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

We thank the reviewer for their thorough and constructive evaluation of our manuscript. The comments have been invaluable in helping us clarify the scope of our study, better communicate the limitations of our modeling framework, and improve the presentation of the results. In revising the manuscript, we have explicitly noted caveats regarding missing processes in the model (e.g., fixed tropospheric photolysis, absence of ammonium/nitrate aerosols), clarified which conclusions apply specifically to the ARISE scenarios, and removed language that overstated the comprehensiveness of the model. We have also improved the abstract and figures, added references and explanations where needed, and corrected wording and formatting issues. We have also implemented many of the text, wording, and grammar edits that have been suggested by the reviewer. 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.

This paper claims that it is better than previous work because it includes a more comprehensive treatment of the climate system. But it needs to make clear right at the beginning what it does not include. There is no treatment of UV changes (now possible with TUV incorporated in WACCM) and tropospheric chemistry does not include changes in photolysis. So all the conclusions have to be tempered by these omissions, and this has to be made clear in the abstract. The abstract focuses on SAI impacts, but that is not correct. This specific SAI scenario, with more forcing in the SH, produces direct effects there. So the results here are not general results for SAI, and that also needs to be made clear in the title and the abstract.

We thank the reviewer for this important clarification. We realized that stating that the model does not include photolysis changes is incorrect. The photolysis rates are calculated using lookup tables that take into account the overhead ozone column and clouds (Emmons et al., 2020; Kinnison et al., 2017). However, this approach does not include the direct effects on actinic flux from aerosol scattering and absorption. In response, we have revised the abstract to acknowledge the limitations of our study more explicitly. We also emphasize that our analysis is specific to the ARISE-SA-1.5 scenario, which applies stronger forcing in the Southern Hemisphere.

We would also like to point out that even though a full interactive TUV scheme is now being incorporated into WACCM (as opposed to look-up table approach), it has only just been released and was not available at the time these CESM2(WACCM) simulations were carried out. Hence, if one wanted to use the new version, one would need to re-run both the 10-member ensemble of the control SSP2-4.5 simulations and all the SAI simulations, which is beyond the scope of the current study.

It is certainly not correct to present results in the abstract to three significant figures without error bars, particularly since 10-member ensembles were used here.

In the revised abstract, we now report results to two significant figures rather than three. In addition, we have included the ensemble range of PM2.5 and ozone-related mortality estimates to better reflect the uncertainty across the 10-member ensemble.

In the title → "Small" → as compared to what?

We appreciate the reviewer's request for clarification. By "small," we mean the additional air quality-related mortality impacts of SAI relative to those projected under the baseline SSP2-4.5 scenario. As shown in the abstract, the percentage changes in air-quality–related mortality under SAI are comparable to the magnitude of the changes projected under SSP2-4.5 alone.

L 58: "comprehensive" → NO. It does not account for changes in UV due to ozone depletion, and its effects on ozone chemistry in the stratosphere or troposphere.

"Comprehensive" is a term often used in connection with WACCM (including WACCM page: https://www2.acom.ucar.edu/gcm/waccm). We also would like to point out that the reviewer is mistaken, and that the current photolysis does take into account the overhead ozone column and, as such, can represent changes in UV from ozone depletion. We have therefore removed phrasing that described the model as "comprehensive" throughout the manuscript.

L 74: How well does this model simulate PM2.5? Ozone? Please show evaluations of the control simulations as compared to observations. If not well, this model should not be used.

We appreciate the reviewers' request to demonstrate how well CESM2-WACCM6 simulates PM2.5 and ozone in the control climate. Our simulations use CESM2–WACCM6 with the MOZART-T1 tropospheric chemistry scheme and the MAM4 aerosol module, which have been extensively documented and evaluated in prior work. Gettlemen et al. (2019) showed that WACCM6 can reproduce observed climatology of trace constituents, in particular ozone, in the middle atmosphere. They concluded that it was capable of reproducing the evolution of ozone in the 20th and 21st centuries. Tilmes et al. (2019), Emmons et al. (2020) and Schwantes et al. (2020) further describe and evaluate the new chemistry mechanisms in CESM2(WACCM6) using both previous model versions and observations such as the NASA ATom aircraft mission, surface ozone data from the Tropospheric Ozone Assessment Report (TOAR) and carbon monoxide from the Measurements of Pollution in The Troposphere (MOPITT), finding good agreement with ozonesonde data and seasonality of surface ozone, while finding some spatial biases in some specific regions. Furthermore, Griffiths et al. (2020) benchmarked CMIP6 models, including WACCM, against observed tropospheric ozone distributions and found overall agreement in the spatial and seasonal variability. For aerosols and PM2.5, Hancock et al. (2023) used WACCM to study air quality over India and found that while the model underestimates PM2.5 in some regions due to missing secondary species (i.e., ammonium), the model

reproduces the observed spatial patterns and long-term trends. Taken together, these evaluations demonstrate that CESM2–WACCM6 reproduces the global and regional distributions of ozone and aerosol species with skill comparable to or better than previous CESM versions, and within the performance range of current state-of-the-art chemistry–climate models.

In response, we now state more clearly in the methods and conclusion sections the explicit caveats of WACCM6 while also emphasizing that our conclusions rely on scenario differences rather than absolute concentrations.

L 106: "For PM2.5" → Doesn't it depend on the chemistry of the particles and not just their concentration?

The toxicity of PM2.5 can vary depending on its chemical composition and source, and there is active research on quantifying such differences (Lippmann et al., 2013; Stanek et al., 2011). In this study, we follow the Global Burden of Diseases (GBD) framework, which applies a concentration–response relationship for PM2.5 exposure that does not distinguish by composition. While this represents a simplification, it enables consistency with widely used health impact assessments.

L 118: What are each of these physically? What are the units?

The parameters θ, α, μ, and *v* are not physical quantities with units, but rather empirical coefficients of the Global Exposure Mortality Model (GEMM) function that defines the concentration–response relationship between PM2.5 exposure and relative risk of mortality. These parameters are estimated by fitting the IER function to epidemiological data for each cause of death and age group, and they are dimensionless. Their role is to shape the curve of the exposure–response relationship (e.g., slope, curvature, and inflection point) rather than representing a physical process. Table 1 serves as a reference for the specific parameter values.

Equation 2 has multiple variables that are not explained. What do each of them mean? What are the units? And what is the science behind this equation? Furthermore, where does the equation come from? What is the reference?

We thank the reviewer for pointing this out. In the revised manuscript, we have added explicit descriptions for all variables in Eq. 2, including units (e.g., PM2.5 concentration and relative risk (RR). The scientific basis for equation 1 is that it follows the standard health impact function commonly used in air pollution epidemiology, which relates changes in pollutant concentrations to attributable health impacts. This formulation is consistent with the methodology presented in the 2015 EPA BenMAP User's Manual (US EPA), which we now cite directly in the text.

How can OSMDA8 (the highest daily 8-hour average ozone concentration during the ozone season) be important for mortality? Shouldn't the impact of ozone on mortality be

the amount of ozone times the exposure? What if there are many days in a season with a little less ozone, and hence a lower OSMDA8, and in a different season only one day with a high OSMDA8 value and all the other days very low? Wouldn't the first case be worse for health? Please explain why the metric you are using makes sense.

OSMDA8 is a common ozone metric used for human health in recent studies, such as in the Global Burden of Disease, an international effort, as well as many other studies (Malashock et al., 2022; Murray et al., 2024; Sun et al., 2024). In order to match the methods used in these newer studies, we have decided to use it in this study as well.

OSMDA8 is not the highest daily 8-hour average ozone concentration during the ozone season, but rather the *highest 6-month rolling average* daily 8-hour average ozone concentration. The text has been revised to reflect this. This means that the OSMDA8 reflects the highest average MDA8 over a 6-month period, and is a good metric to use for long-term exposure to elevated ozone levels. The typical ozone season is from March to August in the Northern Hemisphere, and from September to February in the Southern Hemisphere, corresponding to their respective Spring and Summer months, but the 'ozone season' paradigm breaks down in the Tropics. The method of checking, for each grid box, where the 6-month rolling average MDA8 is maximized, allows for a better reflection of ozone exposure risk in the Tropics than just a blanket 'March to August North of the Equator and 'September to March' South of the Equator.

Also, how much would ozone exposure affect mortality as a function of time over a person's lifetime? Does it matter at what age they are exposed?

Ozone exposure, similar to PM2.5, would likely have different RR values for different age groups. However, due to the lack of data on age-specific RR values, the Global Burden of Disease uses a single value for all adults (Murray et al., 2020). This assumption of a constant RR value with age has since been used in other studies as well (Malashock et al., 2022; Murray et al., 2024; Sun et al., 2024). However, while the RR (i.e., the concentration–response function linking pollutant exposure to mortality risk)) is constant across age groups in our study, the baseline mortality rates (BMR) we apply are age-specific. For ozone, BMR refers specifically to mortality from respiratory and cardiovascular diseases, which determines how the exposure–response relationship translates into actual deaths. BMR for respiratory and cardiovascular diseases are much higher for older adults compared to younger adults. Thus, for the same RR value, premature deaths from respiratory and cardiovascular diseases due to ozone exposure are still higher for older populations.

Lines 148-150. You can't just choose to ignore uncertainty that you know about. This will give you the wrong answers. This just reinforces that the numbers in the abstract to 3 significant figures and no error bars can't possibly be correct.

We thank the reviewer for this important point. Our analysis is designed to assess relative differences in mortality outcomes between scenarios using a consistent set of SSP2-4.5 inputs, rather than provide absolute projections with full quantified uncertainty bounds. We agree that

reporting numbers to three significant figures without error ranges overstated the precision of our results. In response, we have rounded all mortality estimates to two significant figures in the abstract.

Fig. 3 has multiple issues:
- You have to use the same scales for all the panels in each row, like you did for Fig. 1, so that they can be compared. Otherwise the same color means different things in each panel.
- The shading in Figs. 2-3 is hard to make out, as only two colors are used, and the boundaries between the different values are not clear. Use distinct different colors.
- Needs stippling like in Fig. 1 to indicate which results are significant.
- Is it height above sea level? How can you have values under the ice in Antarctica?
- Mark the latitude in more increments, and use natural ones, every 15 or 30 degrees.
- The font in the figures is too small to see.
- You plot water concentration, but is it water vapor or total water, including liquid and solid? If water vapor, you have to use normal meteorological units of mixing ratio or absolute humidity. And you show large changes in the Tropics, but the ITCZ has a large seasonal cycle and spatial variations. Showing zonal-mean annual-mean values obscures much of the signal.

We thank the reviewer for these detailed suggestions. Figure 3 has been revised to address the concerns raise: i) all panels within each row now share the same colorbar scale, enabling direct comparison, ii) stippling has been added, consistent with Fig. 2, to indicate regions that are not statistically significant across ensemble members, and iii) latitude increments are now shown at 30 degree intervals, and font size has been increased for readability. Regarding the plotted variable in Fig. 3j, k, and l, the figure shows the percentage change in water vapor concentration.

We agree that the ITCZ exhibits a strong seasonal cycle and spatial variability, and acknowledge that zonal-mean, annual-mean values inevitably obscure aspects of this signal. Because our health impact analysis is based on annual, population-weighted concentrations of OSMDA8 and PM2.5, the zonal-mean changes presented here are the most relevant for our study design. To provide additional context, we now include seasonal zonal-mean figures of water vapor mixing ratio (% changes) in the Supplementary Information to show how seasonal variations compare to the annual-mean changes, while keeping the main analysis focused on the health outcomes.

H2O levels [What does it mean? Is it just water vapor?] Why would surface cooling produce this, in the ignorance of other factors, including circulation changes? —> "This reduces chemical ozone loss in the free-troposphere, as indicated by an increased net photochemical ozone production (Fig. 3c)."

By 'H2O levels', we are referring to water vapor concentrations. Surface cooling would reduce ambient water vapor concentrations generally due to the Clausius-Clapeyron relationship. While

there could be regional changes in water-related quantities such as precipitation and soil moisture, the relationship between ambient water vapor concentrations and surface temperature are robust and identified as a mechanism that influences tropospheric chemistry under SAI (see for instance, Visioni et al., 2017).

Explain how this happens: "This reduces chemical ozone loss in the free-troposphere, as indicated by an increase in net photochemical ozone production". What does water have to do with it?

In daytime conditions, a primary route of ozone loss is via gas-phase photochemistry. Ozone (O_3) can undergo photodissociation reactions to generate energetic oxygen radicals ($O(^1D)$), which in turn can react with water vapor to generate hydroxyl radicals (OH).

$O_3 + hv \rightarrow O_2 + O(^1D)$
$O(^1D) + H_2O \rightarrow 2OH$

In unpolluted regions (aka regions with low nitrogen oxide ($NOx = NO + NO_2$) concentrations, such as the remote troposphere), further reactions between O_3 and OH can lead to catalytic O_3 losses:

$OH + O_3 \rightarrow HO_2 + O_2$
$HO_2 + O_3 \rightarrow OH + 2O_2$
Net: $2O_3 \rightarrow 3O_2$

The most recent IPCC AR6 report (Szopa et al., 2021) states with high confidence a positive relationship between increasing surface temperatures and increased O_3 loss via increasing water vapor concentrations (which, in turn, occurs due to the Clausius-Clapeyron relationship) over unpolluted areas. The reverse is also true: cooling reduces water vapor, weakens this pathway, and thereby reduces chemical ozone loss (Visioni et al., 2017; Bednarz et al., 2023). This appears as an increase in net photochemical ozone production in our simulations.

L 186: "modifications in photolysis rates due to changes in stratospheric radiation…" What does this mean? In the stratosphere only? What wavelength radiation? Is it temperature dependent reactions? And you ignore changes in UV.

We thank the reviewer for pointing out this contradiction. We have now rephrased this to read: "modifications in photolysis rates due to changes in actinic flux from changes in the overhead ozone column and aerosol absorption and scattering"

As mentioned, photolysis rates are included in the model but they are pre-calculated using lookup tables that take into account overhead ozone column and clouds However, there are limitations with this approach in that they do not include changes caused by aerosol absorption and scattering, which we have now included in the text.

In addition, we have performed preliminary offline analyses of surface UV radiation differences between ARISE-SAI-1.5 and SSP2-4.5 using the Tropospheric Ultraviolet and Visible (TUV) model developed at NCAR to explore TUV changes in response to aerosol scattering and absorption. These results, now included in the Supplementary Information, indicate small changes but highlight a promising direction for further investigation.

Figure A4: Text is much too tiny to read. Make the panels much bigger and use fewer per row. Since the color bar is the same for all the panels, get rid of the small ones and just use one large one. And what are the ////? It looks like Fig. A4 was done with GrADS, and it looks much better than the others, with better labeling of the axes and distinct colors for the shading. But why is there no white box behind each number in the contour labels, so they can be more easily read?

Figure A4 and other figures in the Appendix are more appropriate for the supplemental materials and have been moved to the supplement. We have also replotted Figure A4 taking the reviewer suggestions into account.

Figure 4: It is really hard to compare the two columns, as they need to be plotted with the same color scale. But it looks like the values in the left column for some countries like India and larger than the standard deviation. So why are they indicated as being significantly different from zero?

Figures 4 and 6 have been replotted with the same color scale. The color scales now show that the values on the left column are larger than the standard deviation. The stippling is added over countries where the country's mean change is not statistically significant across ensemble members at 95% confidence interval.

The paper uses "notably" randomly. These should all be deleted. Every sentence should be noted or it should not be in the paper.

The use of "notably" and derivations of the word have been removed from the text.

The paper references Fig. 7 before Figs. 5 and 6. This is confusing. Figures have to appear in numerical order in a paper.

References to Fig. 7 in the first paragraph of Section 3.2 have been removed.

Bibliography

Adam, O., Bischoff, T., & Schneider, T. (2016). Seasonal and interannual variations of the energy flux equator and ITCZ. Part I: Zonally averaged ITCZ position. Journal of Climate, 29(9), 3219-3230.

Bednarz, E. M., Visioni, D., Kravitz, B., Jones, A., Haywood, J. M., Richter, J., ... & Braesicke, P. (2023). Climate response to off-equatorial stratospheric sulfur injections in three Earth system models–Part 2: Stratospheric and free-tropospheric response. Atmospheric Chemistry and Physics, 23(1), 687-709.

Emmons, L. K., et al. (2020). The Chemistry Mechanism in the Community Earth System Model Version 2 (CESM2). Journal of Advances in Modeling Earth Systems, 12(4), e2019MS001882. https://doi.org/10.1029/2019MS001882

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., ... & Randel, W. J. (2019). The whole atmosphere community climate model version 6 (WACCM6). Journal of Geophysical Research: Atmospheres, 124(23), 12380-12403.

Griffiths, P. T., Murray, L. T., Zeng, G., Shin, Y. M., Abraham, N. L., Archibald, A. T., ... & Zanis, P. (2021). Tropospheric ozone in CMIP6 simulations. Atmospheric Chemistry and Physics, 21(5), 4187-4218.

Hancock, S., Fiore, A. M., Westervelt, D. M., Correa, G., Lamarque, J. F., Venkataraman, C., & Sharma, A. (2023). Changing PM2. 5 and related meteorology over India from 1950–2014: a new perspective from a chemistry-climate model ensemble. Environmental Research: Climate, 2(1), 015003.

Lippmann, M., Chen, L. C., Gordon, T., Ito, K., & Thurston, G. D. (2013). National Particle Component Toxicity (NPACT) Initiative: integrated epidemiologic and toxicologic studies of the health effects of particulate matter components. Research Report (Health Effects Institute), (177), 5-13.

Malashock, D. A., Delang, M. N., Becker, J. S., Serre, M. L., West, J. J., Chang, K. L., ... & Anenberg, S. C. (2022). Global trends in ozone concentration and attributable mortality for urban, peri-urban, and rural areas between 2000 and 2019: a modelling study. The Lancet Planetary Health, 6(12), e958-e967.

Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., ... & Borzouei, S. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The lancet, 396(10258), 1223-1249.

Murray, C. J. (2024). Findings from the global burden of disease study 2021. The Lancet, 403(10440), 2259-2262.

Schwantes, R. H., Emmons, L. K., Orlando, J. J., Barth, M. C., Tyndall, G. S., Hall, S. R., ... & Bui, T. P. V. (2020). Comprehensive isoprene and terpene gas-phase chemistry improves simulated surface ozone in the southeastern US. Atmospheric Chemistry and Physics, 20(6), 3739-3776.

Stanek, L. W., Sacks, J. D., Dutton, S. J., & Dubois, J. J. B. (2011). Attributing health effects to apportioned components and sources of particulate matter: an evaluation of collective results. Atmospheric Environment, 45(32), 5655-5663.

Sun, H. Z., Van Daalen, K. R., Morawska, L., Guillas, S., Giorio, C., Di, Q., ... & Archibald, A. T. (2024). An estimate of global cardiovascular mortality burden attributable to ambient ozone exposure reveals urban-rural environmental injustice. One Earth, 7(10), 1803-1819.

Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., Collins, W. D., ... & Zanis, P. (2021). Short-Lived Climate Forcers (Chapter 6).

Tilmes, S., Hodzic, A., Emmons, L. K., Mills, M. J., Gettelman, A., Kinnison, D. E., ... & Liu, X. (2019). Climate forcing and trends of organic aerosols in the Community Earth System Model (CESM2). Journal of Advances in Modeling Earth Systems, 11(12), 4323-4351.

Visioni, D., Pitari, G., Aquila, V., Tilmes, S., Cionni, I., Di Genova, G., & Mancini, E. (2017). Sulfate geoengineering impact on methane transport and lifetime: results from the Geoengineering Model Intercomparison Project (GeoMIP). Atmospheric Chemistry and Physics, 17(18), 11209-11226.

US Environmental Protection Agency. (2018). Environmental Benefits Mapping and Analysis Program: Community Edition (BenMAP-CE) User Manual and Appendices.