statistically significant, one cannot infer much from this.
It is only when the cumulative excess mortality is used as an outcome variable that estimates of β become negative and statistically significant, indicating that the masking is reducing excess mortality. However, the observation of an effect with a lag of only one week (Supplementary Table 1 in the paper) is a red flag, since if masks reduce transmission, any impact on mortality would be expected to manifest only after a longer delay. Since estimates of the interval between cases and deaths vary in the literature (https://doi.org/10.1007/s43071-022-00027-6),

we decided to settle the matter by using the code supplied by Silva et al. to compute the estimates of the parameter β for lags equal to 0 and −1 (Table 2).

Table 2. Extended lag analysis

Lag	Weekly excess mortality β (95% CI)	Cumulative excess mortality β (95% CI)
−1 week	0.0438 [−0.1246; 0.2123]	**−0.0639 [−0.1251; −0.0027]**
0 weeks	0.0523 [−0.1093; 0.2140]	**−0.0658 [−0.1264; −0.0051]**
1 week	0.0524 [−0.1008; 0.2057]	−0.0674 [−0.1270; −0.0077]
2 weeks	0.0426 [−0.1015; 0.1867]	−0.0688 [−0.1269; −0.0107]
3 weeks	0.0269 [−0.1107; 0.1644]	−0.0702 [−0.1263; −0.0140]
4 weeks	0.0134 [−0.1207; 0.1476]	−0.0713 [−0.1253; −0.0172]

As Table 2 shows, estimates of β and corresponding confidence intervals obtained for lags of 0 and −1 are similar to the estimates obtained for lags between 1 and 4 weeks. **Since it is extremely unlikely that masking in a given week would have an effect on cumulative excess mortality on that same week and it is impossible that masking in a given week has an effect on excess mortality on the preceding weeks, this conclusively demonstrates that the effects reported by Silva et al. are an artifact created by including pre intervention data in the outcome variable. Hence, the effects reported by Silva et al. are spurious and not true causal effects of masking on excess mortality. This renders their analysis invalid.** In addition, we show below that important assumptions of the model used by Silva et al. are violated. This might explain why the model produces a spurious effect of masking on excess mortality during the pre-masking period.

1.1 The assumptions underlying the model are invalid

The two-way fixed effects model used by Silva et al. has the following form

$$Y_{it+d} = a_i + b_t + \beta M_{it} + \varepsilon_{it},$$

where Y_{it} is the outcome variable (either excess mortality or cumulative excess mortality) for country i on week t, d is the fixed chosen time lag between 1 and 4 weeks, a_i is a parameter that represents country specific time independent effects, b_t is a parameter that represents time specific country independent effects, β is the parameter of interest, M_{it} is the average masking for country i on week t and ε_{it} is a random error term. In short, aside from the effect of masking and random noise, the outcome variable is the sum of a country specific time independent term a_i with a time specific country independent term b_t.

It is well known that the validity of the additive structure $a_i + b_t$ of the main effect is crucial for an adequate estimation of the parameters of interest in a two-way fixed effects model (http://www.doi.org/10.1017/pan.2020.33). In fact, it is easy to prove that if this additivity assumption fails then the usual OLS estimator for β is often biased and the bias does not tend to zero as sample size goes to infinity, i.e., the estimator is inconsistent.

In the case at hand, it is clear that the additivity assumption is unrealistic. At the very least, time lags between infection waves in different countries make the time effects country dependent. Moreover, country dependent effects are not in general expected to act additively. Small differences in viral reproductive numbers, for instance, would generate complicated non-linear effects on infection waves. Also, differences in age structure, general baseline health of the population and quality of health infrastructure would change the infection fatality rate generating multiplicative country specific effects.

In order to assess the validity of the model assumptions, we plotted residuals against time for each of the 24 countries. This is a standard graphical diagnostic for linear models: when the assumptions are satisfied, residuals should behave like random noise fluctuating around zero. In the present case, the opposite is observed. The 24 residual plots are shown in the accompanying file *residuals_cum_lag1.pdf*, where cumulative excess mortality is used as the outcome variable Y with a time lag $d = 1$. The same qualitative pattern appears for lags $d = 2, 3$, and 4. These plots provide conclusive evidence that the model assumptions are severely violated. **Had Silva et al. performed this basic diagnostic check, they would have seen that their model is fundamentally flawed**.

2 Incorrect conclusions derived from Figure 1d

One of the central claims advanced by Silva et al. is that countries exhibiting higher levels of mask usage did so in response to more severe pandemic waves. Figure 1d is presented as the sole piece of evidence supporting the assertion that mask adoption increased reactively following worsening outbreaks. As they state:

> *To further contextualise the findings, we included data from Brazil (Fig. 1d) showing the excess mortality (monthly) and mask usage (30-day moving average). The figure suggests that an increase in mask adoption typically occurs after peaks in mortality, indicating a reactive response to worsening outbreaks rather than a causal effect, contradicting Tausk and Spira's conclusions.*

However, in addition to the fact that the legends on the Y-axes are swapped - the orange curve with pronounced peaks corresponds to excess mortality, not mask usage, the blue curve, which represents mask usage, shows no meaningful increase across pandemic phases. For instance, mask usage during the second wave, peaking in April 2021, was virtually identical to that during the first wave in June 2020, remaining near 70 percent, despite the second wave being far more lethal, with excess mortality of approximately 80 percent compared with about 27 percent in the first wave. Once the colours are correctly assigned, this contradiction becomes immediately apparent. The data presented by Silva et al. therefore refute, rather than support, their claim of reverse causality, undermining the credibility of their interpretation.

3 Misrepresentation of a key reference (ref. no. 8)

Silva et al. state that

> *Substantial evidence shows that masks, especially high-grade masks used correctly, are effective barriers to viral transmission or, at worst, do not increase infections or mortality.*[6–11]

One of the references cited in support of this claim is the Cochrane review on non-pharmaceutical interventions (ref. no. 8). However, the Cochrane authors explicitly concluded that:

> *The pooled results of RCTs did not show a clear reduction in respiratory viral infection with the use of medical-surgical masks. There were no clear differences between the use of medical-surgical masks compared with N95-P2 respirators in healthcare workers when used in routine care to reduce respiratory viral infection.*

It is therefore unclear how this review could be cited as evidence of mask effectiveness. Moreover, the conclusions regarding masks are qualitatively identical to those of the previous Cochrane review published in late 2020 (https://doi.org/10.1002/14651858.CD006207.pub5), which reached essentially the same assessment. It is understandable why Silva et al. sought to include a Cochrane review among the studies purportedly supporting their thesis, given that Cochrane reviews are internationally recognised as the gold standard for high-quality evidence on the effectiveness of healthcare interventions. However, by portraying the Cochrane review as supportive of mask use for reducing viral transmission, Silva et al. fundamentally mischaracterise its conclusions and thereby mislead readers about the actual state of the evidence.