Reviewer comments are in **bold** and the authors' responses are in blue.

We thank the reviewer for their evaluation of our manuscript and for raising several important concerns. We have taken these comments seriously and have made revisions throughout the paper. In particular, we have clarified the limitations of CESM-WACCM6, adjusted language that overstated the comprehensiveness of the chemistry representation, rounded and reformatted numerical results to avoid overprecision, and expanded the conclusion to provide more interpretation of the results. We appreciate the reviewer's concern regarding the length of the manuscript. While the overall message may appear conceptually straightforward, arriving at this conclusion requires careful and detailed analysis across multiple facets. We have thoroughly reviewed the manuscript and find that the content presented is necessary to support our conclusions rigorously. Below, we provide point-by-point responses to the reviewer's comments and describe the corresponding changes made in the revised manuscript.

Major comments

A key issue with this paper is the use of CESM2-WACCM6 for the evaluation of health impacts of stratospheric aerosol injection (SAI). Previous studies have relied on models with shortcomings noted in the introduction, but unfortunately the shortcomings of the model used here are downplayed. Main issues are with the fixed photolysis rates and lack of ammonium and nitrate aerosols, essential to air quality assessment. Without these terms, the author's conclusions are, at best, incomplete.

These are not minor caveats, but fundamentally constrain the reliability of this study, something the authors should be upfront about. Claims of "comprehensive chemistry" should therefore also be removed.

We thank the reviewer for raising this important point. We agree that CESM2-WACCM6 has limitations that need to be acknowledged when interpreting our results. In particular, 1) photolysis rates are calculated using lookup tables that account for overhead ozone column and clouds but that do not include direct effects of dynamic aerosol distributions, and 2) the MAM4 aerosol module does not represent ammonium and nitrate, which are important contributors to PM2.5 in many regions. These are important caveats, and we have revised the manuscript to emphasize them more clearly.

While "comprehensive" is a term often used in connection with WACCM (including on the WACCM model webpage: https://www2.acom.ucar.edu/gcm/waccm), we recognize that its use here could be misleading. We have therefore removed phrasing that described the model as "comprehensive" throughout the manuscript.

The use and interpretation of the 10-member ensembles in this manuscript raises several problems. The paper stresses variability across ensemble members yet lacks a systemic metric for uncertainty analysis. Without this, the role of ensembles remains descriptive, rather than analytical.

We appreciate the opportunity to clarify the interpretation of the 10-member ensembles. For geographical figures (Figs. 1-4 and 6), our focus is on the spatial patterns and whether simulated changes are robust relative to internal variability. In these cases, we use stippling based on a two-sided t-test across ensemble members at the 95% confidence level to indicate regions where results are not statistically significant. For Fig. 2, we are interested in identifying the dominant PM2.5 species for each geographical region, so we stipple over areas where fewer than 90% (9 out of 10) of the ensemble members agree on the dominant species. For Figs. 5, 7-9, we focus on summarizing the overall magnitudes and regional differences in mortality burdens. Here, error bars show ensemble spread, since these quantities are spatially averaged and the spread directly conveys the variability in the estimated mortality burden. Thus, uncertainty is presented in a manner appropriate to the scientific question addressed by each figure.

Simultaneously, there is the issue of overprecision. Results are given such as "-149,397 to -177,296" (line 227) which is indefensible when variability and uncertainty are ignored, as the authors admit to in lines 148-150. Results should be rounded and expressed as mean +- standard deviation or 90% confidence interval. Not as exact integers or with unrealistic significance. Unfortunately, variability is highlighted when it dilutes signal but downplayed when the results look robust. Such inconsistency weakens the conclusions.

We appreciate this observation and agree that reporting excessively precise integers can be misleading. In the revised manuscript, all mortality estimates have been rounded to avoid the appearance of overprecision. For uncertainty, we have chosen to present the ensemble spread (minimum-maximum across members) rather than a +- standard deviation of 90% confidence intervals. Our primary approach to uncertainty is to present the spread across ensemble members, which reflects the internal variability captured by the model. The following text has been added to make this clear:

"In the figure, this variability is quantified by the standard deviation across mortality estimates, while the text emphasizes the overall spread in the projections."

Furthermore, the statement that "mortality impacts do not scale with SO2 injection" is unsupported. Only two scenarios are compared, over a relatively short time period. A more nuanced treatment would recognize that non-linearity is plausible but cannot be demonstrated here.

We thank the reviewer for this comment. Our analysis is based on two distinct large ensembles (ARISE-SAI-1.5) and ARISE-SAI-1.0), which span a wide injection range– from near 0 up to ~20 Tg So2/yr– sufficient to encompass a range of cooling from present-day conditions to ~3 degrees C. Over the 50-year simulation period, these scenarios provide a reasonable basis for assessing differences in health outcomes under SAI as a function of injection rate. However, we have revised the manuscript to clarify that while our results suggest no scaling of mortality impacts of SO2 injection across these scenarios, a more systematic assessment of non-linear

responses would require additional scenarios beyond those available here (last sentence of Section 3.2).

The scope of this work is narrow. While the paper claims that "this study focused on the air quality-related health impacts of SAI", only ozone and PM2.5 are considered. The abstract (and title) should reflect the scope. Unfortunately, the paper glosses over the significant regional increases in mortality (Figure 9), these results deserve more emphasis, as focusing on global aggregates risks misinterpretation of the overall findings.

We thank the reviewer for raising this important point. PM2.5 and ozone are the two pollutants most commonly used to represent surface air quality in global health assessments (Pandley et al., 2019), as they account for the vast majority of air quality-related mortality (GBD (2019), WHO, 2016 & 2021)). Thus, by focusing on PM2.5 and ozone, our study focuses on air quality-related health impacts of SAI.

While many prior studies (e.g., Eastham et al., 2018 and Harding et al., 2024) have focused primarily on global aggregates, we sought to go further by including Section 3.3, which presents mortality burdens for each GBD super-region. This regional perspective highlights heterogeneity in outcomes, including regions where mortality increases under SAI, even though the paper does not explore each regional change in detail. To avoid misinterpretation, we have revised the abstract to explicitly note that our analysis is limited to ozone- and PM2.5-attributable mortality and that both global and regional changes are considered. We believe this provides sufficient context for readers without changing the overall framing of the paper. We have included figures showing the regional changes PM2.5 and ozone-related mortality from SAI in the Supplemental.

The discussion largely restates results rather than interpreting them. The reader is left with little beyond "impacts are modest".

We agree with the reviewer that the discussion should go beyond restating results. To address this, we have expanded the conclusion to interpret our findings in the broader context of climate change and air quality policy. Specifically, we highlight the impacts of SAI relative to internal variability and policy-driven improvements, and relate our results to the "climate penalty" literature. We now emphasize that white SAI alters the spatial distribution of ozone and PM2.5, the dominant determinant of future health outcomes remains the strength of air quality policies. This additional discussion clarifies that our main conclusion isn't simply that the impacts are modest, but they are modest relative to variability and policy effects (SSP2-4.5), underscoring the importance of emissions reductions for long-term air quality and health (see last paragraph of conclusion).

Minor comments:

Figres are dense, inconsistently referenced, and hard to interpret. At worst they are misleading (e.g., different scales across panels in Figure 3). Figure 5(b) is never referred to in the text. Remove this from the paper or discuss the meaning in the main body.

We thank the reviewer for this helpful feedback. Figure 3 has been revised so that all panels now show percent changes and each row has consistent percentage-change scales across the three panels. We have also reviewed and corrected all figure references to ensure consistency. In addition, Fig 5b is now explicitly discussed in the main text.

Check the citations. E.g., line 123: WHO cited as "(Organization et al., 2021).

The citation has been corrected.

Line 16: I strongly recommend against using uncommon words like "ameliorate".

The word "ameliorate" has been replaced with the word "offset"

Line 47&48: awkward use of "they" to refer to Harding et al., I suggest referring to the studies instead of the authors when critiquing methods used.

In Line 47&48, the pronoun "they" refers to solar dimming simulations rather than to Harding et al. To avoid ambiguity, we have revised the sentence to explicitly state "solar dimming approaches" instead of "they.

Line 54: "the the"

The extra "the" has been removed from the text.

Line 158: "these three-way comparison".

The text has been revised to refer to "this three-way comparison" rather than "these three-way comparison"

Line 184: this is phrased rather unprofessionally: I suggest replacing ", as in" with: "i.e.,".

The text has been revised accordingly.

Bibliography

GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 19902019: a systematic analysis for the Global Burden of Disease Study 2019 Lancet 2020; 396: 1223–49.

Pandey, A., Brauer, M., Cropper, M. L., Balakrishnan, K., Mathur, P., Dey, S., ... & Dandona, L. (2021). Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. The Lancet Planetary Health, 5(1), e25-e38.

World Health Organization. (2016). Ambient air pollution: A global assessment of exposure and burden of disease. Clean Air Journal, 26(2), 6-6.

World Health Organization. (2021). WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.