DISSERTATION

QUANTITATIVE HEALTH IMPACT ASSESSMENTS AS A TOOL FOR EXPLORING PUBLIC HEALTH DIMENSIONS OF ENVIRONMENTAL EXPOSURES

Submitted by

Daniel Dean

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: David Rojas-Rueda
Co-Advisor: G. Brooke Anderson

Jennifer Peel
James Hurrell

ABSTRACT

QUANTITATIVE HEALTH IMPACT ASSESSMENTS AS A TOOL FOR EXPLORING PUBLIC HEALTH DIMENSIONS OF ENVIRONMENTAL EXPOSURES

Public health is influenced by a population's built and natural environment in both negative (e.g., natural disasters or ongoing stress from heat) and positive (for instance, heat-moderating effects of vegetation) ways, as well as interactively with behavioral and social dynamics. One framing of policy priorities and urban resilience is a "triad" consisting of exposure reduction (limiting the extent to which community members are exposed to environmental hazards—including "ambient" ones like stressful temperatures), vulnerability reduction (mitigating the impacts of sustained hazards), and hazard reduction (actively reducing the frequency or intensity of hazards) (Hoegh-Guldberg et al. 2018). Because any such measures carry tradeoffs in financial and other resources, it is important that policymakers and other stakeholders weigh comparative benefits of potential environmental hazards or interventions with consistent, quantifiable metrics. In this body of work, we applied quantitative health impact assessments, an epidemiology framework that provides a valuable tool here, allowing researchers to project health outcome changes for a population of interest given predicted changes in a relevant exposure and using epidemiological evidence, including exposure-response functions (exposure-response functions), which link exposure and health outcomes. In this body of work, we use HIAs to explore three different resilience-relevant systems spanning a range of intervention types, environmental systems, and spatiotemporal scales:

Project 1: Health Impacts of Future Tropical Cyclones in the Eastern United States: While tropical cyclones are among the most damaging natural disasters faced by the United States, the temporal and spatial rarity of these events impedes traditional frequency-based estimates for public health and related risk projections, leading to potential oversights in risk characterization. In addition, mortality associated with tropical cyclones may not be readily apparent between delayed onset and indirect causes (e.g. stress, disrupted medical care, infections), meaning that immediate mortality counts often underestimate full attributable mortality. In this project, we performed a pilot quantitative health impact assessment designed to address aspects of these limitations. First, we tested extending the historical tropical storm dataset using a pool of 10,000 simulated, or "synthetic" tropical cyclone seasons from the widely used and open-source STORM algorithm, trained from and intended to represent the "gold standard" of

historical International Best Tracks Archive for Climate Stewardship (IBTrACS) data. To the extent that STORM represents real-world conditions, this vastly expanded 'sample population provided information on potential tropical cyclone exposure risk than would be possible from historical data alone. For the second challenge of accounting for delayed and indirect attributable mortality, we combined the synthetic data with a recently developed exposure-response function: an integrated Bayesian causal-predictive model trained on Medicare Claims data (simplified to Americans aged 65 and older), featuring an integrated model approach to combine a whole-population causal inference model for central trends with county-specific predictive models for give county-specific estimates. This model also tracked up to 21 days of "lag time" in health outcomes after a TC exposure to capture delayed mortality. This combination of methodologies promises a comprehensive, county level picture of tropical cyclone-associated all-cause mortality risk among older adults. This approach provided insights including regions of the country at the greatest risk for tropical cyclone-related exposures among older adults. However, as our study represented a new application of the STORM algorithm (in particular, our emphasis focusing on post-landfall behavior of tropical cyclones), we also assessed the level of agreement between STORM and the historical dataset, finding some discrepancies including lower overall frequency, and considerably 'smoother' spatial distribution in exposures; some discrepancies were in line with previously noted limitations. This project used recent innovations in atmospheric science and epidemiology modeling to explore the utility of a quantitative health impact assessment framework for present-day risk and could inform policy and planning decisions in terms of tropical cyclone preparedness and response measures.

Project 2: Health impacts of Urban Tree Canopy policy scenarios in Denver and Phoenix: We explored potential health impacts (in terms of all-cause mortality, stroke, and dementia) of standing policy goals in Denver, Colorado and Phoenix, Arizona, for increasing the urban tree canopy coverage in these relatively arid cities. We projected health benefits (in terms of reduced attributable all-cause mortality, stroke, and dementia incidence) at a census block group level using several existing exposure-response functions based on the widely used Normalized Difference Vegetation Index (NDVI). Because the cities expressed policy goals in terms of percentage urban tree canopy, we generated predictive models to "translate" between this metric used in policy goals and the NDVI metric. We modeled the public health impacts of proposed real-world policies for near-future policy interventions in the form of increasing urban tree canopy, using current populations, and modeling an "overnight" change in exposure, with policy scenario benefits modeled for populations in year 2020, rather than demographic projections

for the 2030 (Phoenix) and 2050 (Denver) dates in the policy goal timelines. We also considered socioeconomic dimensions by using the census-based Social Vulnerability Index to trace the equity of current UTC and NDVI exposures, as well as of potential benefits. We determined that each city could, by reaching its standing policy goals, could avert hundreds of all-cause mortality cases, with even a partial attainment scenario (halfway between current and desired UTC levels) having appreciable benefits, with roughly half the captured mortality prevention; with respect to equity of UTC access, more-vulnerable communities in the cities saw lower access to current canopy cover, and consequently greater potential per-capita benefits under successful intervention scenarios.

Project 3: Health Impacts of Future Temperature Extremes Under a Solar Climate Intervention Scenario. In this project, we explored potential all-cause mortality implications of a proposed climate intervention effort intended to counteract anthropogenic warming, modeling the years 2050-2060 under alternate climate scenarios. Specifically, we projected temperature-associated mortality under a stratospheric aerosol injection (SAI) intervention scenario, as well as a corresponding scenario of "middle-of-the-road" climate change. We used a study population of 65-and-older Americans in eight major US cities (Seattle, Chicago, New York, Philadelphia, Los Angeles, Phoenix, Houston, and Miami) spanning a range of local climates. We built our analysis on widely used models and the shared socioeconomic pathway platform, allowing our two scenarios to be compatible, differing only in the SAI intervention itself. We focused on two age groups (65-75, and 75+) to reflect elevated heat- and cold-associated mortality risks among this population, finding broadly similar trends between age groups. We explored city-specific exposure-response functions for the temperature-mortality association, using a widely used modeling, comparing the anticipated number of cold- and heat-related deaths under each scenario, and highlighted tradeoffs for either policy scenario, finding considerable heterogeneity in trends between these cities. To make our analysis more specific to the mid-21st century, we incorporated existing estimates for population growth and mortality rate changes based on the same climate modeling scenarios as the SAU exposure scenarios. We observed dramatic variability in minimum mortality temperatures and temperature-attributable mortality between cities and found that SAI was not associated with decisive reductions in all-cause mortality among either age group. While SAI did effectively reduce heat-attributable mortality, this effect was counterbalanced by lower cold-attributable mortality under the warmer, non-SAI scenario. This observation could help inform planning and resilience efforts as far as types of temperature-related stress under each scenario, as well as provide insights for larger cost-benefit analyses for the overall proposition of SAI.

Together, these projects demonstrated how quantitative health impact assessments can help form a methodological foundation for exploring epidemiology and resilience-relevant systems. The variety of projects covered demonstrated the utility of this methodology in a variety of spatial scales, ranging from census block groups (comparable to neighborhoods) in Project 2 to county-level characterizations of tropical cyclone-associated risk for much of the eastern United States in Project 1. We also explored a range of time periods, ranging from Project 1's focus on characterizing tropical cyclone risk representative of the past several decades (as represented by the STORM resampling algorithm), through our attempts to explicitly model mid-21st century populations and temperature-related mortality trends using both climate and demographic projections. The modularity of the quantitative health impact assessment framework enabled our projects to leverage of existing research and datasets for low-cost, comparatively rapid assessments, as well as to lay infrastructure for future research and introduce several specific innovations in their respective designs.

ACKNOWLEDGEMENTS

My experience in the Environmental Research and Health Sciences department has been interesting and rewarding, and I've appreciated the experience to learn and challenge myself in this context. I want to thank my advisor, Dr. David Rojas-Rueda, and co-advisor Dr. Brooke Anderson for their guidance, support, and generous sharing on their areas of expertise, especially for the design and implementation of health impact assessments, and instruction and guidance on R programming respectively. I also wanted to thank my other committee members, Dr. Jennifer Peel and Dr. Jim Hurrell, for their insightful questions and advice on the preliminary exam, and Dr. Hurrell and Dr. Lantao Sun for their advice and input on the One Health Institute Stratospheric Aerosol Injection project. Beyond my immediate committee members, I've had a great experience with ERHS faculty and peers and found it a very welcoming and open environment. I appreciate the time I've been able to spend here. I also wanted to thank my family for their support and interest during the process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1. Background

A population's public health can be powerfully shaped by its natural and built environment, which carries both risks (e.g. natural disasters, temperature extremes) and potential protective effects (e.g. ecological services like water and greenspace access) (Briggs 2008). While important even when considering ambient present-day conditions, this becomes especially important in the context of climate change and general projected environmental trends, in that environmental risks for public health will be influenced not only by novel events (e.g. extreme storms or temperatures), but synergistically with existing "baseline" exposures (e.g. seasonal storms, the urban heat island effect) and second-order effects like infrastructure and power disruptions (Nethery et al. 2020). While climate change-associated public health risks are interdependent and multifaceted stressors in coming decades, strategies to improve communities' ability to weather external shocks and recover quickly could likewise see synergistic benefits. The concept of "urban resilience" is invoked frequently in the context of climate change preparedness and adaptation, and should be intentionally cultivated by a population and its institutions (Leichenko 2011). Resilience is a multifaceted concept with a range of spatial and institutional scales, and mechanisms including policy implementation, physical infrastructure, financial systems, and personal behaviors. While it is an important component of community continuity under any circumstances, climate change poses an especially noteworthy challenge, carrying irreversible impacts in the present and near future, especially for already-vulnerable communities, including increases in extreme and unpredictable weather events setting in even at the 1.5°C warming target set in the 2008 Paris Agreement, which is to date well behind the prerequisite benchmarks (Hoegh-Guldberg et al. 2018). Societies will have to anticipate and address both novel, potentially high-intensity climate-linked risks and existing baseline stressors, with synergistic interactions; these can be divided into absolute, "hard" limits to adaptation (e.g. the integrity of ecosystems and baseline habitability in the face of processes like flooding) and "soft" limits (e.g. the ability of populations to economically weather new stressors) (Mechler et al. 2020).

In the context of climate change, this could encompass a variety of strategies. Strategies could include promoting preparedness for natural disasters by improving detection and response systems, reducing health burdens of more routine stressor through changes to the built environment or behaviors, or more drastically, mitigation

efforts to greenhouse gas emission mitigation efforts to limit future instability, or even direct climate interventions intended to modify earth systems (e.g. atmospheric reflectivity, carbon cycles) to counteract warming or other dimensions of climate change (Hoegh-Guldberg et al. 2018). Given this range of dynamics, policy researchers suggest that resilience measures should balance considerations for population vulnerability (e.g. the "intrinsic" ability of a population to weather and recover from shocks in terms of institutions, economic diversification, and similar factors), exposure reduction (the ability to detect and if necessary evade extreme weather events or similar hazards), and hazard reduction (the ability to actively reduce the incidence and severity of potential hazards) (ibid.). Unfortunately, any such measures will compete for economic and other resources, not only within this proposed triad of resilience components and alternate solutions (e.g. increased tree canopies vs artificial shading structures), but the many other costs needed to run a city or larger society. The Health Impact Assessment (HIA) framework provides a flexible set of epidemiology tools to support this kind of decision-making, with a systematic, replicable strategy for understanding positive or negative implications of scenarios like competing policy proposals, or natural disasters.

HIAs are used to estimate potential health implications of a socially relevant policy/project (e.g. construction, environmental regulations), or other change in exposure (e.g. a natural disaster), for a population of interest, ideally through a holistic lens considering vulnerable populations and other equity concerns (Harris-Roxas et al. 2012a). While originally envisioned largely as an extension of environmental impact assessments, focusing on potential positive and negative impacts of proposed policies and projects (Mindell et al. 2003), the framework extends to events like natural disasters or hazards, and in addition to projecting future impacts can estimate past or ongoing impacts of an event or policy of interest. Especially in a policy context, the HIA framework's steps include screening (determining whether an HIA is appropriate), scoping (establishing operational boundaries like the population and geographic region, exposure and outcome definitions), appraisal or risk assessment (some combination of interpretive methods and using data and dose-/exposure-response functions to characterize health effects on the population of interest), forming recommendations, sharing these with stakeholders, and ongoing monitoring (Harris-Roxas et al. 2012b). In this body of work, screening is implicit, so we focus mainly on the scoping through risk assessment steps (with recommendations being addressed in conclusions, and eventual publication and further research addressing the last components. HIA risk assessment components can include a mix of qualitative and quantitative methods, with the former being especially useful for hard-to-define exposures or

allowing community feedback and guidance (Harris-Roxas et al. 2012b). While qualitative approaches may explore the probability and impact of risks through symbolic scales (e.g. 1-5, "minor" through "severe" impact or the like), and even use these in calculations for determining e.g. priorities, they generally do not attempt to obtain exact count estimates for health or other outcomes, or quantify uncertainty (Burkov et al. 2018). This approach contrasts with more quantitative approaches, as used in our body of work, and discussed below.

We focus on quantitative health impact assessments, which use a quantitative, rather than qualitative, risk assessment for this portion of the HIA framework. Risk assessments quantitatively model the implications of a predetermined change in exposure to health outcomes in a population of interest as a standalone approach. Risk Assessments consist of hazard identification (determining or selecting relevant health outcomes), exposure assessment (identifying relevant exposure(s), operational definitions, and levels experienced by a population), exposure-response assessment (obtaining the relationship between the health outcome and exposure levels), and risk characterization (combining the previous elements to estimate total health outcomes in the population via the exposure) (Briggs 2008). To briefly distinguish quantitative health impact assessments from comparative risk assessments, which can address similar research questions, the latter compare counterfactual outcomes of multiple risk factors under a given level of exposure(s) but do not necessarily consider a specific mechanism for obtaining a given exposure level or social context to the same extent as an HIA (Campbell and Woodruff 2006); comparative risk assessments can also be used to estimate the relative contributions of a range of risk factors to a known outcome.

As far as relative merits of qualitative and quantitative methods (elements of which could be used synergistically), an advantage of quantitative methods is that implementations can be faster (by drawing on existing empirical models and data resources), the results can be more readily compared between methods, and it is possible to quantify and discuss the relative certainty of a given set of conclusions with more confidence (e.g., extrapolating uncertainty from model parameters compared to weighing the relative merit of potentially competing citizen surveys). On the other hand, choosing to rely on remote sensing or other preexisting data does impose a necessary degree of abstraction, obscuring potentially relevant community-specific knowledge and information. For a concrete example, in relying on census data, we must accept that unhoused or otherwise unrecorded (e.g., undergraduate students with a different primary residence) inhabitants of a block group or county are unlikely to be represented in impact estimates, with additional resources or methodologies needed to capture these populations. There could also

be more nuanced implications—for instance, modeled benefits of an intervention like increased green spaces may be modified by legal (e.g., property laws) or cultural (e.g. inclination to spend recreational time outside) considerations that could be difficult to capture or translate for a given community without at least prior qualitative knowledge, making it ideal to engage with community members and other stakeholders where feasible. It is also important to note that while quantitative health impact assessments return numeric estimates, the validity of these estimates depends on the consistency of the framing assumptions with real-world conditions, making it important to develop well-founded operational definitions and functions on relevant literature and ideally community feedback and context to inform e.g., relevant model variables.

The models also generally rest of the assumption of causality—that is, that outcomes of interest (like all-cause mortality) are causally downstream of exposure metrics (e.g. air temperature), with any relevant interactions, indirect effects, etc., either captured in the model function or conceded as model error. While this limitation is less of a concern in situations where there is a well-characterized mechanism linking an exposure and outcome (e.g. for situations like medications where thorough experimental studies are feasible), the environmental exposures addressed in our body of work, and many quantitative health impact assessments more generally, do tend to lean more heavily on this assumption, where in some cases readily-captured metrics are understood to be indicative of more granular mechanisms. For instance, in Project 2, we use percentage urban tree canopy and corresponding Normalized Difference Vegetation Index scores (a metric for, effectively, chlorophyll levels in an area) based on methodological uniformity and availability as remote sensing data, but the health benefits are understood to proceed via specific mechanisms like cooling properties to offset the urban heat island effect or a level of air particular removal (Rojas-Rueda et al. 2019), less readily-quantified effects like apparent stress reduction (Ward Thompson et al. 2012), and potentially even effects like beneficial microbiome influences (Flies et al. 2017). In effect, this health impact assessment treats tree canopy as directly causing the change in mortality observed in the previous studies we sourced our exposure-response functions from, but it effectively acts as an indicator variable for multiple mechanisms assumed to be effectively correlated with NDVI.

2. Objectives and Scope

Our primary objective in this body of research is to carry out three quantitative health impact assessments for effects of environmental exposures on all-cause mortality in United States populations, spanning multiple environmental systems, spatial scales, timeframes, and specific methodologies for exposure-response functions and

other health impact assessment elements. We present each project as a chapter in the format of an academic paper. To briefly summarize, quantitative health impact assessments provide a useful tool in the context of environmental risks by providing quantitative, evidence-based estimates of the relative health costs or benefits of a range of environmental stressors or proposed policy interventions. In this body of work, we aim to explore this framework by applying it to three different resilience-relevant systems spanning a range of action types, spatial and temporal scales, and exposures.

Our first project, arguably focusing on the exposure reduction portion of the "triad" proposed in Hoegh-Guldberg et al by promoting risk awareness and potential long-term preparedness for tropical cyclones, is a near-term, preparedness-oriented quantitative health impact assessment that models all-cause mortality risks among older adults to eastern United States counties of Tropical Cyclone (TC) activity using a pool of 10,000 years of resampled synthetic tropical cyclone tracks resampled from high-quality historical storm records (1980-2018) to provide a more comprehensive picture of tropical cyclone-associated risk than could be inferred only by looking at past events. Because high-quality historical tropical cyclone records only extend to around 1980 and tropical cyclones occur as comparatively rare, discrete events (compared to, e.g. spatially and temporally continuous exposures like heat or air quality), this historical record provides a limited basis for standard probability-based risk assessments, with the last several decades not necessarily being representative of long-term trends.

We explore a potential countermeasure to this limitation by leveraging output from a leading machine learning-based algorithm that resamples individual storm data from the high-quality International Best Track Archive for Climate Stewardship (IBTrACS) archive to generate novel but plausible tropical cyclone events (specifically, hurricanes and tropical storms; as will be discussed, this has some implications for cumulative estimated impacts), effectively augmenting the historical record and yielding a more comprehensive picture of possible tropical cyclone distributions under roughly current climate projections at a county resolution, providing greater specificity than would be available under e.g. global-scale climate and impact models (Meiler et al. 2022). In this project, we integrate 10,000 simulated hurricane seasons from this "synthetic" dataset of tropical cyclone tracks with an existing exposure-response function, which uses a Bayesian framework to integrate two model components: the first is a casual inference model linking tropical cyclone binary exposure and maximum sustained windspeed (as the primary exposure of interest) with county-level 1999-2015 Medicare claims data for (in our case) all-cause mortality statistics with a daily resolution, allowing insight into dynamics like delayed effects (over up to two

weeks), and the second is a linked predictive model incorporating these factors, in addition to a range of socioeconomic and demographic traits, as well as additional features of the tropical cyclone exposure, to provide weightings to modify outputs from the first model, as well as allowing for more accurate propagation of uncertainty (Nethery et al. 2020).

Unlike the following projects, Project 1's exposure-response function, being developed specifically for our population of interest, already implicitly contained the relevant mortality incidence statistics, generating risk results in terms of excess events per 100,000 residents, which represents a slight departure from the standard the quantitative health impact assessment approach insofar as not needing a separate source for baseline population mortality, although this does come at the cost of making the exposure-response function more specific to this population, and the overall quantitative health impact assessment less modular. We combined the simulated exposure data and exposure-response function with 2020 census results to provide a simulated profile of tropical cyclone-associated mortality profiles under present-day (or more accurately recent past) climate conditions. Here, as with the other projects, the quantitative health impact assessment framework also provides a flexible groundwork from which other related investigations could spring—for instance, exploring alternative climate scenarios, recurrent climate patterns like El Niño Southern Oscillation phases, or different health outcomes.

Our next project focuses on the vulnerability reduction aspect through standing policy goals for expanding the urban tree canopy (UTC) in Denver, CO, and Phoenix, AZ. By drawing on literature sources for exposure-response functions associating vegetative density as measured by the remote sensing metric Normalized Difference Vegetation Index (NDVI) with mortality, stroke, and dementia, our quantitative health impact assessment framework models the relative impacts of full or partial attainment of policy goals. Here, one challenge was developing a model to provide a bridge between the UTC metric used in the policy goals and the NDVI metric used the exposure-response functions we sourced from literature; the resulting model, which uses natural cubic splines to capture nonlinear effects as well as spatial autocorrelation to allow locations within the city to modify the association (an approximation of local climate and land use differences), let us translate the policy metrics in the terms compatible with the health outcome component of the HIA. We took a much more spatially granular approach in this project than the others, using remote sensing and census data to assess exposures and outcomes at the scale of census block groups (integrating 30-meter-resolution exposure data). Unlike Project 1, which tracks existing patterns of tropical cyclone-associated risk, in this project, we explored proposed near-future policy goals (although still

using a 2020 baseline for population levels and distributions), reflecting the capacity of quantitative health impact assessments to explore interventions in addition to natural exposures. We also modeled alternative UTC scenarios—halfway between the proposed and current UTC levels—to explore potential benefits of more modest canopy improvements considering current and future constraints in water and other resources.

Our final project focused on a much more ambitious proposed hazard reduction intervention—using the strategic release of sulfur dioxide in stratospheric aerosol injection (SAI) to increase the albedo of the atmosphere, alleviating the effects of the greenhouse gas effect. In this project, we returned to the county level, modeling the attributable fraction of all-cause mortality in eight major US metropolitan areas (modeled at the county level) during 2050-2060 either under unmitigated climate change following a "business-as-usual" scenario from the standardized Shared Socioeconomics Pathways (SSP; specifically SSP2-4.5), or under conditions modified by an intervention in the form of SAI, as represented by the Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI) model, which attempts to keep global average temperatures within 1.5°C of pre-industrial levels (Richter et al. 2022). Unlike the previous project, where we treated the UTC intervention as occurring essentially "overnight" for modeling purposes, we used NASA projections for the populations and age compositions of our counties of interest and census estimates for future mortality rates to better model the mid-21st century setting of the study. As with the first project, our exposure-response function in this case (using an implementation of the distributed lag nonlinear model (DLNM) framework developed by Antonio Gasparrini and others (Gasparrini 2011)) can account for lag times of up to three weeks after a given day's temperature exposure (an important consideration given the different etiologies of heat- and cold-related mortality). Because cities exhibit considerable adaptation to their local climates in terms of temperature-associated all-cause mortality (with e.g. hotter cities showing lower risks of heat-related mortality), we use a modeling framework adapted from Vicedo-Cabrera et al 2019 to generate city-specific exposure-response functions, rather than relying on single, "universal" models as in Project 2, or using a model previously developed specifically for a population of interest as in Project 1.

While these three projects vary widely in respects like the exposures, their spatial scale, and specific modeling methods, they follow a common quantitative health impact assessment framework for predicting health outcomes given information on a natural or artificial change in exposure, relevant traits for the population of interest, current rates of mortality incidence (or in the case of project 2, of neurological health outcomes), and an

existing or newly-derived function to track the association between changes in the exposure and changes in the outcome. Our projects also share a common grounding in exploring interactions between the environment and health, whether in the form of existing hazards like tropical cyclones, urban planning decisions like UTC growth, or larger modifications like the global-scale SAI explored in Project 3. In designing these projects, we aimed to use publicly accessible data (or at least standardized and widely used, as in the case of the formerly publicly accessible NMMAPS dataset for temperature-mortality associations in Project 3) to make the projects serve not only as investigations into these specific questions, but as frameworks transferrable to other contexts. However, the underlying data sources dictate the prerequisite degree of modification needed. For instance, our approach in Project 3 could be immediately applied to other cities in the NMMAPS data with no changes. Still, one would need an equivalent source of daily temperature and mortality, or other health outcome figures would be necessary outside of this context, and alternative demographic projections for populations outside of the United States. The methodology of Project 2, which used an exposure-response function sourced from a global meta-analysis, could also translate to any American city, or internationally provided compatible census and baseline mortality data. While Project 1's machine learning-based exposure-response function model was trained on specifically US Medicare claims data, it could readily be used to explore alternate climate conditions (e.g. the STORM team has released datasets modeling mid-21st century scenarios under a variety of climate models), or three other health outcomes (hospitalizations linked to respiratory illnesses, a chronic obstructive pulmonary disorder, or cardiovascular disease) within this context, and provided an equivalent source of daily health outcome data, the underlying exposure-response function approach could theoretically be repeated elsewhere. In addition to providing flexibility for future studies, this leverage of existing datasets provides a metaphorical "return on investment" from the initial studies and data collection efforts by providing frameworks to explore new public health-relevant questions. Overall, quantitative health impact assessments provide a valuable tool for researchers and practitioners in exploring current or potential environmental exposures in a public health context.

3. Theoretical Framework and Methodological Approach

While quantitative health impact assessments can take a variety of forms, in this implementation, we use several shared components for the risk assessment step (Figure 1), which focus on a single exposure in each case and can be described as projecting a future health outcome (e.g. mortality incidence rates in a population of interest)

as the effective "product" of current health outcomes (integrating observational or literature-based data on current incidence rates, as well as population counts and relevant demographics like age), the predicted change in exposure expected from the event/intervention, and the exposure-response function (exposure-response function, also called a dose-response function) (Balbus et al. 2016).

The exposure-response function, generally derived from associations between recorded exposure levels and health outcome data either in the research population and context or previous studies on compatible systems, describes the relationship between a change in exposure and the corresponding change in risk or incidence of the outcome—e.g. a hypothetical output might be a risk ratio of 1.1 for heat stroke in some population for every 1°C increase in the daily maximum temperature (binary exposure states could also be considered in some cases). We use a variety of specific exposure-response function models ranging from a straightforward linear regression in our project on urban tree canopy and mortality to a nonlinear (spline-based) distributed lag model to account for time since exposure in our project exploring mortality implications of proposed stratospheric aerosol injection.

By creating a consistent, quantifiable outcome estimate, quantitative health impact assessments enable the comparison of relative impacts for alternate proposals or scenarios—e.g., in our second project, the benefits of fully attaining a policy goal to increase tree cover, or of falling 50% short of this target. While we take a fairly consistent approach to the structure (if not individual models and components) of our quantitative health impact assessments, the method can take a variety of forms; for instance, where our projects each consider a single exposure of interest (tropical cyclone winds, presence of urban vegetation, and mean daily temperatures for the three projects), quantitative health impact assessments can accommodate multiple exposure types for a proposed intervention (e.g., considering rainfall as well as maximum sustained windspeed for projecting tropical cyclone health impacts). However, we do consider multiple health outcomes in Project 2. This is more a matter of project scope and intent than any intrinsic technical or methodological limitation.

HIAs can also take more qualitative forms, which frame impacts more narratively than quantitatively; these methodologies often integrate more smoothly with a participatory community input, although quantitative HIAs can also follow an initial or ongoing community engagement process. In our projects, we rely on generally publicly accessible datasets (with one exception for Project 3 in using the previously public NMMAPS dataset, and indirectly in Project 1 with using an exposure-response function derived from privileged United States Medicare claims, although these have been anonymized, and only derived data is available as part of the model), but did not engage in

a participatory approach, with two of our projects taking place on a national scale, and the third using current standing policy goals rather than seeking new potential measures.

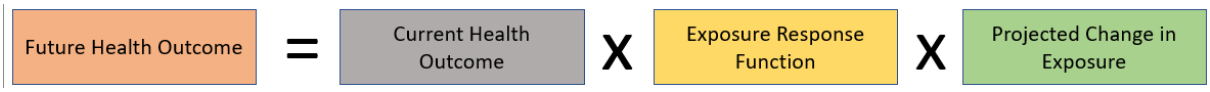

Figure 0.1: Overall conceptual diagram of the Quantitative Health Impact Assessment/Risk Assessment component as implemented in our projects, after Balbus et al 2016; future health outcomes are obtained by feeding data on current health outcomes (integrating population, demographic, and incidence data) and the projected change in exposure driven by an intervention or other event of interest into an exposure-response function, which models the observed association between exposure levels and outcomes in this or a comparable population and context.

Before delving into individual projects, it may be beneficial to discuss some overall strengths and limitations of the quantitative health impact assessment approach and commonalities across our projects. One overarching benefit of this body of work as *quantitative* health impact assessments is that the resulting impact projections are both "concrete" and transferrable between theoretical alternative interventions or scenarios, allowing us to calculate and articulate the relative degree of certainty for a given outcome. While more qualitative approaches have their ideal use cases and benefits, especially in the context of more community-led or participatory quantitative health impact assessments (Mindell et al. 2008), our approach in this case facilitates rapid and scalable assessments, and extensibility to other systems or outcomes.

In each of our projects, we employ exposure-response functions derived from observational studies, which means that while there are relevant known biological pathways, each analysis rests on an assumption of causality. While it is possible to adjust for known, quantifiable confounders when developing exposure-response functions, it remains likely that residual variance remains in each function. For instance, our second project uses an exposure-response function tying exposure to vegetative "greenness" (operationally determined with the remote sensing-based Normalized Difference Vegetation Index) to health outcomes like all-cause mortality. While multiple plausible and verified biological pathways exist(Rojas-Rueda et al. 2019), green spaces are also highly correlated with higher socioeconomic status through mechanisms like property value and historical discrimination (Nardone et al.), meaning that while our source study can control for a range of potential factors (e.g. income, age, sex, geographic areas), some dimensions of SES are unlikely to be fully captured. Similarly, our third project develops community-specific exposure-response functions linking temperatures to all-cause mortality (allowing for a lag in exposure time). At the same time, high and low extremes are known to increase the risk of death, among other adverse health

outcomes (Näyhä 2005), "off-target" dynamics like increased disease transmission from people staying indoors during cold weather may likewise introduce uncertainty or bias. While quantitative health impact assessments could theoretically use, for instance, experimentally-derived exposure-response functions from clinical trials, this approach rarely works with the kind of systematic, environmental exposure-based questions we explore in this body of work, and are especially well-served by quantitative health impact assessments (Briggs 2008).

In this work, we also use all-cause mortality (i.e., death from all causes in aggregate, as opposed to identified causes like cardiovascular disease, injuries, and so on) as a primary health outcome of interest, although two projects use additional outcomes. This metric has the advantage of presenting a rather final, concrete metric as far as health outcomes are concerned. Still, it should be acknowledged that mortality does not necessarily give a full picture of public health, and that other health outcomes like disease incidence and injuries play an important role in community wellbeing. However, especially in our approach of population-scale quantitative health impact assessments, mortality (especially all-cause) has the advantage of being reliably and uniformly counted across communities, and much less likely to be overlooked or misdiagnosed than e.g. chronic diseases. While as mentioned it does not represent a totality of public health (e.g. some diseases can exert an appreciable impact on quality of life without changing life expectancy), mortality also has a wide range of potential causes, allowing it to act as something of a proxy for a range of etiologies, a useful trait in the kind of systematic exposures we explore in these projects.

Given the modular nature of the quantitative health impact assessment framework, our projects all draw closely on previous literature and methodologies; Project 1 uses a simulated exposure dataset of the 10,000-year synthetic tropical cyclone seasons (Bloemendaal et al. 2023), which we filtering to storms in our area of interest—operationally, passing within 250km of the contiguous United States coastline. We sourced the exposure-response function directly from Nethery et al (2020). However, we used 2020 decennial census data in place of the older training data and replaced the historical tropical cyclone track inputs (converted into county-level exposures) with synthetic equivalents. In our project the novelty of the project lies in combining these elements, using an empirical exposure-response function based on historical tropical cyclone-mortality trends (with maximum sustained windspeed as the primary exposure of interest) to project all-cause mortality under alternative conditions (even, as in this case, what is effectively an "augmented" historical record) to provide a more comprehensive profile of tropical cyclone-associated risks in the region. As mentioned above, this framework could be readily expanded to

accommodate other climate scenarios, including projected mid-21st century conditions also modeled by Bloemendaal et al (2023), input from alternative models to approach more of an ensemble-style consensus, ideally smoothing out the influence of any given model's biases, or novel datasets trained using STORM or other algorithms.

However, while future scenarios would be mechanically compatible, the training data for the exposure-response function grounds it in a scenario of Medicare recipients in the present or near-present eastern United States, making it more feasible to model interventions or climate scenarios as "overnight" changes as in Project 2, rather than attempting to account for demographic shifts and related changes as in Project 3. While the STORM data, in this case trained using historical data, would be unable to capture novel events, the higher granularity afforded by the large, simulated dataset still enables us to provide a more detailed temporal and spatial risk estimates within an average modeled year, as well as identify potentially overlooked at-risk regions which could see tropical cyclone exposures under current conditions. While "behind the scenes" and potentially more subjective in value, another component of the project was reworking the code for the methodology laid out in Vicedo-Cabrera et al (2019) to follow a more modular format and move from a robust but at times unintuitive "base R" approach (often combining analytical and modeling steps in the same processes) to one using the *tidyverse* and *ggplot2* frameworks commonly used by at least our lab.

Synthetic tropical cyclone datasets, as a whole, have already been used for other forms of impact assessments for some time, with this being one of the driving motivations for the creation of this method. To use a recent and emblematic example, Meiler et al. (2022) performed a metanalysis comparing STORM alongside three other synthetic tropical cyclone algorithms (these relying on coupled statistical-dynamical models which incorporate physical simulations of climate systems in addition to deducing "rules" of tropical cyclone behavior from the resampled datasets) and a simpler perturbation-based method (adding noise to historical data), tracking both the physical characteristics of the simulated events and corresponding economic impact assessments from the CLIMADA (CLIMate ADAptation) suite of tools, an open source framework providing economic damage assessments for a range of climate-associated exposures natural disasters, which has been used both by researchers and practitioners in recent years (Bresch and Aznar-Siguan 2021). In this sense, beyond the immediate value of this or similar quantitative health impact assessments, researchers could integrate a desirable end state into a similar suite of tools, where it could support impact assessment efforts and integrate with tools developed by experts in other

fields. This kind of integration could help better capture the multifaceted nature of climate change, which is difficult to address in any discipline.

Although Project 2 fairly directly implements an approach from an existing paper (Kondo et al. 2020)—we made several notable methodological changes to better meet our study population (e.g. Denver and Phoenix have a wider range of block group sizes and greater inter-city vegetation variance, so we changed the exposure metrics and modeling accordingly) reflect our goal of developing a more robust and extensible framework, as well as unique features of our intended study systems. We based the initial framework closely on Kondo et al (2020), which also used the exposure-response function developed in Rojas-Rueda et al (2019) in the context of a census-based HIA based on urban tree canopy goals, featured a UTC-to-NDVI conversion model, used Landsat8 imagery for NDVI data, but modified other aspects of the approach to reflect this project's specific setting and goals. From the outside, this was intended to be a "pilot" for an extensible framework as much as a specific investigation into these cities, so we used the United States Forest Service tree canopy dataset (United States Forest Service) and National Land Cover Database (Jin et al. 2019) to provide spatially and methodologically consistent data across the United States, if likely at the cost of accuracy and specificity in more granular regional estimates like the one used in Kondo et al (2020). For instance, early in the study we considered using municipal zoning data to map population concentrations more precisely but found that the formats and structures didn't match between our two test cities, an especially relevant consideration given our intent to develop an expandable, generalizable pipeline. Another choice on this front was to replace the linear regression model for the UTC-NDVI association with a more flexible model capable of reflecting nonlinearities and variation within each city; we found that while a model of similar complexity could capture Denver's distribution fairly well, the same was not true for Phoenix, and in any case, a single model did not appear to capture both cities well, meaning that hypothetical future users would need to fit models for additional cities manually. Phoenix especially exhibited a much wider range of local landscape types than Philadelphia, covering everything from well-watered golf courses and parks to riparian woodlands to arid mountain preserves, making local variation even within a city desirable.

In this respect, our UTC-to-NDVI predictive modeling solution should balance generalizability across cities and specificity to any single city. Another important consideration was that while Philadelphia's census tracts had a mean area of 81 m^2, which the authors considered sufficiently comparable to the circular 500 m^2 buffers used for the exposure metric in Rojas-Rueda et al., Denver and especially Phoenix were more unevenly distributed especially at

the census tract level, featuring a wide range in sizes. We intended to use both the 500 m all-cause mortality buffers, and the 250 m definition used for stroke and dementia outcomes in Paul et al (2020), meaning that the same approximation would be untenable. To this end, we used the more granular census blocks as a starting point, further focusing on NLCD-assessed "developed" (operationally, effectively interchangeable with >20% impervious surfaces) land as a proxy for population centers within block groups, a distinction mostly relevant in large examples on the edges of cities, and drew buffers of the appropriate radius around these "population-weighted" regions to calculate mean NDVIs and UTC values. Beyond better reflecting the spatial variability in block groups, this approach also carried the advantage of allowing the landscape surrounding block groups to be accounted for—e.g. in the Kondo et al. (2020) approach, the exposure profile of a long, thin block group adjacent to a park would only reflect the interior NDVI, while our implementation would reflect both this and the nearby vegetation.

Naturally, administrative boundaries like block groups or census tracts are not the only option for approximating spatial exposures; Iungman et al. (2023) used the same exposure-response function, but based their estimate on a gridded 2015 population estimate dataset with a 250 m resolution, which would have the advantage of preserving more granularity in environmental exposures (especially in larger block groups like those seen outside of Phoenix), and potentially capturing more detail in how populations are distributed (e.g. our model would treat high-density housing blocks as equivalent to single-family housing). Near the end of the process of developing project 2, we identified one potential candidate for a similar approach in a Meta (parent company of Facebook) dataset using Sentinel satellite imagery to estimate 2020 population density at a 30 m resolution (Fibæk et al. 2022); an informal comparison revealed a fairly good agreement with census tract and block group estimates (with some appreciable outliers especially in lower-density areas), and an advantage in being able to differentiate residential from non-residential developments generally, and reflecting high-density dwellings like apartments or dorm buildings.

We used a somewhat simplified, "optimistic" operational representation of UTC policy scenarios, assuming block groups would individually (and with priority for inhabited areas) reach at least the cities' standing UTC goals, and not accounting for logistical limitations like water or financial resources, barriers posed by impervious surfaces, or specific tree species (in fact, we implicitly assume a largely non-native complement in both cases by extrapolating from the current canopy). As an example of how these considerations could be addressed more explicitly, a recent paper from the Rojas-Rueda lab (Garber et al. 2023) includes spatially explicit scenarios for planting native species to a certain level in e.g. riparian corridors, parking lots, or stormwater drainage systems in

Denver, with the scenarios designed with input from local stakeholders and planners. This approach presents a more detailed and context-appropriate representation of policy scenarios. In contrast, our approach makes some concessions in the name of generalizability between cities (with the predicted outcomes, while well-founded, arguably acting more as general indicators).

One challenge, and unavoidable decision in this project is that there is, as of yet, no standard temporal metric for health-relevant NDVI exposure; e.g. Kondo et al (2020) opted to use a single day, selected *a priori* based on clear skies and being in a known relative "green" period of the year, while others might use single- or multiyear averages (either across the year or across fixed reference months, as in Iungman et al. (2023). To some extent this variety is defensible in that especially high NDVIs tend to be fairly closely correlated at least between adjacent years (Helbich 2019), meaning that even a single representative timepoint could be understood as a sufficient indicator for general greenness. Our approach of using a "greenest month", rather than a fixed date, was intended to reflect the fact that different vegetation trends exist between cities—e.g. Phoenix's NDVI levels in particular are driven by complex interactions between native vegetation, agriculture, and urban greenery, resulting in a winter or early spring peak (Buyantuyev and Wu 2012). Still, in retrospect a simple annual mean may capture a similar level of specificity (if potentially 'diluting' the effect with low-greenery seasons) and be more within the range of existing literature, because it does not seem like an unmistakable advantage over existing metrics. The effective "exposure time" of NDVI in a public health context has also been relatively little-studied, which is understandable in that barring a study design explicitly focusing on a highly mobile population with multiple locations, or e.g. environmental changes like the aftermath of a wildfire, it would be difficult to disentangle the effects associated with NDVI over a period of interest from the likely correlated NDVI exposures in previous timeframes.

Along similar lines, using fixed addresses is something of a concession to considerations like participant privacy, data collection logistics, and scalability, in that environmental conditions around ones' home may not necessarily reflect total cumulative exposure (e.g. in some professions, or cases like a student with extracurricular activities, a majority of waking hours might be spent outside the home). Remote sensing data, while providing a methodologically uniform an objective (as opposed to e.g. subjective metrics, which may nonetheless be useful in contexts like more qualitative research or satisfaction with public parks—e.g. Sefcik et al. 2019), does come with the caveat of abstracting legal or logistical barriers to use—for instance, living near a private golf course or corporate campus may have different health implications than a physically identical park, as might living across a busy road

from the same park; this known uncertainty has led to some calls for the development of a more "function-oriented" metric for green space, which could incorporate factors like crowding and accessibility. However, this would (at present) necessarily entail a more detailed data collection process (Ekkel and de Vries 2017).

These and similar limitations point to a perhaps fundamental consideration for environmental quantitative health impact assessments, which is that while the best available data for a given question (e.g. between collection logistics and privacy, daily-resolution mortality data as used in Project 3 or the exposure-response function for Project 1 is available only at the county level and with special access in a United States context) may require concessions concerning precision or accuracy (ideally taking the form of Berkson Error, where the bias is non-directional), "over-fitting" the operational precision for a given exposure-response function could arguably become counter-productive concerning modeling real-world health impacts to in the sense that necessary or useful abstractions and indicators in observational studies (e.g. using mean NDVI around a participant's home address as a proxy for cumulative lifetime exposure) do not necessarily conform to specific mechanisms for health outcomes. The quantitative health impact assessment approach in Project 3 is based on three major "pillars" from a data and methodology perspective. In this context, we wanted to model mid-21st century scenarios explicitly given the projected increase in older (65+) adults and their greater sensitivity to adverse temperature-associated health outcomes (Greaney et al. 2016), so it was important to model this population of interest explicitly, rather than assuming an "instantaneous" change in exposure as in Project 2, or the near-past conditions of Project 1. To this end, we used projected demographic data from NASA's Socioeconomic Data and Applications Center (SEDAC) (Hauer and Center For International Earth Science Information Network 2020)

These considerations are illustrative of tradeoffs inherent in using preexisting public data, with some benefits including greater cost and time effectiveness over collecting new data, better replicability, and transferability to new datasets, and (in an abstract sense) obtaining more value from the resources originally invested in generating the datasets. For instance, the historical dataset used in Project 3, which had its public accessibility discontinued in 2012, had been used in the development of 67 scientific papers published by that the time (Barnett et al. 2012), and continues to be a valuable resource for testing temperature-mortality associations (Bobb et al. 2014).

4. Limitations

It is important to acknowledge some limitations intrinsic to the quantitative health impact assessment framework overall, as well as specific challenges and limitations of our approach. Our overall implementation's basis in preexisting datasets posed several universal limitations. At a 'strategic' level, our study lacks the kind of stakeholder engagement that would be ideal in especially more comprehensive analyses (covering e.g. multiple exposures and outcomes), with this level of specificity being something of a tradeoff for the scalability and generalizability we pursued in design choices like readily available remote sensing and census data. The gap is perhaps most notable in Project 2, where more detailed 'ground-truthing', at least in a single city, could shed light on factors like scenario practicality in specific areas given e.g. land use codes, or inform assumptions about uniform benefits from NDVI exposure across the cities (e.g. in practice, a public park could conceivably be more beneficial than physically equivalent private land due to physical accessibility supporting mechanisms like psychological benefits or exercise). These are more abstract in the larger-scale Projects 1 and 3, but, for instance, consulting local experts may point to e.g. elements of infrastructure and other community resilience measures that may influence the relative impact of tropical cyclones beyond the census-based parameters (median income, fraction of home ownership, population density, etc.) used in the exposure-response function. In practice, such datasets may be harder to obtain in a systematic fashion (making this level of detail more practical for geographically smaller studies).

More granularly, these limitations tended to entail a level of abstraction for potentially health-relevant considerations (e.g. using point temperatures to stand in for larger areas in Project 3, or aggregated census data throughout). Aggregated data does not necessarily introduce bias (e.g. Berkson Error) but can influence interpretation. For instance, while the Couple Earth Systems Modeling Project 2 (CESM2) (Danabasoglu et al. 2020) datasets are widely used and founded on carefully-researched assumptions, the spatial resolution of a roughly 1-degree grid means that even given the best available data, the nominal temperatures in Project 3 reflect averages over a larger, potentially heterogenous region. In contrast, in the historical NMMAPS data, the temperature estimates represent readings from specific onsite monitors within the county (Health Effects Institute 2000), and are treated as representative of the wider county-level data used for population and mortality statistics. For example in Project 1, while estimating county-level impacts of tropical cyclones provides a much better spatial resolution than grid-style climate models (Bloemendaal et al. 2020a). Our exposure-response function choice allows us to account for relevant features of specific counties, we would still be unable to make judgments about, for instance, specific areas or populations within the counties that would see higher risks. In the case of Project 1, where we hewed

closely to a model developed for our population of interest, much of this detail is arguably implicit in the training data. Still, this limitation becomes more apparent when adapting models to new contexts, as in Project 2.

While the all-cause mortality exposure-response function used a large, international composite cohort as training data, one could expect that specific NDVI-health associations could vary by population, necessitating further abstraction. In this case, we assumed county- or even state-level (in the case of neurological health outcomes) mortality rates applied uniformly. Still, in practice one could expect each city's specific socioeconomic and demographic composition and spatial distribution to influence this trend, in addition to potentially other environmental exposures, would modify this trend. The model also assumes an essentially linear relationship between NDVI and health, where some alternative approaches have predicted that at least tree-based cooling follows more of a nonlinear trend with diminishing returns over a certain level (Zhou et al. 2021). In general, we follow an underlying assumption that the selected exposure-response function adequately represents the population and range of exposures in question. In this sense, our predicted averted mortality or neurological health condition data may be more of an indicator unit than an exact mechanistic prediction (especially given that our methodology is essentially "atemporal" in abstracting the necessity for newly added trees to grow and survive over potentially decades.

In another sense of abstraction, our projects also demonstrated the need, or at least utility, of abstracting potentially relevant features—for instance, in Project 2, we did not model important considerations like potential economic and water resource limitations to planting trees, and in Project 3 elected to treat population data (both historical and projected) as fixed values in Project 3 because no uncertainty intervals were provided, and without "in-house" demographics expertise, attempting to deduce reasonable uncertainty ranges seemed like an overstep. To some extent any model requires this kind of simplification by available data (as noted above) if nothing else, but this does highlight the valuable role that interdisciplinary collaborations can play in designing quantitative health impact assessments.

More fundamentally, comprehensively modeling even a relatively near-term (e.g. 2030s) future scenario along relevant parameters would be ambitious, given the many potentially interacting factors like population growth, local and international politics (e.g. zoning policies, internal migration) or conflict, short-term climate and weather, technological changes, economics, public health outside the sphere of the quantitative health impact assessment, and so on. To use a dramatic recent example, even a well-founded short-term model made in the years before 2020 would unlikely have anticipated the scope of the COVID-19 epidemic and its wide-ranging implications e.g.,

mortality rates and economic trends. While we focused on a mid-2050s timeline for Project 3, this was already apparent in the earlier section of the SEDAC projection, developed in 2015, and the SSP2-4.5 scenario on which they were based, where the now- "historical" early 2020s naturally lack this disruption. In this respect, it may be most prudent to use fairly simple projected with well-defined assumptions, trying to include only relevant and fairly well-grounded features of a future scenario is integral to the analysis like the larger fraction of older Americans in Project 3, given the different influences of temperature on mortality by age.

There are also limitations intrinsic to the data and specific methods selected for these projects. For example, in Project 1, the STORM algorithm carries known caveats, including an overestimation of strong hurricanes (Category 3 or higher), which could increase the estimated impact of these events, although they are fairly rare in the dataset; a previous exploration also revealed that tropical cyclone genesis rates along the equator differed appreciably from historical estimates. In Project 2, we chose to use United States Forest Service tree cover estimates, rather than city-level estimates, because of the spatial and methodological uniformity across the United States, which would facilitate better scalability to new contexts. However, we selected this dataset ofr its ability to give a reasonable approximation of canopy cover across the entire United States, so gaps like the appreciable mismatch between the reported and predicted mean UTC for Phoenix could represent a failure of the model for an outlier. While our relative change-based approach could help mitigate this to some extent, it remains an underlying limitation of the data, and an area that could benefit from more detailed datasets if methodological adaptability across cities was not a priority. Along similar lines, in Project 3, some of the simulated temperature data in the Community Earth Systems Model 2 (CESM2) used for 2050-2060 predictions differed appreciably from the temperatures in our historical temperature and mortality dataset for 1987-2005 (if to a smaller degree than with the UTC data) even for now "historical" modeled data (2015 in the 2015-2100 model), which could lead to effective exposure mischaracterization in the context of CESM2-based predictions. Specifically, the historical data is aggregated from specific on-site sensors, while the CESM2 data uses 1° (roughly equivalent to 111 km) gridded cells, leading to larger spatial aggregation, and less influence from micro-climates around sensor sites (especially relevant in coastal climates, or e.g. areas with dramatic elevation gradients; in our case, Miami, FL, and Los Angeles, CA, differed appreciably between the versions). While the historical dataset uses the sensor-based temperatures, there is no "mechanical" linkage between these and the corresponding daily county-level mortality counts, so as a future step in the overall analysis, we intend to explore substituting historical reanalysis data more

methodologically cohesive with CESM2 to see if we can obtain correspondingly better agreement in the exposure-response functions.

The process of performing multiple quantitative health impact assessments at vastly different spatial scales also highlighted the inevitability of these kinds of abstractions; for instance, in Project 2, we made decisions to attempt to preserve 30-meter nuances of population and environmental patterns at the block group scale in Phoenix, while in Project 3, our projected temperature data and the need for a daily resolution of mortality dictated that the same city was (nominally) represented by Maricopa county as a whole in the historical NMMAPS data, and physically by a roughly 1-degree square (more than 100 km on a side). In this sense, it is important to understand that the best available data for a given purpose may come with considerable caveats on a given front, and to communicate this to readers, which we have attempted to do in this body of work.

5. Outline of Dissertation

The three central chapters of the dissertation each contain a self-contained, paper-style description of each project as described above. The first and third are drafts in preparation (the first is nearly ready for feedback from coauthors, while we antitipate more extenisve improvements to the third before seeking publication considering some of the findings under the methodolgy as designed), while the second has been published online in *Environmental Research* (Health implications of urban tree canopy policy scenarios in Denver and Phoenix: A quantitative health impact assessment - ScienceDirect) at time of writing. These project-based chapters are followed by a general conclusion, and finally supplemental sections for each of the proejcts in sequence providing additional plots, tables, and in some cases methodological detail.

References

Balbus J, Crimmins A, Gamble JL, Easterling DR, Kunkel KE, Saha S, et al. 2016. Ch. 1: Introduction: Climate Change and Human Health. The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment 25–42.

Barnett AG, Huang C, Turner L. 2012. Benefits of Publicly Available Data. Epidemiology 23:500; doi:10.1097/EDE.0b013e31824d9ef7.

Bloemendaal N, Haigh ID, De Moel H, Muis S, Haarsma RJ, Aerts JCJH. 2020. Generation of a global synthetic tropical cyclone hazard dataset using STORM. Sci Data 7:40; doi:10.1038/s41597-020-0381-2.

Bobb JF, Peng RD, Bell ML, Dominici F. 2014. Heat-Related Mortality and Adaptation to Heat in the United States. Environmental Health Perspectives 122:811–816; doi:10.1289/ehp.1307392.

Bresch DN, Aznar-Siguan G. 2021. CLIMADA v1.4.1: towards a globally consistent adaptation options appraisal tool. Geosci Model Dev 14:351–363; doi:10.5194/gmd-14-351-2021.

Briggs DJ. 2008. A framework for integrated environmental health impact assessment of systemic risks. Environmental Health 7:61; doi:10.1186/1476-069X-7-61.

Buyantuyev A, Wu J. 2012. Urbanization diversifies land surface phenology in arid environments: Interactions among vegetation, climatic variation, and land use pattern in the Phoenix metropolitan region, USA. Landscape and Urban Planning 105:149–159; doi:10.1016/j.landurbplan.2011.12.013.

Danabasoglu G, Lamarque J -F., Bacmeister J, Bailey DA, DuVivier AK, Edwards J, et al. 2020. The Community Earth System Model Version 2 (CESM2). J Adv Model Earth Syst 12:e2019MS001916; doi:10.1029/2019MS001916.

Ekkel ED, de Vries S. 2017. Nearby green space and human health: Evaluating accessibility metrics. Landscape and Urban Planning 157:214–220; doi:10.1016/j.landurbplan.2016.06.008.

Fibæk CS, Keßler C, Arsanjani JJ, Trillo ML. 2022. A deep learning method for creating globally applicable population estimates from sentinel data. Transactions in GIS 26:3147–3175; doi:10.1111/tgis.12971.

Garber MD, Guidi M, Bousselot J, Benmarhnia T, Dean D, Rojas-Rueda D. 2023. Impact of native-plants policy scenarios on premature mortality in Denver: A quantitative health impact assessment. Environment International 178:108050; doi:10.1016/j.envint.2023.108050.

Greaney JL, Kenney WL, Alexander LM. 2016. Sympathetic regulation during thermal stress in human aging and disease. Autonomic Neuroscience 196:81–90; doi:10.1016/j.autneu.2015.11.002.

Hauer M, Center For International Earth Science Information Network. 2020. Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100.; doi:10.7927/DV72-S254.

Health Effects Institute. 2000. National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and Mortality from Air Pollution in the United States. Health Effects Institute. Available: https://www.healtheffects.org/publication/national-morbidity-mortality-and-air-pollution-study-part-ii-morbidity-and-mortality-air [accessed 22 February 2023].

Helbich M. 2019. Spatiotemporal Contextual Uncertainties in Green Space Exposure Measures: Exploring a Time Series of the Normalized Difference Vegetation Indices. IJERPH 16:852; doi:10.3390/ijerph16050852.

Iungman T, Cirach M, Marando F, Pereira Barboza E, Khomenko S, Masselot P, et al. 2023. Cooling cities through urban green infrastructure: a health impact assessment of European cities. The Lancet 401:577–589; doi:10.1016/S0140-6736(22)02585-5.

Jin S, Homer C, Yang L, Danielson P, Dewitz J, Li C, et al. 2019. Overall Methodology Design for the United States National Land Cover Database 2016 Products. Remote Sensing 11:2971; doi:10.3390/rs11242971.

Kondo MC, Mueller N, Locke DH, Roman LA, Rojas-Rueda D, Schinasi LH, et al. 2020. Health impact assessment of Philadelphia's 2025 tree canopy cover goals. The Lancet Planetary Health 4:e149–e157; doi:10.1016/S2542-5196(20)30058-9.

Meiler S, Vogt T, Bloemendaal N, Ciullo A, Lee C-Y, Camargo SJ, et al. 2022. Intercomparison of regional loss estimates from global synthetic tropical cyclone models. Nat Commun 13:6156; doi:10.1038/s41467-022-33918-1.

Mindell JS, Boltong A, Forde I. 2008. A review of health impact assessment frameworks. Public Health 122:1177–1187; doi:10.1016/j.puhe.2008.03.014.

Nardone A, Rudolph KE, Morello -Frosch Rachel, Casey JA. Redlines and Greenspace: The Relationship between Historical Redlining and 2010 Greenspace across the United States. Environmental Health Perspectives 129:017006; doi:10.1289/EHP7495.

Näyhä S. 2005. Environmental temperature and mortality. International Journal of Circumpolar Health 64:451–458; doi:10.3402/ijch.v64i5.18026.

Rojas-Rueda D, Nieuwenhuijsen MJ, Gascon M, Perez-Leon D, Mudu P. 2019. Green spaces and mortality: a systematic review and meta-analysis of cohort studies. The Lancet Planetary Health 3:e469–e477; doi:10.1016/S2542-5196(19)30215-3.

Sefcik JS, Kondo MC, Klusaritz H, Sarantschin E, Solomon S, Roepke A, et al. 2019. Perceptions of Nature and Access to Green Space in Four Urban Neighborhoods. IJERPH 16:2313; doi:10.3390/ijerph16132313.

United States Forest Service. USFS Analytical 2016 Tree Canopy Cover CONUS (Image Service). Available: https://data-usfs.hub.arcgis.com/datasets/75ae6ca6cb954e2bb15dac042ecc05d2 [accessed 30 September 2023].

Zhou W, Huang G, Pickett STA, Wang J, Cadenasso ML, McPhearson T, et al. 2021. Urban tree canopy has greater cooling effects in socially vulnerable communities in the US. One Earth 4:1764–1775; doi:10.1016/j.oneear.2021.11.010.

CHAPTER 2: PROJECT 1: HEALTH IMPACT ASSESSMENT OF TROPICAL CYCLONE-RELATED EXCESS MORTALITY IN EASTERN UNITED STATES USING STORM FOR SYNTHETIC STORM TRACKS

1. Introduction

Tropical cyclones, or low-pressure storm systems forming over tropical waters, represent a major public health hazard, especially as they reach hurricane strength. Between 1980 and 2009, tropical cyclones globally caused a reported 412,644 deaths, 290,654 other injuries, and 466.1 million people otherwise impacted (Doocy et al. 2013). Tropical cyclone health impacts are often delayed—e.g., a survey-based reassessment of mortality from Hurricane Maria in Puerto Rico estimated 4,645 (95% confidence interval 793-8,498) excess deaths to the initial official count of 61 (Kishore et al. 2018). Roughly one in three of these deaths was attributed to delayed or disrupted healthcare, compared to just one in ten directly to the hurricane itself. Tropical cyclones can introduce hazards like floodwater-borne pathogens and toxins (Erickson et al. 2019), displacement-associated increased infectious disease exposures, and interrupted hygiene (Shukla et al. 2018). Tropical cyclones are often associated with power outages, which can disrupt services like air conditioning or medical equipment, and raise the risk of compounding hazards with heat waves or similar exposures (Feng et al. 2022). Even gasoline and diesel generators carry risks like carbon monoxide poisoning from improper indoor usage (Stevens and Ashley 2022) or, when used at a community scale, potentially hazardous levels of carbon dioxide and sulfur dioxide exposure (Subramanian et al. 2018). The storms are especially dangerous for vulnerable groups like older adults due to factors like mobility challenges, susceptibility to physiological stress, and reliance on potentially-disrupted systems like regular medical care and indoor climate control (Parks et al. 2021). Given these and related complexities, the National Academies of Science, Engineering, and Medicine have called for uniform, scalable analytic approaches to evaluating population-level health impacts in the aftermath of natural disasters as a key component in American disaster preparedness and resilience (National Academies of Sciences et al. 2020)

Considering the often delayed and indirect nature of tropical cyclone-associated mortality, spatially and temporally large-scale studies of population-level mortality trends, paired with concurrent tropical cyclone records, are useful for estimating the cumulative impact of tropical storms. Health impact assessments are a useful tool in this respect. They estimate health implications of exposures like tropical cyclones by combining demographic and health outcome data with functions from epidemiologic studies to link exposures with a health outcome "response". In this

investigation, we focus on "excess" mortality, or deaths exceeding a counterfactual exposure where the exposure event had not occurred, an intuitive but statistically complex proposition (Harris-Roxas et al. 2012b). The mortality implication of any single tropical cyclone can be quite difficult to assess, but powerful population health-level analytical models integrating epidemiological and atmospheric science are emerging to fill this gap. For this investigation, we've selected an exposure-response function using Bayesian modularization to integrate two complementary models. The first is a causal inference model using 1980-2015 tropical cyclone records, and all contemporaneous Medicare recipients in the Eastern United States, to model sustained windspeed-mortality associations over a two-week window around exposure events. The second component is a predictive module which updates estimates from the first using relevant county-level and storm-specific characteristics to yield more precise estimates (Nethery et al. 2020). This is one in a small but growing body of comparable models, including a gridded global excess mortality model for the 2 weeks after tropical cyclone exposure (Huang et al. 2023) and a United States-specific model tracking longer-term county-level trends for a variety of health outcomes among all age groups at a monthly resolution (Parks et al. 2022).

However, a challenge for health impact assessments is that tropical cyclones, as an exposure class, are fairly rare and unevenly distributed, making it difficult to apply traditional probability-based estimates for long-term risk. With an estimated global total of 37 annual events between 1980 and 2009 (Doocy et al. 2013), tropical cyclones are much rarer, and hence harder to model, exposures than, for instance, air temperature or quality. While e.g. mid-19th century observations of tropical cyclones in the Indian Ocean or parts of the North Atlantic, and paleoclimatology can provide much longer-term estimates for some regions, effective health impacts assessments require the kind of standardized methodology and spatial continuity of modern datasets like the International Best Track Archive for Climate Stewardship (IBTRACS), which is available from 1980 onward (Knapp et al. 2010). This relatively sparse historical data raises the possibility of single high-impact events, or the absence of events, skewing traditional probability-based risk estimates, potentially leading to over- or underestimates of regional risk profiles relative to hypothetical "true" long-term trends.

At the same time, tropical cyclones are not random events, but shaped by consistent physical laws. These constraints facilitate simulated extensions of historical data, alleviating the bottlenecks associated with the scarcity of high-quality historical data, at least to the extent that the models are able to accurately model the exposures (Meiler et al. 2022). There are a variety of simulation approaches, including models incorporating relevant physical climate

dynamics, fully statistical resampling of existing storm records, and combinations of both types (Meiler et al. 2022). Fully statistical simulations infer plausible tropical cyclone "rules" from historical data using machine learning methods like Markov Chains, and require fewer computational resources than models incorporating physical simulations (Bloemendaal et al. 2020a). While the possibility of producing thousands, or tens of thousands, of years of data for sampling, it is important to understand any systematic biases present in simulations before accepting results at face value.

In this investigation, we combine a leading tropical cyclone resampling algorithm with an empirical exposure-response function for the association between tropical cyclone exposure and all-cause mortality, to generate county-level predictions of tropical cyclone-attributable mortality among American adults 65 and older. Our synthetic tropical cyclone dataset is 10,000 pre-generated seasons from the Synthetic Tropical Cyclone geneRation Model (STORM), a leading open-source, fully statistical tropical cyclone resampling algorithm method (Bloemendaal et al. 2020a). Specifically, we use a pre-generated dataset of 10,000 synthetic seasons generated by resampling historical (1980-2018) historical storms (Bloemendaal et al. 2022a). Our exposure-response function model's exposure metric is estimated county-level maximum sustained windspeed; this was obtained from a wind field generated from the central windspeeds provided in the historical data (Nethery et al. 2020). We also compare STORM estimates against historical (1980-2022) data to assess the extent of agreement, setting general context for trustworthiness.

2. Methods

2.1: Overview of Analysis

To briefly describe the structure of our analysis, we aimed to estimate hurricane-attributable mortality counts (and derived metrics like rates per 100,000 residents) among older Americans under synthetic North Atlantic Basin tropical cyclones exposures representative of late 20th to early 21st century historical data from the International Best Tracks Archive for Climate Stewardship (IBTrACS), considered a "gold standard" for methodologically consistent tropical cyclone data. In order to obtain this, we needed population data (including relevant age and socioeconomic demographics), predicted county-level wind exposures experienced as a result of the cyclones, and a model to link these two elements in a health impact assessment context. For the first, we used 2020 census data. The next component required a two-step process: first, filtering simulated STORM tropical cyclone tracks to our geographic area of interest, and second using a model to convert the central windspeeds and coordinates of this data into corresponding wind

fields describing the exposures that would be experienced by county-level populations due to the simulated storms, a more health-relevant metric. Finally, we combined these elements using an existing exposure-response model based on historical storm and mortality that uses the wind exposure intensity (maximum sustained windspeed) and population characteristics to predict the excess tropical cyclone-attributable mortality for each storm event. From there, we were able to explore spatial and temporal patterns in tropical cyclone-related mortality risks. With this being an experimental and somewhat transformational application of the STORM algorithm, we also took steps to compare the predicts against historical data for our region of interest and consider potential limitations and caveats in this application.

2.2: Study System and Population Data

Our study area (Figure 1) was the continental Eastern United States—operationally 2,209 counties across 25 states exposed to Tropical Cyclones over the course of the simulation, and a population of older adults (65 years and older). We obtained county-level demographics data from the 5-year 2020 American Community Survey United States Census, accessed via the *tidycensus* R package (US Census Bureau 2022; Walker and Herman 2022).

To create aggregate regional-level estimates of tropical cyclone impacts, we used multistate regions as defined by the National Centers for Environmental Information (NCEI), which subdivide the United States into nine regions sharing similar climates for weather forecasting and related applications (Karl and Koss 1984). Our study area coincided with four of these regions (Figure 1): the Northeast (Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont), the Southeast (Alabama, Florida, Georgia, North Carolina, South Carolina, and Virginia), the South (Arkansas, Kansas, Louisiana, Mississippi, Oklahoma, and Texas), and a portion of the landlocked Ohio Valley region (Indiana, Kentucky, Ohio, Tennessee, and West Virginia) (Karl and Koss 1984).

2.3 Simulated Tropical Cyclone Tracks

To obtain health-relevant tropical cyclone-associated exposures, we extrapolated from 10,000 seasons of STORM algorithm-generated tracks to generate corresponding county-level sustained maximum windspeeds. In doing so, we explore the viability of using synthetic tropical cyclone seasons to augment the rare, high-quality historical data. As we will discuss, we took steps to assess the relative agreement between STORM and historical data along a

number of relevant metrics for this use case. This process began with the synthetic STORM hurricane seasons (June–November) resampling algorithm (trained on historical records from 1980-2018) made available by the team who created the algorithm, which is an updated version of the output used to validate the STORM model (Bloemendaal et al. 2022b).

In a recent comparative study, STORM's basin-wide accuracy (relative to historical benchmarks globally and in the North Atlantic basin specifically) was comparable to three analogous statistical-dynamical synthetic tropical cyclone models on metrics including tropical cyclone probability density, frequency and intensity, and return periods (time between storms of a certain intensity), with some caveats including overrepresenting the frequency of Category 3 and higher hurricanes relative to historical data (Meiler et al. 2022). STORM's authors also found that at the basin level, it was able to approximate historical (1980-2018) tropical cyclone season traits such as number of storms per year and average maximum wind speeds at the basin level, if with some noted discrepancies including a lower-than-historical rate of landfall events (Bloemendaal et al 2019). While this analysis focused on the entire North Atlantic basin, for our application we geographically subset the storms to an area relevant the eastern United States, potentially changing some of the relationships.

Each synthetic hurricane season in the STORM-generated dataset is rendered at a 0.25° × 0.25° (approximately 27.75 km) resolution and consists of individual storm tracks and a range of modeled relevant parameters at 3-hour time steps across the storm's lifetime. Of the simulated tropical cyclone traits, we used latitude, longitude, maximum sustained windspeed per timepoint, identifying variables for month of genesis, year (years are self-contained, with e.g. no modeled continuity from a nominal Year 1 to Year 2), and unique storm identities. For our analysis of STORM data, we focused on tropical cyclones generated in the North Atlantic Basin and passing within 250 km of the continental United States' coastline. Operationally, this included all tropical cyclone track timesteps between the first entry into, and last departure from, this area of interest.

2.4 Windfield Modeling and Exposure Classification

For each synthetic storm, we modeled a corresponding wind field throughout the storm's history. This wind field provides estimates of the peak winds that a storm incurred at population-weighted county centers throughout our study area over the storm's recorded lifetime (with STORM terminating tracks once they transition to e.g. an extra- or subtropical cyclone or weaken to a disturbance). To accomplish this, we interpolated the 3-hour wind speed and

coordinate timesteps into 15-minute intervals, and then modeled the wind field at each time point using the Willoughby wind field model, with both steps applied through the open-source R package *stormwindmodel* (Anderson et al. 2021; Willoughby et al. 2006). This provided estimates for our main exposure of interest: maximum surface-level (10 m) sustained windspeeds throughout the storm (Figure S1.1) at predetermined population-weighted county centers (Figure S1.2).

While tropical cyclone tracks themselves can be categorized using the Saffir-Simpson wind scale, which is based on central windspeeds, this derived metric and shift in perspective to county-level windspeeds is better captured by the Beaufort wind scale, which is designed to reflect conditions experienced by a stationary observer, rather than describing central storm track characteristics (World Meteorological Organization (WMO) 1970). For clarity, we will refer to exposures as "hurricane-force" if they fall in this highest category of the Beaufort scale, use "gale-force" to describe exposures are lower than hurricane-force but above 17.5 m/s (34 knots).

For this, we generated of county-level exposure events, or cyclone days—one day with tropical cyclone–associated maximum sustained wind speeds of ≥17.49 m/s (equivalent to gale-force winds using the Beaufort scale) (Parks et al. 2022). Given the focus on maximum sustained windspeed per storm, there was a maximum of one cyclone-day per county-storm combination (i.e. a storm that stayed over a county for multiple days would only be assigned one exposure event). Because the Nethery et al (2020) exposure-response function model uses a minimum exposure threshold of 17.4 m/s, we excluded county-level exposures below this limit for modeling purposes; this threshold also closely coincides with the 18 m/s minimum threshold used in the STORM training data (Bloemendaal et al. 2020b), although in practice weaker winds—both central and county-level—were present in the dataset. We are also interested in the interaction between physical tropical cyclone exposures and populations, so calculated a derived "cyclone person-days" metric by taking the product of exposed county populations and cyclone events over the simulated period.

2.5 Comparison of Synthetic and Historical Hurricanes

An in-depth assessment of STORM's agreement with historical tropical cyclone activity was outside the scope of our analysis, with diagnostic assessments against historical data in the initial methods, and a subsequent inter-model comparison deeming it to have accuracy equivalent to comparable paired statistical-dynamical models and suitable for economic impact assessment models (Bloemendaal et al. 2020b; Meiler et al. 2022). However, to provide

context, we performed several comparisons relevant to our eastern US and post-landfall specific application. Given STORM's focus on tropical storms and hurricanes. This is a relevant distinction because while weaker tropical depression may not account for significant hazards, other classes of storms, like storms that tat have made an extratropical transition, may display appreciably different behavior, like persisting and traveling for longer over land.

We compared STORM data against historical data filtered to these categories using the "USA_STATUS" designation in the historical International Best Tracks Archive for Climate Stewardship (IBTrACS) dataset (Knapp et al. 2010). In addition, we compared this subset against unfiltered historical to give context for possible implications of this design choice. We compared the number of unique storm events per year, on average, over the simulated period to equivalent historical data using historical 1980-2022 storm track data. We also modeled windspeed decay in the hours after landfall to contextualize how accurately STORM could represent this type of exposure, and the ratio of tropical cyclone track coordinates spent overland (using geometric intersection to track landfalls). Finally, we performed wind field and exposure-response function modeling on both the filtered and full historical data to compare the geographic extents and estimated excess mortality figures, providing context for likely biases of the STORM model given its operational scope of modeling only hurricane and tropical storms.

2.6 Exposure Response Function and Health Impact Predictions

We used an existing epidemiological model for the association between county-level tropical cyclone-associated maximum sustained windspeeds and corresponding excess all-cause mortality over the 12 days after exposure. This model uses national-scale trends to makes full use of this sample size, while using a complementary sub-model to make county-level adjustments based on more granular data (e.g. relevant demographic traits, storm-specific features like peak, rather than sustained, windspeeds), providing greater specificity. This model was developed using 1998-2018 HURDAT historical tropical cyclone data, and American Medicare claims data (Nethery et al. 2020). The model uses a Bayesian framework to combine national-scale and county-scale sub-models. Specifically, the framework incorporates an integrated causal-inference and predictive model, leveraging both the large sample size enabled by the national scale and a range of socioeconomically- and exposure- relevant variables, using census data (proportion of residents under the poverty line, proportion identifying as racially white, proportion of 65-and-older residents, median age, graduate without a high school diploma, population density, proportion of owner-occupied houses, and median home value). The exposure-response function model allows lag time between

exposures and health outcomes, covering a 14-day lag period. This period is 2 days before, and 11 after, exposure, a choice reflecting evidence of increased respiratory illness-based hospitalizations in the two days before a hurricane (Yan et al. 2021). The model uses a minimum exposure threshold of 17.5 m/s for maximum sustained windspeeds, corresponding to a tropical storm on the Saffir-Simson scale.

In our implementation of the model, we used 2020 decennial census figures as accessed with the *tidycensus* package (US Census Bureau 2022; Walker and Herman 2022) as well as several additional exposure-relevant variables (maximum surface-level gust windspeeds, duration of sustained surface-level wind speeds above 20 m/s, duration of gusts, number of previous exposures over the 1988-2015 training data, the year of exposure, and an indicator variable for Census-categorized coastal counties). Given the temporally ambiguous nature of the STORM data, we opted to uniformly assign 2015 (the last year in the Nethery model's training data) to each synthetic tropical cyclone event, as well as to the filtered historical data to maintain consistency (so strictly speaking, our historical data is modeling the outcomes of storms equivalent to historical tracks impacting a roughly present-day population); it should be noted that of these variables, only maximum sustained windspeed had a consistent, uniformly positive association with excess all-cause mortality, although this was not uniformly the case for other health outcomes (Nethery et al. 2020). We excluded two counties (Kenedy, TX and Loving, TX; total estimated populations of 391 and 119, respectively) on the basis of missing median house value data and matched the exposure-response function's minimum inclusion threshold by excluding two additional Texas counties with fewer than 100 over-65 residents (King and McMullen Counties, at 31 and 96, respectively).

While the model was fit for several health outcomes, we selected all-cause mortality, (Figure 1.1), which shows a somewhat distinctive profile for excess mortality rate, with a protective effect for lower windspeeds; other outcomes like hospitalization generally increased steadily with greater maximum sustained windspeed exposures (Nethery et al. 2020). The model expresses this as excess events (deaths in this case) per 100,000 residents, with factors like underlying mortality rates captured implicitly in the training process. The exposure-response function for all-cause mortality (Figure S1.1) is non-monotonic, with attributable mortality decreasing until protective effect of ~12 fewer deaths per 100,000 is reached in the low 20 m/s range, before increasing back to around zero by 30 m/s and increasing monotonically with windspeeds from this point, reaching a maximum central tendency of around 150 excess deaths per 100,000 residents by 54 m/s, albeit with widening confidence intervals on account of the growing rarity of such high sustained windspeeds in the training data. We used the exposure-response function to project excess

tropical cyclone-attributable mortality events by combining our calculated wind field-based county exposures (maximum sustained surface-level windspeeds for the primary casual inference model component, and several related metrics for the predictive model component) with corresponding population and demographic data relevant to the model.

2.7 Uncertainty Propagation

With our approach incorporating multiple sources of uncertainty, we needed to take a comprehensive approach to uncertainty, capable of reflecting both variability in the tropical cyclone-related exposure, and in predicted excess mortality outcomes of this exposure. To this end, we used a Monte Carlo approach, where the interacting components (the STORM-derived county exposures and the Bayesian exposure-mortality model) are run many times, with summary statistics being generated from the pool of samples, which empirically represent the probability space of the system; we report a 95% "empirical confidence interval" (eCI) representing the 2.5th through 97.5th percentiles of the data, using the median (50th percentile) as a metric of central tendency. The integrated exposure-response function has two sets of variability. First, the causal inference module generates a central tendency for excess mortality

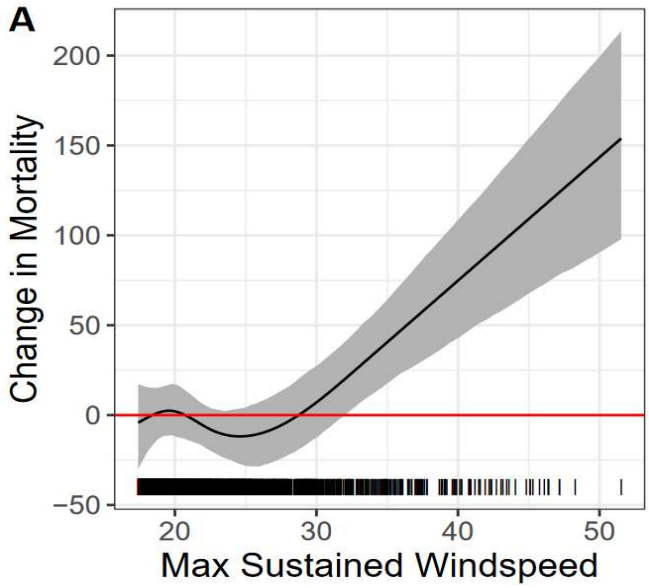

Figure 1.1: exposure-response function curve linking maximum sustained windspeed (meters per second) to change in excess mortality rates per 100,000 county residents; (Reproduced from Nethery et al 2020; Figure 4A; CC-BY 4.0 2021)

(per 100,000 residents) based on the entire study population, which is modified by the predictive model component based on county- and storm-specific traits representing a range of possible results in the model space (Nethery et al.

2020). In addition to uncertainty in the exposure-response function probability, the synthetic tropical cyclone seasons effectively represent another component in a composite probability space by representing a range of plausible tropical cyclone activity. We combined these approaches to generate more comprehensive uncertainty intervals by resampling the exposure-response function model 100 times per storm event, theoretically representing 10,000 seasons of data. For a given geographic level of analysis, we performed the . We took an empirical 95% confidence interval (2.5%, 50%, and 97.5% percentiles) of the resulting datasets.

2.8 Summary Statistics

Because we were interested in assessing overall patterns of tropical cyclone-associated impacts in the region of interest, rather than of any single storms, we generated average numbers of county-level exposure events (cyclone days) per decade, along with associated cyclone person-days, and total numbers of attributable deaths as calculated by the exposure-response function model (incorporating both maximum sustained windspeed and inputs for the predictive component of the integrated model). We took single-timepoint maximum sustained windspeed for each county-and-storm combination. We divided the sum of these exposure events by the number of decades to obtain a per-decade exposure metric. The exposure-response function generated 1000 estimated attributable mortality (excess deaths per 100,000) rates per cyclone day, so to obtain corresponding mortality counts, we calculated a 95% uncertainty interval using the 2.5^{th}, 50^{th}, and 97.5^{th} percentile values for each set of estimates, and taking temporal sums across each county, as well as spatially for state and regional totals. In addition to performing the above summary steps across all tropical cyclones, we also performed summaries within each genesis month (June-November in the STORM data, corresponding to ~97% of historical storms in a 1988-2018 HURDAT comparison).

3. Results

3.1: Agreement Between Synthetic and Historical Datasets

With our investigation addressing the baseline viability of STORM (as a representative synthetic tropical cyclone generation algorithm) as an HIA component, it is important to address the extent of agreement, and ground expectations for the model. We observed roughly 45% fewer synthetic than historical (1980-2022—specifically tropical storms and hurricanes) in our specific study region (2.71 vs 4.86 per year, respectively). By contrast, the original STORM methods paper found a closer agreement for North Atlantic basin-wide historical genesis rates at

10.9 (standard deviation of 3.9), compared to 10.8 (standard deviation of 3.3 historically) in the authors' estimation (Bloemendaal et al. 2020b). In our region of interest, there was closer agreement among all included storms that make landfall (2.71 STORM vs 4.07 historical), but a larger gap for hurricanes (reaching central track windspeeds of at least 64 knots) at 1.10 vs 2.29, respectively. This suggests that for our use case, STORM likely underestimates the net impacts (negative or positive) of a typical hurricane season; At the same time, the genesis rate could theoretically be adjustable, the fact that the basin-wide values were closer could reflect a more systematic spatial bias, and may reflect the lower relative rate of predicted landfall events (6.0 with a standard deviation of 3.3 in STORM vs 8.2 with a standard deviation of 5.6 historically) (Bloemendaal et al. 2020b).

Within our study region and use case, the relevant characteristics of individual simulated tropical storms and hurricanes tracks were quite close to their historical counterparts. Central STORM track windspeeds had an interquartile range of 20.7 - 35.9 m/s, compared to 23.2 - 36 m/s in the historical tropical storms and hurricanes; roughly 0.02% of STORM windspeeds fell below 18 m/s (likely an interpolation artifact, as 18 m/s was the minimum windspeed for STORM training data), but this did not influence the interquartile range to at least 2 decimal places.

As far as windspeeds after landfall, STORM again resembled the filtered historical data when we resampled the data in 42-year subsamples, an interval chosen to match the available period of historical data. This metric gave a mean 26.7% of track coordinates over land (full range of 24.5% to 29.6% across subsamples), compared to 24.3% of historical track coordinates after filtering to tropical storms and hurricanes only. However, the unfiltered (i.e. beyond tropical storms and hurricanes) historical tracks had 36.9% of coordinates over land, a difference consistent with the longer persistence of storm events like extratropical and subtropical depressions. STORM landfalls also showed a considerably more even spatial distribution than the historical equivalent (Figure 1.3), with e.g. Florida, the Carolinas, and New England having a smaller difference in expected exposures per decade than in the historical data. This greater homogeneity could be a result of statistical smoothing from the larger sample size but may also indicate a loss of relevant geographic and current-based drivers of tropical cyclone risk. Windspeed decay and (nominal) storm persistence after landfall was broadly similar in both STORM and the filtered 1980-2022 dataset (Figure S1.3, showing a composite central tendency of windspeeds) although STORM showed a more convoluted windspeed trend over time than the historical data, with a steeper immediate decline and lower windspeeds in roughly the first day post-landfall, which returned to comparable windspeeds before a secondary

departure—this time to higher windspeeds than this historical precedent—around 100 hours post-landfall through a maximum of around 150 hours, which applied to both datasets. As one might expect given this pattern in wind decay, this translated into a shorter distance traveled inland for both STORM and the filtered historical data than the full historical dataset (Figure S1.3), which reached as far as the Great Lakes region for some storms like Hurricane Ike (2008).

3.2 Population Distributions

Our operationally defined study area (counties experiencing at least one gale-force exposure in the STORM dataset) encompassed 1490 counties with a composite 65-and-older population of roughly 27.8 million as of the 2020 census, compared to 919 counties in the 1980-2022 data after filtering to tropical storms and cyclones, or 1361 without filtering. Of these, 1054 counties (population of 9.6 million) saw at least one hurricane-force exposure during the simulated period. Between underlying population density and age demographics, population distribution was highly heterogeneous, with the 10 most-populated counties accounting for more than 15% of our 65-and-older study population. Harris County, TX, and Miami-Dade County, FL had the largest single 65-and-older populations at 491,390 and 440,203, respectively, although the five constituent counties of New York City had a combined population of nearly 1.5 million; at the state level, Florida had the largest population at 4,347,912 exposed age 65+ residents, followed by New York at 2,908,071 and Texas at 1,991,991. Figure 1.2 shows this population, and Table S1.1 presents a summary of county populations and relevant socioeconomic and demographic traits used in the exposure-response function.

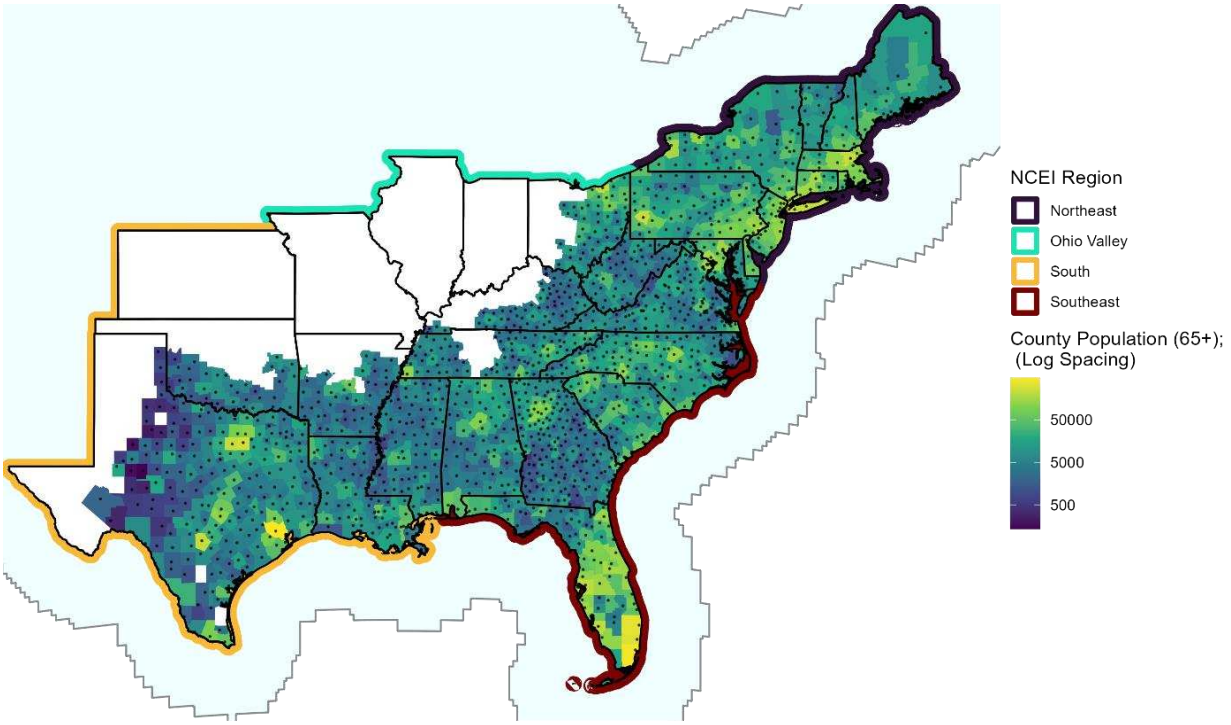

Figure 1.2: Map of Study Population, with shading to indicate population (log spacing) of 65-and-older residents as of the 2020 5-year American Community Survey; areas used for regional analysis are indicated in exterior outlines, with the 250km buffer used in inclusion criteria indicated as a shaded region surrounding the coast. Points indicate population-weighed county centers. Unshaded counties indicate either those not projected to experience exposures at our tropical cyclone inclusion threshold or counties excluded due to having <100 65-and-older residents, to meet exposure-response function inclusion criteria.

3.3 Exposures (Cyclone Days) and Cyclone Person-Days

Over the 10,000 synthetic STORM-generated seasons used in our analysis, there were 17,959 unique tropical cyclones meeting filtering criteria (~18 per decade), compared to 136, or 32.4 per decade, in the 1980-2022 data (after filtering to only tropical storms and hurricanes). The STORM tropical cyclones gave rise to 739,652 unique county-level "cyclone day" exposure events, 77,483 of which had at least hurricane-force (≥32.9m/s) windspeeds across counties in the study area, or roughly 77.5 county-level exposure events per decade, compared to 84.0 in the filtered historical data. Unsurprisingly, there were more cyclone days along the coasts, with an even falloff in exposures over roughly 200km inland (figure 1.2); there were notably fewer maximum cyclone-days in any single county, with a maximum around 2.0, where the equivalent 1980-2022 data (filtered to hurricanes and tropical cyclones) had localized

peaks of nearly 7 days per decade near the North Carolina Barrier Islands. STORM exposures were also more uniformly distributed, whereas impacts from specific storms or sets of storms are more visible in the historical data (e.g. even the eastern and western borders of the Louisia coast show distinct 'hotspots').

The large sample size of STORM data also enables exploration of month-to-month cyclone-day exposure trends (Figure S1.2). While there are overall national commonalities like August and September seeing the highest overall expected frequency of gale-force cyclone-days per decade by a large margin, some notable spatiotemporal patterns emerge—for instance, the Gulf of Mexico-facing coasts saw more exposures at the early than late boundaries of the season, hurricane-force exposures were most common in Texas during August. However, gale-force exposures appeared farther inland later in the season, and Maine saw its (admittedly low) peak hurricane risk in August. As noted above, these represent exposures generated specifically by tropical storms and hurricanes, so some lower-windspeed exposures—particularly farther inland and in the north—would not be captured in this dataset compared to the full 1980-2022 dataset (Figure S1.3). In addition, it is worth noting that the specific windspeeds of tropical cyclone exposures, beyond the broad gale-force compared to hurricane-force comparison used here, play a key role in shaping the predicted exposure-mortality connection.

Plotting median windspeeds (Figure S1.4) provides some context (although in even rare extreme outliers could play a strong role); the distributions are largely like the total cyclone-day exposure map (Fig 1.3), with median windspeeds roughly between 17 and 32 m/s, but some differences emerge—e.g. somewhat counterintuitively, the inland Florida peninsular region has relatively low median windspeeds, possibly as a result of more frequent exposure to tropical cyclone events overall, and some regions like the Florida panhandle area or central Louisiana coast see relatively high median windspeeds; even with 10,000 simulated seasons, some seldomly-exposed areas like parts of central Texas see fairly "patchy" distributions of windspeed, likely representing rare one-off events; this could also be a feature of the pattern seen in post-landfall wind decay in STORM, where long-lived storms seem to get a "second wind" and pick up windspeed even over 100 hours after landfall, departing somewhat from the historical precedent (if not completely outside the uncertainty range).

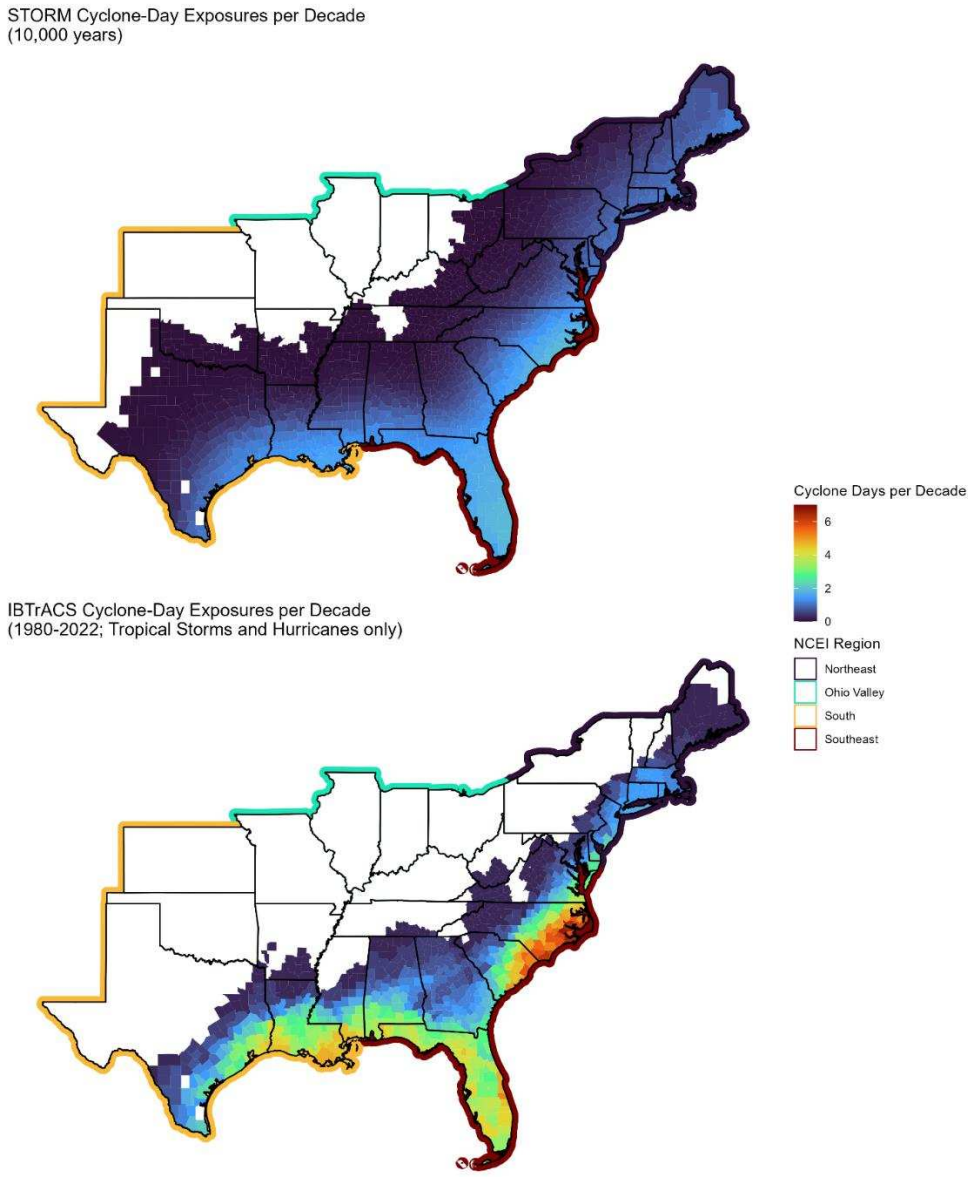

Figure 1.3: Comparison of Tropical Cyclone Exposures (Cyclone Days) per Decade in STORM (10,00 seasons) and filtered IBTrACS (1980-2022) datasets. Map of Cyclone-Day Exposures (sustained county-level windspeeds of 17.5 m/s or greater) per Decade in 10,000 seasons of simulated STORM data (top) compared to 1980-2022 historical data (bottom) filtered to tropical storms and hurricanes only to approximate STORM training data. Here, we present total exposures (sum of hurricane-force and sub-hurricane [gale through violent storm on Beaufort Scale). Counties not expected to experience at least gale-force maximum sustained windspeeds are unshaded. Maximum STORM cyclone-days per decade is 1.91, maximum historical cyclone-days is 6.67.

We combined cyclone-day exposures and corresponding county populations into a derived "cyclone person-days" metric (Figure 1.4) for perspective on the combined risk profile of physical hazards and population distributions. Florida, given its combination of high-population counties and frequent tropical cyclone exposures,

had the most cyclone person-days (among our 65-and-older study population) by a wide margin, with 7,343,068 predicted cyclone person-days in a decade to New York's 2,144,032 and Texas' 1,622,122, or 21,787,383 cyclone person-days per decade nationwide. Miami-Dade County, with 440,203 residents and frequent tropical cyclone exposures, saw the most cyclone person-days in the nation at an average 791,925 per decade. However, population generally had a stronger influence than storm exposure. For instance, the high population of the Houston metropolitan area and surrounding regions yielded a conspicuously high cyclone person-day metric for the wider coastal region, with New York City and the surrounding region showing a similar population-driven trend. Conversely, coastal North Carolina's smaller population yielded low cyclone person-days, despite having the second-highest total county-level exposures at 97.13 per decade to Florida's 104.7. While total exposures can give context, the actual hazard is determined by specific maximum sustained windspeeds, leading to counterintuitive results for total all-cause mortality estimates as we will discuss in the following section.

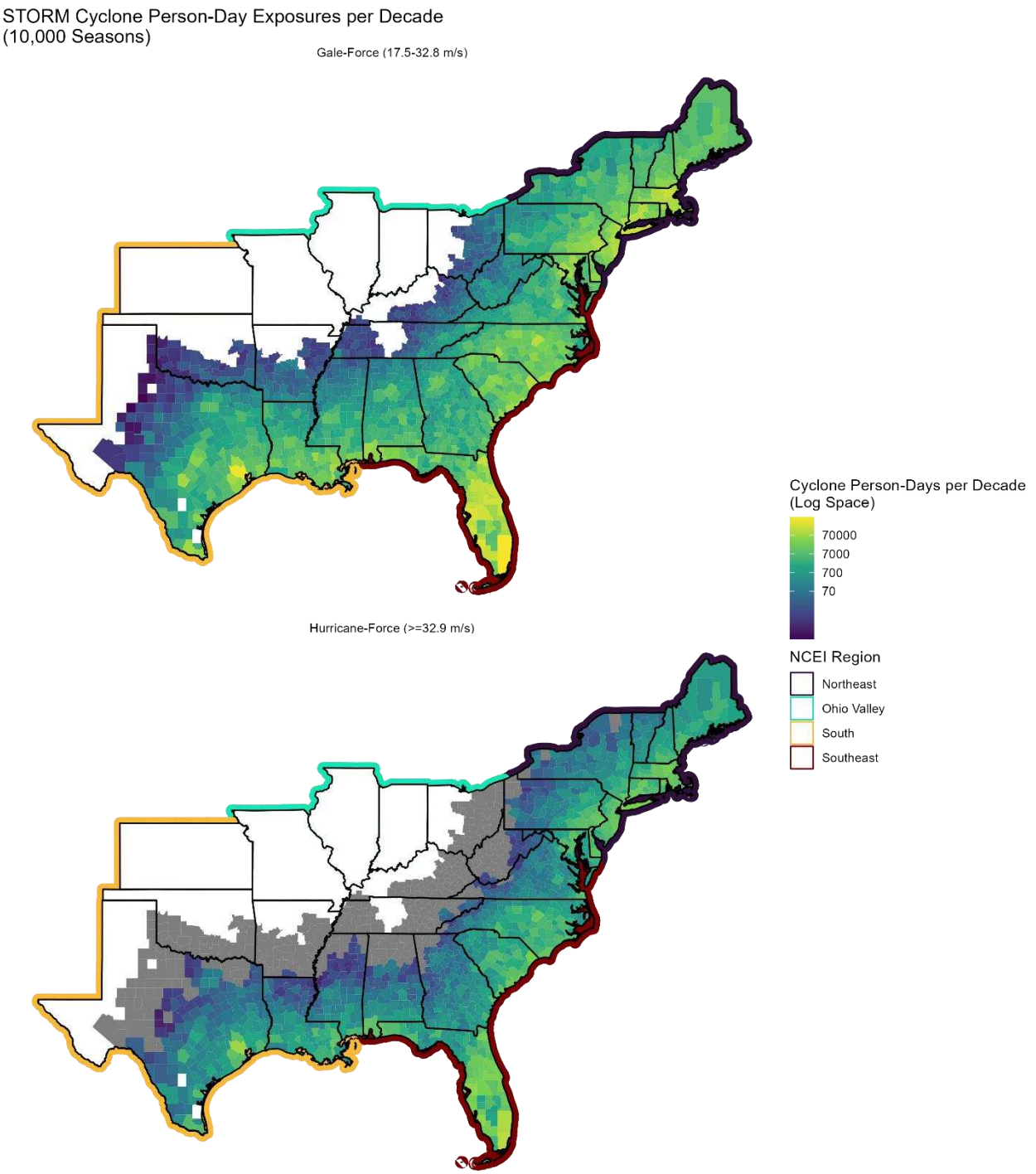

Figure 1.4: Map of STORM Cyclone Person-Day Exposures (sustained county-level windspeeds of 17.5 m/s or greater multiplied by county population) per Decade in 10,000 simulated seasons. (top). We divide results into gale-force (operationally, maximum sustained windspeeds of 17.5 m/s through 32.8 m/s, corresponding to a gale through violent storm on the Beaufort scale) and hurricane-force cyclone person-days (bottom; maximum sustained windspeeds of ≥32.9 m/s). We indicate exposures per decade using shading, with log-spaced intervals to communicate the range of values better; grey-shaded counties indicate locations not expected to see hurricane-force exposures.

3.4 Hurricane-Attributable Mortality

Nationwide, our analysis returned an expected 1,823 (95% empirical confidence interval 962, 2745) excess deaths per decade attributable to hurricane-force exposures, but -2944 (95% eCI -8448, 1696) from gale-force exposures, leading to a broad net predicted impact of -1197 (95% eCI -7709, 4216). Given the counterintuitively nonlinear nature of the windspeed-mortality exposure-response function used in this investigation, it may be useful to discuss hurricane-specific and net tropical cyclone-associated projected mortality impacts separately. Focusing on hurricane-force exposures (sustained maximum windspeeds of ≥32.9 m/s), where a monotonic trend exists between windspeed and excess mortality rates, the Southeast saw the highest expected excess mortality at 1151 (95% eCI 735, 1703) excess deaths per decade, followed somewhat distantly by the Northeast at 491 (95% eCI 129, 780) and South at 171 (95% eCI 96, 264).

However, these regional trends obscure some more granular trends, in some cases differing from the cyclone person-days discussed above, with e.g. Florida seeing the highest mortality rate at a predicted 815 (95% eCI 536, 1205) excess deaths per decade, followed by North Carolina at 135 (95% eCI 89, 202), with the intensity of exposures counterbalancing the smaller cyclone person-day metric than e.g. New York, which was nonetheless third in the nation at 133 (95% eCI 133 (22, 210) predicted deaths, and Texas at 106 (95% eCI 53, 166) as one of the few highly-impacted Southern state (with Louisiana at 52 (95% eCI 33, 83) being surprisingly low), and South Carolina at 101 (95% eCI 65, 153). Our model did not show the Ohio Valley region having deaths from these exposures, with only the 95th percentile of estimates passing a single death. However, to focus on the gale-force part of the tropical cyclone exposure spectrum, the exposure-response function's predicted protective effect at these windspeeds means that the cumulative effect of hurricane seasons looks very different than the above results would suggest; this effect led to an effective reversal in the above order, with seldom-exposed states (e.g. Kentucky, Oklahoma) seeing near-zero expected attributable deaths (e.g. -1 (95% eCI -1, 0) in Arkansas), while more-impacted states tended to see negative central estimates like -90 (95% eCI -492, 334) excess deaths per decade in Texas, Florida with an expected net impact of -60 (95% 1993, 1973) excess deaths per decade, New York at -46 (95% eCI -778, 657) and North Carolina at -54 (95% eCI -446, 389). In these cases, it is apparent that a given tropical cyclone season could have potentially highly damaging public health effects within the predicted confidence intervals, even with a net 'protective' effect observed.

We also considered region-level mortality trends throughout the season (Figure 1.5). August and September consistently saw the deadliest impacts for hurricane-strength exposures, with the Southeast's 50th percentile estimates approaching 500 expected excess deaths per decade, and the Northeast also seeing similar rates at 227 and 210, respectively, although in both cases the two months' empirical confidence intervals largely overlapped; the South differed from this trend in having more deaths in August than September, with nearly double the expected attributable deaths. Cumulatively, the mixed-to-protective mortality effects of sub-hurricane-strength exposures, combined with the greater frequency of these events, yielded fewer net attributable mortalities (i.e. more deaths would be averted than caused on average when considering all tropical cyclone strengths).

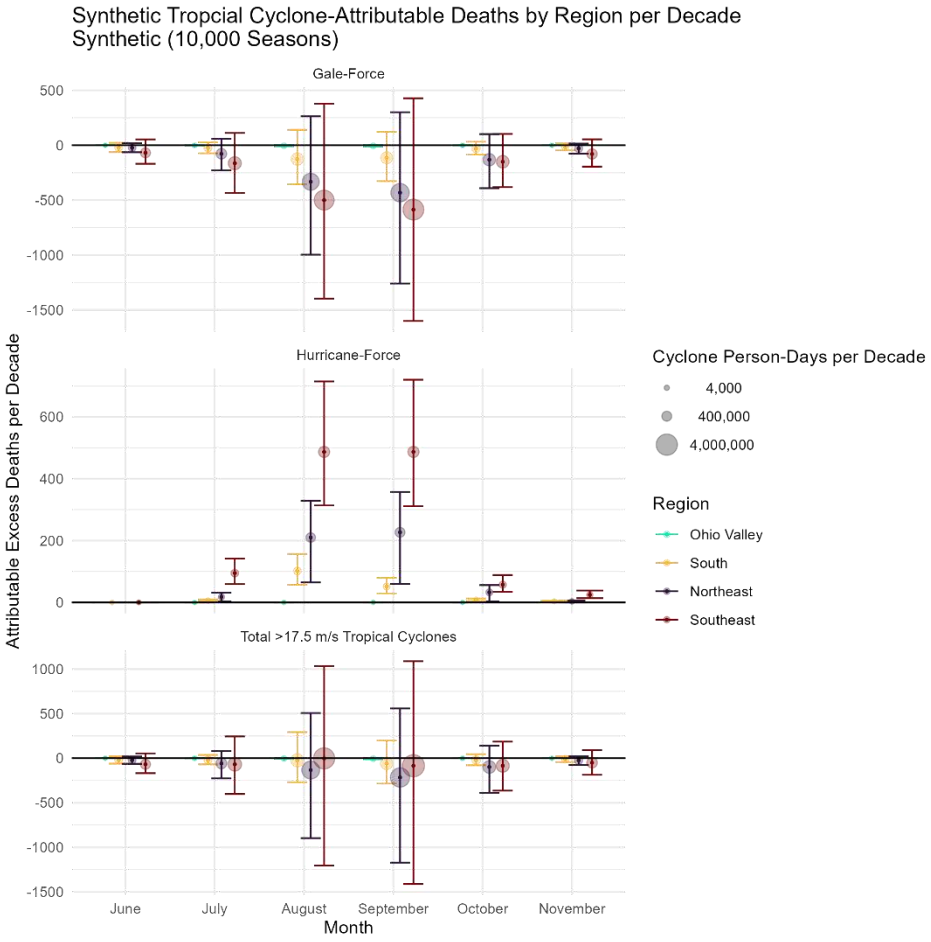

Figure 1.5: Summary of NCEI regional-level attributable mortality from monthly hurricane-force exposures ("cyclone days" with a maximum sustained windspeed of at least 17.5 m/s). Points and error bars represent excess mortality per decade as a 95% empirical confidence interval (2.5%, 50%, and 97.5% percentiles), with the point sizes indicating several corresponding cyclone person-days per decade.

Returning to comparisons between historical and STORM-generated tropical cyclone exposures, our national comparison of estimated excess mortality in the 1980-2022 data with all storm types, compared to only

tropical storms and hurricanes, showed similar hurricane-force associated mortality trends at an expected 214 (95% eCI 96.4, 340) annual excess deaths in the full data compared to 209 (95% eCI 97.5, 326) after filtering, but a wider deviation for sub-hurricane-force exposures at -677 (95% eCI -1895, 730) and -543 (95% eCI -1550, 622) annual averted excess deaths, respectively, corresponding to a roughly 20% shortfall (with relative absence of lower windspeeds, and of total exposures given the shorter duration, resulting in a smaller protective effect). Assuming this difference translates to the STORM data, one might expect a similar mortality rate from hurricane-force exposures on a per-storm basis, although lower total mortality due to the under-representation of storms in general, and about 25% higher averted mortality for gale-force winds (again, with roughly 30% higher net impacts in terms of cumulative storms).

4. Discussion

In this investigation, we explored the viability of the STORM algorithm as a tool for projecting all-cause mortality risks among 65-and-older Americans associated with tropical cyclones in an integrated causal-predicted Bayesian model as part of a quantitative health impact assessment. This investigation is, to our knowledge, the first use of synthetic tropical cyclone tracks in this use case. This approach has several strengths, notably the modularity of the quantitative HIA framework. Using a population-scale exposure-response function (exposure-response function) model that incorporates two weeks of lag-time post-impact (allowing more time for delayed effects like disrupted medical and electrical infrastructure, or illnesses associated with floodwater and other physical hazards, to be detected) also proved to be useful, projecting a national average of 1823 (95% eCI 963, 2745) expected attributable hurricane-associated deaths per decade in the simulated data, and 2095 (95% eCI 975, 3263) in the equivalent historical data. This suggests the exposure-response function used in this model was able to detect about four times the 48 deaths per year (480 per decade) of deaths estimated by the US National Weather Service estimate (US Department of Commerce) than could be directly attributed from e.g. injury in the aftermath of a hurricane, although the disparity could be greater given our study's focus on 65-and-older adults. In a larger departure from a short-term impact-based analysis, the exposure-response function also included a protective effect for some sub-hurricane-force tropical cyclone exposures, which carries over to cumulative hurricane season effects.

By combining the exposure-response function with a large (10,000-season) synthetic dataset of tropical cyclone Tracks, we were able to explore these patterns spatially and temporally (throughout a hurricane season) with a more comprehensive picture of tropical cyclone-associated exposures and health outcomes than purely historical data could achieve. While this exposure-response function model is tailored to a narrow population (Americans enrolled in Medicare from 1999-2015) and building specifically on US Census data collection and methodology (meaning that e.g. transferability to even adjacent countries would be limited), the exposure and health outcomes (within the 3 types of hospitalizations explored in the paper could be expanded.

Our results suggest that STORM is capable of approximating, on a storm-by-storm basis, hurricanes and tropical storms compared to similar historical (1980-2002) data. However, STORM data differs in having a lower average annual frequency of storms, as well as a slightly lower-than-historical fraction of tropical cyclone tracks over land. The large sample size (10,000 synthetic seasons) made possible by STORM enabled us to explore a spatially and chronologically granular map of tropical cyclone exposure risks at the county scale, as well as effectively "smoothing out" the heterogeneity introduced by specific high-impact storms, or lack thereof, in analogous historical data.

For instance, while the Carolinas are among the most-impacted regions in both cases, the STORM data puts the frequency at a more comparable level to the Florida peninsula, rather than 1.5-2x the rate as would be inferred from a straightforward frequentist reading of 1980-2022 storm tracks. While atmospheric or oceanic science field experts may be able to make these inferences easily, tools like STORM (to the extent that they reflect real-world climate interactions) could play a valuable tool in bridging disciplines and allowing wider interdisciplinary exploration of tropical cyclone implications across various fields.

At the same time, STORM, while judged to be similar to the accuracy (in terms of ability to recapitulate historical traits and tropical cyclone distributions by basin) of comparable models in a recent comparison (Meiler et al. 2022), does have known limitations—notably, a lower-than-historical (if within standard deviation) frequency of landfall events, a lower frequency of sub-hurricane-strength tropical cyclones, and an overrepresentation of high-intensity storms. In this respect, it is important to acknowledge the operational scope of the STORM dataset, with hurricanes and tropical storms being represented fairly accurately (if at a somewhat lower genesis rate in our region of interest—because the overall North American basin was well within the historical precedent, this may represent a

spatial bias, rather than issue with genesis rates), while the lack of e.g. extra- and subtropical depressions could be of concern for modeling inland or delayed health or economic impacts.

Our characterization of the expected combined population exposure to all-cause mortality impacts from simulated tropical cyclones among older adults in the eastern United States was another unique feature of the methodology. The present implementation did not prove to be a perfect predictor of expected tropical cyclone-attributable mortality in our implementation, with population having an outsize influence. Weighting windspeeds as part of a derived metric could provide useful context for long-term, low-frequency risk profiles associated with this exposure type among our population of interest. Along similar lines, our use of simplified Beaufort Wind Scale categories (compressing all sub-hurricane wind speeds into a single generic "gale" level) gave wide confidence intervals in the "gale" category yielding wide confidence intervals. This is to some degree driven by the relatively high uncertainty for lower-windspeed gale exposures, but making e.g. an intermediate subdivision of perhaps 25 knots (e.g. perhaps the Beaufort Scale's Storm through Violent Storm categories, spanning 48 through 63 knots, or 24.7-32.4 m/s) could capture the nadir of the exposure-response function curve (and hence the most reliably "protective" effects, giving a more informative picture of risks.

Regarding the counterintuitive protective effect of especially intermediate-windspeed exposures, future research in international contexts (or even different age groups in the US) could shed light on whether this is a broadly generalizable feature, or something specific to our study system. For instance, the extent to which factors like favorable geography, effective infrastructure and emergency preparedness and responses help buffer US populations from the physical impacts of tropical cyclone exposure is made clear by comparing this estimate against other regions in the study—for instance, the authors' operational definition of Latin America and the Caribbean, despite the proximity to the tropical cyclone basins, with the equivalent period seeing 21.3 (95% empirical CI 15.8, 28.2) deaths per 100,000, only slightly lagging Southeast Asia (29.3) for the highest rate of the surveyed regions. Given the difference even for nearby Florida in our study, this supports the idea that the protective effect observed for of sub-hurricane-force exposures in the Nethery et al (2020) model should not be taken for granted, and instead likely represents a combination of factors including effective infrastructure, preparedness, and emergency and recovery response measures, in addition to more intrinsic features like geography and water temperature.

Overall, our results support the viability of using a statistical tropical cyclone projection algorithm like STORM for health impact assessments, if with some caveats about implementation given the lower overall frequency of storms in the study area, and absence of events like extra- -and subtropical depressions, which pose a different risk profile given their longer overall effective lifetimes, lower windspeeds, and persistence post-landfall. In this context, STORM should fairly accurately project hurricane-driven all-cause mortality (given that this band of exposures, and post-landfall decay, is modeled relatively accurately) on a per-storm basis, although the annual average is likely lower than its historical equivalent. On the other hand, despite the similarly accurate representation of tropical storms between 18 and 32.9 m/s, the absence of e.g. sub- and extratropical storms from simulated hurricane seasons will lead to a weaker "protective" than would be predicted using a fully modeled scenario given the comparative scarcity of lower-windspeed exposures.

With the protective effects for relative risk at lower windspeeds is specific to the association between maximum sustained windspeeds all-cause mortality in the Nethery model, with e.g. respiratory illness hospitalizations increasing more monotonically with windspeed, meaning the equivalent STORM estimate would be more of a straightforward underestimate of true impacts; similarly, the full historical dataset shows storms moving considerably farther inland, so exposures to inland regions like the Ohio Valley and even the Great Lakes region are essentially unrepresented in the STORM dataset. An important note is that while the STORM data differs from equivalent historical data in these respects, 1980-2022 data shows hurricanes and tropical storms account for the majority of overall predicted hurricane season mortality (negative and positive), and more fundamentally, these are less intrinsic limitations of a statistical model like STORM than a reflection of STORM's intended use case (which focuses more on behavior in oceanic basins up to the point of landfall) and the authors' selected performance metrics, which include features like basin-wide storm genesis rates and maximum sustained windspeeds (Bloemendaal et al. 2020b). In this respect, it wouldn't be unfeasible for a similar algorithm, or modified version of the STORM code (given its open-source nature) to be developed with more emphasis on representing post-landfall behavior and non-tropical cyclones and depressions.

Of course, the study (especially in its current state) carries several limitations, both intrinsic to the overall approach and associated with our specific implementation. As a quantitative HIA, our methodology inherits any limitations of the underlying models and datasets. Most prominently, STORM has a narrower operational definition of a hurricane season (with the lack of e.g. extra-and subtropical depressions), and more seriously appears to

underestimate the baseline frequency of storms (even when comparing only tropical storms and hurricanes) by about 45%, and 50% for hurricanes specifically. However, given the weaker wind speeds and lower frequencies, the former was projected to have a small but appreciable impact on expected mortality trends, with about 80% of the protective effect of sub-hurricane-force exposures, and mortality counts attributable to hurricane-force exposures appeared largely unchanged. As an example of a more granular limitation, using 65-and-older residents as a proxy for Medicare recipients is a close, but not exact, match, with around 10% of recipients being below this age and qualifying through other means. Our approach also makes some operational assumptions—for instance, maximum sustained windspeed is assumed to cause the excess mortality for modeling purposes, while it is in some respects more of a useful indicator of general storm intensity (meaning that e.g. precipitation and floodwaters are effectively implicit in the relationship), as well as intermediate casual mechanisms (for instance, disrupted electrical grids and transport) which could theoretically be weakened by 'hardened' infrastructure or other resilience measures. At least within the modeled population and system, the fact that e.g. gust wind strength in the predictive model did not have a consistent effect does provide supporting evidence for the strength of this assumption.

More broadly, our exposure-response function, while well-defined, does narrow the scope of generalizability more than would be the case with some datasets, although this is a tradeoff with the high temporal resolution—for instance, a similar modeling method was recently applied at the county level to the entire United States using an interpolated (concerning mortality) monthly resolution (Parks et al. 2022); however, this is not impossible—for instance, another study drawing on a collaborative global database (the MCC Collaborative Research Network) was able to perform an international analysis on a daily resolution tracking lag times of up to 14 days and using a distributed lag nonlinear model framework for the exposure-response function (Huang et al. 2023). At present, we are only focusing on all-cause mortality. Still, we could expand the analysis to include additional modeled health outcomes (hospitalizations from respiratory conditions, cardiovascular diseases, and chronic obstructive pulmonary disorder) in Nethery et al. (2020) for a fuller picture of tropical cyclone-associated health impacts. This approach has several strengths, notably the modularity of the quantitative HIA framework. While our specific exposure-response function model is tailored to a narrow population (Americans enrolled in Medicare from 1999-2015) and building specifically on US Census data collection and methodology (meaning that e.g. transferability to even adjacent countries would be limited), the exposure and health outcomes (within the 3 types of hospitalizations explored in the paper could be expanded.

Our study fits in a growing ecosystem of interdisciplinary research into both synthetic tropical storms and exploring delayed and indirect health impacts of tropical storms. For instance, Meiler et al (2022) use STORM alongside other synthetic tropical cyclone models as input for the open-source CLIMADA (CLIMateADAptation) impact modeling platform (Bresch and Aznar-Siguan 2021) to model economic impacts of storms with a 10km grid of projected "asset values", rather than e.g. population and mortality as input. On the epidemiological side, a recent United States-based study modeled the association between historical tropical cyclone data and an all-age National Center for Health Statistics mortality dataset at an interpolated monthly resolution (Parks et al. 2022). Another study applied a different but comparable exposure-response methodology to global 0.5° (roughly 55km) gridded data, making a similar delineation between hurricane-strength and sub-hurricane-strength exposures, with mortality data and associations extrapolated from a 1630-site population dataset sourced by a global collaborative effort (Huang et al. 2023). Our exploration of using synthetic exposure data sets it apart from these and comparable studies, allowing researchers to potentially augment the limited scope of historical data, or explore alternate climate scenarios.

This flexibility makes the methodology outlined in this paper a strong basis for future exploration—for instance, a classification metric like NOAA's multivariate MEI.v2 scale (Kobayashi et al. 2015), could be used to explore health implications of strong or weak trends in the El Niño Southern Oscillation, or interactions with other recurrent climate patterns. We could also draw on demographic projections to completement future climate scenarios like the mid-21st century projections produced by the STORM research team (Bloemendaal et al. 2022a). Along similar lines, while we have focused on the cumulative impacts of hurricane seasons over time (expressed in terms of confidence intervals for excess mortality), the existing dataset could be readily used to explore other presentations of tropical cyclone-associated risk—e.g. estimating the time that could be expected between storms of a given magnitude (their return period) for a region of interest, as in (Meiler et al. 2022).

Our approach prioritized modularity, but it may be beneficial to consult community members and other stakeholders to future analyses on potential oversights at this level of analysis—for instance, the exposure-response model does not explicitly consider specific geographic or infrastructure features which may influence the relative impact of a given maximum sustained windspeed. Even dynamics not readily "translatable" to this type of analysis could inform limitations and biases, better contextualizing the estimates. A complementary approach could explore

adapting this methodology to a more granular spatial scale, although suitable mortality dataset would need to be located, which could limit the geographic or temporal extent of such an investigation.

Our findings in this pilot investigation can be categorized into those immediately transferrable to public health, and those more relevant to modeling and exposure assessments. For the first group, the STORM data indicates elevated tropical cyclone risk during August and September with some region- and state-level variation, and dramatic variation in attributable mortality under the current windspeed-mortality association. However, at this stage, the main implication is that while exposure-response functions based on historical observations can be successfully applied to synthetic tropical cyclones for more spatially and temporally comprehensive impact estimates, it is important to understand model limitations, with STORM simulations showing some departures from expected historical trends (e.g. a smaller overall number of storms per decade) in our region of interest and somewhat transformative use case.

These limitations are not necessarily an indictment of using STORM or other fully statistical algorithms in this capacity. Rather, it suggests that different modeling parameters to be considered for this specific use case, rather than the basin-wide benchmarks used in the methods paper (Bloemendaal et al. 2020b). For instance, the STORM methods paper tested a number of performance parameters (whole-basin average maximum windspeeds, landfall frequency, TC genesis rates, etc.), with STORM estimates falling within standard deviation of the historical benchmarks (Bloemendaal et al. 2020a). A subsequent economic impact-focused model comparison also concluded that STORM's overall accuracy was sufficient for the authors' use case, and comparable to other tested models (Meiler et al. 2022). With STORM's source code being available on request, it seems plausible, if not trivial, to modify the algorithm with a higher premium on these health-relevant features. Alternatively, the modularity of our overall approach could allow researchers to seamlessly use output from a more landfall-specific model, if developed or located, in place of the STORM data, or take an ensemble-style approach using multiple models. Overall and even with the current uncertainty, this investigation highlights potential use cases for, and limitations of, synthetic tropical cyclone algorithms.

5. Conclusion

In this pilot study, we explored combining synthetic tropical cyclone and epidemiological models in a health impact assessment framework for a detailed risk profile of all-cause mortality among 65-and-older residents

of the Eastern United States. We also considered the intersection of physical tropical cyclone hazards and population centers to generate a weighted person-exposure metric to explore integrated public health risks. Some relevant findings include the Southeast, followed by the Northeast, seeing the greatest typical tropical storm-associated mortality risk, and storm exposure frequency peaking in August and September, with slight regional variation in timing. Due to a predicted protective effect (specific to mortality) for sub-hurricane-strength maximum sustained wind speeds, which were considerably more common than hurricane-strength exposures, the central tendency for modeled seasons was, perhaps counterintuitively, for fewer deaths than a counterfactual baseline scenario, if with wide uncertainty intervals potentially allowing up to several hundred excess deaths in a given season and region. STORM also specifically models tropical storms and hurricanes, which would have a somewhat moderate impact on especially hurricane-specific mortality, but present a partial picture of the complete impacts of a typical hurricane season, especially since the excluded storm types can travel farther than strictly-defined tropical cyclones Some of these limitations may be addressable through parameter modifications to the open-source STORM algorithm, and more fundamentally, the modularity of this quantitative health impact assessment would allow the study's groundwork to be transferred other climate scenarios, health outcomes, or populations with varying degrees of modification.

References

Anderson B, Schumacher A, Guikema S, Quiring S, Ferreri J, Tennant E, et al. 2021. stormwindmodel: Model Tropical Cyclone Wind Speeds. Available: https://github.com/geanders/stormwindmodel [accessed 27 February 2023].

Bloemendaal N, Haigh ID, De Moel H, Muis S, Haarsma RJ, Aerts JCJH. 2020a. Generation of a global synthetic tropical cyclone hazard dataset using STORM. Sci Data 7:40; doi:10.1038/s41597-020-0381-2.

Bloemendaal N, Haigh ID, de Moel H, Muis S, Haarsma RJ, Aerts JCJH. 2020b. Generation of a global synthetic tropical cyclone hazard dataset using STORM. Sci Data 7:40; doi:10.1038/s41597-020-0381-2.

Bloemendaal N, Haigh ID (Ivan), de Moel H (Hans), Muis S, Haarsma RJ (Reindert), Aerts JCJH (Jeroen). 2022a. STORM IBTrACS present climate synthetic tropical cyclone tracks.; doi:10.4121/12706085.

Bloemendaal N, Haigh ID (Ivan), de Moel H (Hans), Muis S, Haarsma RJ (Reindert), Aerts JCJH (Jeroen). 2022b. STORM IBTrACS present climate synthetic tropical cyclone tracks.; doi:10.4121/12706085.V4.

Bresch DN, Aznar-Siguan G. 2021. CLIMADA v1.4.1: towards a globally consistent adaptation options appraisal tool. Geosci Model Dev 14:351–363; doi:10.5194/gmd-14-351-2021.

Doocy S, Dick A, Daniels A, Kirsch TD. 2013. The Human Impact of Tropical Cyclones: a Historical Review of Events 1980-2009 and Systematic Literature Review. PLoS Curr; doi:10.1371/currents.dis.2664354a5571512063ed29d25ffbce74.

Erickson TB, Brooks J, Nilles EJ, Pham PN, Vinck P. 2019. Environmental health effects attributed to toxic and infectious agents following hurricanes, cyclones, flash floods and major hydrometeorological events. Journal of Toxicology and Environmental Health, Part B 22:157–171; doi:10.1080/10937404.2019.1654422.

Feng K, Ouyang M, Lin N. 2022. Tropical cyclone-blackout-heatwave compound hazard resilience in a changing climate. Nat Commun 13:4421; doi:10.1038/s41467-022-32018-4.

Huang W, Li S, Vogt T, Xu R, Tong S, Molina T, et al. 2023. Global short-term mortality risk and burden associated with tropical cyclones from 1980 to 2019: a multi-country time-series study. The Lancet Planetary Health 7:e694–e705; doi:10.1016/S2542-5196(23)00143-2.

Information (NCEI) NC for E. International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4. Available: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C01552 [accessed 26 February 2023].

Karl T, Koss WJ. 1984. Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983.

Kishore N, Marqués D, Mahmud A, Kiang MV, Rodriguez I, Fuller A, et al. 2018. Mortality in Puerto Rico after Hurricane Maria. N Engl J Med 379:162–170; doi:10.1056/NEJMsa1803972.

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ. 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. Bulletin of the American Meteorological Society 91:363–376; doi:10.1175/2009BAMS2755.1.

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, et al. 2015. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. Journal of the Meteorological Society of Japan 93:5–48; doi:10.2151/jmsj.2015-001.

Meiler S, Vogt T, Bloemendaal N, Ciullo A, Lee C-Y, Camargo SJ, et al. 2022. Intercomparison of regional loss estimates from global synthetic tropical cyclone models. Nat Commun 13:6156; doi:10.1038/s41467-022-33918-1.

National Academies of Sciences EM, Division HM, Policy BHS, Disasters CBPAMSMFLS, Cork DL, Yost OC, et al. 2020. *A Framework for Assessing Mortality and Morbidity After Large-Scale Disasters*. National Academies Press.

Nethery RC, Katz-Christy N, Kioumourtzoglou M-A, Parks RM, Schumacher A, Anderson GB. 2020. Integrated causal-predictive machine learning models for tropical cyclone epidemiology. arXiv:201011330 [stat].

Parks RM, Anderson GB, Nethery RC, Navas-Acien A, Dominici F, Kioumourtzoglou M-A. 2021. Tropical cyclone exposure is associated with increased hospitalization rates in older adults. Nat Commun 12:1545; doi:10.1038/s41467-021-21777-1.

Parks RM, Benavides J, Anderson GB, Nethery RC, Navas-Acien A, Dominici F, et al. 2022. Association of Tropical Cyclones With County-Level Mortality in the US. JAMA 327:946; doi:10.1001/jama.2022.1682.

Shukla MA, Woc-Colburn L, Weatherhead JE. 2018. Infectious Diseases in the Aftermath of Hurricanes in the United States. Curr Trop Med Rep 5:217–223; doi:10.1007/s40475-018-0162-6.

Stevens BR, Ashley WS. 2022. Fatal Weather-Related Carbon Monoxide Poisonings in the United States. Weather, Climate, and Society 14:373–386; doi:10.1175/WCAS-D-21-0130.1.

Subramanian R, Ellis A, Torres-Delgado E, Tanzer R, Malings C, Rivera F, et al. 2018. Air Quality in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower Cost Air Quality Monitors. ACS Earth Space Chem 2:1179–1186; doi:10.1021/acsearthspacechem.8b00079.

US Census Bureau. 2022. American Community Survey 2016-2020 5-Year Data Release. Census.gov. Available: https://www.census.gov/newsroom/press-kits/2021/acs-5-year.html [accessed 28 October 2023].

US Department of Commerce N. Weather Related Fatality and Injury Statistics. Available: https://www.weather.gov/hazstat/ [accessed 17 January 2024].

Walker K, Herman M. 2022. tidycensus: Load US Census Boundary and Attribute Data as “tidyverse” and ’sf’-Ready Data Frames. Available: https://walker-data.com/tidycensus/ [accessed 7 July 2022].

Willoughby HE, Darling RWR, Rahn ME. 2006. Parametric Representation of the Primary Hurricane Vortex. Part II: A New Family of Sectionally Continuous Profiles. Monthly Weather Review 134:1102–1120; doi:10.1175/MWR3106.1.

World Meteorological Organization (WMO). 1970. The Beaufort scale of wind force (technical and operational aspects). 22.

Yan M, Wilson A, Dominici F, Wang Y, Al-Hamdan M, Crosson W, et al. 2021. Tropical Cyclone Exposures and Risks of Emergency Medicare Hospital Admission for Cardiorespiratory Diseases in 175 Urban United States Counties, 1999–2010. Epidemiology 32:315–326; doi:10.1097/EDE.0000000000001337.

CHAPTER 3: PROJECT 2: HEALTH IMPLICATIONS OF URBAN TREE CANOPY POLICY SCENARIOS IN DENVER AND PHOENIX: A QUANTITATIVE HEALTH IMPACT ASSESSMENT

1.Introduction

By 2050, two-thirds of the global population will live in cities (Flies et al. 2017), so improvements to urban built spaces will be crucial for public health. One key strategy is to increase urban tree canopy (UTC), or the foliage-covered proportion of a surface of interest. In addition to aesthetic and economic considerations, UTC benefits public health in ways like alleviating the urban heat island effect via shade and evaporative cooling (Zhou et al. 2021), sequestering air pollutants like PM2.5 (Nowak et al. 2018), and supporting exercise and socialization (Hunter et al. 2019), with tree-driven cooling alone significantly reducing summertime mortality (Iungman et al. 2023). Population-level analyses can capture the net effects of these benefits, along with some adverse effects like allergens or risks of accidents (Rojas-Rueda et al. 2019). Quantifying these health benefits is important for stakeholder decisions, with health impact assessments (HIAs) providing a versatile framework. HIAs estimate health implications of socially relevant changes in exposure by combining demographic and health outcome data with exposure-response functions (exposure-response functions) from epidemiologic studies to link exposures and outcomes, whether tracking outcomes caused by the exposure change, or that would have happened without a beneficial change—in that sense, making the prevented deaths "premature" in the sense of occurring later with the intervention (Harris-Roxas et al. 2012b). A recent study, for example, used an HIA framework to Philadelphia's 2025 UTC goal of 30%, projecting 403 annual averted deaths, more than half in disadvantaged areas (Kondo et al. 2020).

The aim of this study is to quantitatively predict all-cause mortality, stroke, and dementia implications of UTC policy goals in the cities of Denver, Colorado, USA and Phoenix, Arizona, USA. These cities' climates differ from each other and from Philadelphia, providing an opportunity to assess the approach's robustness and seek to better capture local variation. The cities' biophysical constraints may also preclude full policy attainment, so we consider UTC scenarios halfway between current levels and desired goals. We also used the Centers for Disease Control Social Vulnerability Index (SVI) to assess equity in current UTC and NDVI exposures within cities and potential benefits from UTC policy interventions.

We used exposure-response functions from large-scale meta-analysis to model the protective association between vegetative greenness and all-cause mortality hazard and stroke and dementia as examples of major adverse

neurological health outcomes. These models use the normalized vegetation difference index (NDVI), a standard metric for vegetation health and exposure associated with multiple health outcomes (Rojas-Rueda et al. 2019). To bridge the gap between the cities' UTC-based scenarios and the NDVI-based exposure-response functions, we developed city-specific predictive models.

2. Methods

In this study, we perform a quantitative HIA for city UTC policy goals, including official targets and partial attainment scenarios, for Denver, Colorado, and Phoenix, Arizona. We considered adult (ages 18 years old (y/o) and older) all-cause mortality and major neurological conditions stroke and dementia (ages ≥35 y/o and ≥55 y/o), along with economic impact estimates. We considered socioeconomic dimensions as measured by the Social Vulnerability Index (SVI).

2.1 Study Sites

Denver, the capital of Colorado, had a 2020 population of ~715,522 (U.S. Census Bureau QuickFacts: Denver County, Colorado 2021), and is a semi-arid high-elevation (1609 meters) city built over Western shortgrass prairie and riparian systems along the South Platte River. The urban canopy's trees are primarily deciduous broadleaves, with many non-endemic species imported from farther east in North America, or overseas (Colorado Information Marketplace 2023). Current UTC levels vary widely, with some areas well over the 20% goal and others falling considerably short (Figure S1).

Phoenix, the capital of Arizona, had an estimated 2020 population of 1,608,139 (U.S. Census Bureau QuickFacts: Phoenix city, Arizona 2021) with an arid Sonoran Desert climate and elevation of 330 meters, featuring scrub, mountains, and riparian systems. Phoenix's urban forest is primarily deciduous and evergreen broadleaves, with a small (~60%) majority of native species in a 2014 estimate (Mikulanis 2014). As with Denver, UTC levels exhibited substantial variation across the city (Figure S1).

We defined geography and populations at the census block group's geographic unit, which has roughly 600-3,000 residents (United States Census Bureau 2023). We gathered 5-year 2020 American Community Survey (ACS5) boundaries and demographic data using the *tidycensus* R package (Walker and Herman 2023). We counted populations in age groups relevant for the corresponding health outcome (18+ for all-cause mortality, 35+ for stroke, and 55+ for dementia). We used Denver's official census boundaries (excluding the non-residential area of Denver

International Airport's block group, leaving a citywide area of 301 km^2). We used Phoenix's "Planning Village" scheme, which differs from census boundaries by lacking internal cutouts and extending slightly farther north, with an area of 1,439 km^2 (City of Phoenix 2021). Maricopa County's census block groups intersect with Phoenix's boundaries, so we used within-city portions of relevant block groups, setting a lower inclusion threshold of 0.017 km^2 (20% the smallest block group's area). We adjusted population estimates using a land cover-based process described below.

2.2 UTC Policy Scenarios

We adopted Denver's standing goal of 20% UTC by 2050 (Denver Parks and Recreation 2019), as well as a "halfway" scenario of 16.5% from a 2014 13% baseline (City and County of Denver Technology Services 2014), and a sensitivity analysis to reflect falling UTC (11% in 2020) (City and County of Denver Technology Services 2020). We modeled policy scenarios as an "instantaneous" change in UTC with fixed 2020 ACS5 populations and assumed all block groups UTCs meet or exceed UTC targets. While less uniform implementations are more likely in practical terms, this idealized operationalization reflects the city's emphasis on equitable UTC access. For example, the Game Plan for a Healthy City planning document specifies that "neighborhoods with canopy under the city's goal should be priorities for tree planting" (Denver Parks and Recreation, 2019).

The standing goal for Phoenix is a 2030 target of 25% UTC (from 9% as of 2014) (Phoenix City Council 2010), with a corresponding "halfway" scenario of 17% UTC. Phoenix's 2010 Tree and Shade Master Plan is less explicit in prioritizing low-UTC regions, but more recent communications aim to "achieve tree equity" in "heat vulnerable communities" by 2030 (Urban Heat Island and Tree and Shade Subcommittee 2021).

The cities' survey methodologies differed both from each other and our exposure data sources, so for consistency we defined scenarios as percent change from the current citywide mean UTC (e.g., Phoenix's mean 25% UTC goal is approximately 2.77 times the 9% baseline).

2.3 Environmental Exposure Data

To estimate block group-level exposures, we used publicly accessible and geographically contiguous datasets. We used the 30-meter resolution, "analytical" version of the 2016 United States Forest Service Tree Canopy Cover data for UTC (Yang et al. 2018), which assigns pixels 0%-to-100% UTC, and matching National Land Cover Database (NLCD) data—a modified version of the Anderson Land Cover Classification System—for land-use

characterization (Multi-Resolution Land Characteristics Consortium 2023). NLCD includes a “developed open” category for surfaces like urban parks (<20% impervious cover, but distinct from natural vegetation or crops), and low- through high-intensity “developed" categories (reflecting 20% through >80% impervious cover). The latter stood in for population in Phoenix’s partially enclosed block groups. For example, if 50% of a block group’s total “developed” pixels were contained, we assigned 50% of the 2020 ACS5 population for health impacts. We also used NLCD data to exclude known false positives (open water, crops) from the USFS UTC data (Jin et al. 2019).

To gather data on NDVI, we downloaded a pre-processed Google Earth Engine product for monthly NDVI averages (derived from Landsat 8 Tier 1 imagery bands 4 and 5) at a spatial resolution of 30 meters (Google Developers 2022). We used the highest single month mean for each city as the exposure metric, considering NDVI levels across pixels within operational boundaries. By comparing the 12 monthly averages, we determined this “greenest month” to be June for Denver and January for Phoenix (driven largely by maintained urban vegetation, rather than the native landscape).

2.4 Spatial Analysis

The exposure-response functions defined NDVI exposure as mean NDVI across a fixed-radius (500 m for all-cause mortality, 250 m for neurological outcomes) circle from a participant’s address (Rojas-Rueda et al 2019, Paul et al 2020). We adapted this buffer radius to block groups using NLCD “developed” pixels (excluding the 'open' category, and <3-pixel areas) to represent inhabited areas and calculated mean NDVI and UTC levels in surrounding 250m and 500m buffers (Figure S2). For Denver, these areas covered roughly 248 km^2 (82.1% of Denver’s operational 301 km^2 footprint) with 500 m exposure buffers enclosing 383 km^2 (extending beyond city boundaries). For Phoenix, the developed area was 743 km^2 (51.6% of the total 1419 km^2), with 500m buffers covering 1307 km^2.

The exposure-response functions are based on NDVI, and the policy scenarios are based on UTC. To translate UTC to NDVI so that we could use the exposure-response functions, we developed city-specific UTC-to-NDVI predictive models for each buffer distance, with NDVI at a 500-meter or 250-meter buffer distance as the outcome, UTC and percentage NLCD-assessed open space were represented by natural cubic spline functions (4 degrees of freedom as chosen by Akaike Information Criterion), and a Matérn covariance function by block group longitude and latitude centroids to allow locations to influence the NDVI-to-UTC association. We adjusted for fraction of NLCD “open developed” land to isolate the specific contribution of trees to NDVI from more generalized urban greenspace contributions.

2.5 Health Outcome and Socioeconomic Data

We obtained county all-cause mortality rates 2019 from the CDC WONDER portal (National Center for Health Statistics (NCHS) 2021). While Denver City and County coincide, Maricopa County contains various levels of urbanization, so we excluded "Micropolitan" and "Non-Core" urban densities from Maricopa County to better reflect the City of Phoenix's population. We obtained state-level stroke and dementia incidence rates from the 2019 Global Burden of Disease Study from the Institute for Health Metrics and Evaluation (Institute for Health Metrics and Evaluation 2020), and estimated health outcomes by multiplying incidence rates by block group populations of "at-risk" residents, stratifying by age (1- year increments for all-cause mortality, 5-year for neurological outcomes). For socioeconomic considerations, we used the Socioeconomic Vulnerability Index (SVI), a composite of 12 census variables used by the CDC Agency for Toxic Substances and Disease Registry (ATSDR) (Agency for Toxic Substances and Disease Registry (ATSDR) 2021). SVI is provided at the census tract level, a unit comprised of block groups, so we applied estimates uniformly across constituent block groups. To compare socioeconomic groups, we created citywide SVI quartiles for exposure metrics and health outcomes, testing for differences using two-sided Wilcoxon rank tests.

2.6 Risk characterization

We used exposure-response functions to project block group health outcomes under each scenario by applying their hazard ratios as a function of NDVI exposure. We drew our all-cause mortality exposure-response function from a meta-analysis (9 studies; composite cohort of 8,324,652), with the constituent NDVI-mortality association models adjusting for, at a minimum, age, sex, smoking status, and some type of socioeconomic metric, with some of the studies further adjusting for environmental and geographic factors and additional socioeconomic metrics (Rojas-Rueda et al. 2019). The stroke and mortality exposure-response functions were from a large-scale Canadian study (cohorts of 4.3 and 1.7 million respectively). Each adjusted for age, sex, socioeconomic factors, comorbidities, population density, and several air pollutant concentrations (Paul et al. 2020). The mortality exposure-response function had a hazard ratio of 0.96 (95% Confidence Interval: 0.94, 0.97) per 0.1-unit increase in NDVI within 500m of a participant's residence. The exposure-response functions for stroke and dementia shared a 250-meter exposure buffer, with hazard ratios per 0.12-unit NDVI increase of 0.97 (95% CI: 0.96, 0.98) hazard ratio and 0.96 (0.95, 0.98) for stroke and dementia respectively. All exposure-response functions assume a linear NDVI-

mortality association. We estimated population-attributable health outcomes counts for relevant age groups using the equation:

$$PAN = \sum_{i=1}^{k}(N_i * IR_i)\frac{\left(HR^{\frac{\Delta_{NDVI}}{NDVI_{increment}}} - 1\right)}{\left(HR^{\frac{\Delta_{NDVI}}{NDVI_{increment}}}\right)} \quad (1)$$

where *AN* is the population's attributable cases, *i* is 1-*k* of either 1-year or 5-year (for all-cause mortality and neurological outcomes, respectively) segments of the relevant, N is the at-risk population (e.g., ages 35+ for stroke) from 2020 ACS5 estimates, *IR* is an age- and health outcome-specific incidence rate (from CDC WONDER for mortality, and the global burden of disease study for neurological health outcomes), *HR* is the exposure-response function-specific hazard ratio (e.g. 0.96 for all-cause mortality), Δ_{NDVI} is a change from baseline NDVI to the projected scenario value, and $NDVI_{increment}$ is the exposure-response function's unit of change in NDVI (e.g., 0.1 NDVI for all-cause mortality). To estimate economic impacts, we multiplied projected changes by the inflation-adjusted (2014 to 2019) EPA "value of mortality risk reduction" ($9,710,000) (US EPA 2014) for mortality, and Milken Institute cost of illness estimates for stroke ($5,951 direct, $6,111 indirect) and dementia ($33,084 direct, $14,662 indirect) (Waters and Graf 2018). We also compared mean and median averted deaths (absolute and per 100,000 residents), with pairwise Wilcox tests for significant differences between SVI quartiles.

Uncertainty analysis

We used a 200-iteration Monte Carlo simulation to propagate uncertainty and generate 95% uncertainty Intervals (UIs) of the 2.5th and 97.5th percentiles of resampled estimates. We resampled health outcome incidence rates and exposure-response function risk ratios with 95% confidence intervals, treating the difference between upper and central estimates as 1.96 standard deviations in normal and log-normal distributions respectively. We treated census estimates as fixed values because the provided margins of error put many smaller populations estimates below zero, reducing practical utility. Given the minimal Root Mean Square Errors for deviation between observed and predicted NDVIs (7.02 x 10^{-5} through 2.1 x 10^{-3} depending on city and buffer radius; Figure S2.1), we considered projected NDVIs based on the UTC-to-NDVI model unlikely to introduce substantial variation.

3. Results

Summarizing mean exposures at the 500-meter block group level, Denver's June 2019 mean NDVI was 0.35, and mean UTC was 13%, while Phoenix's January 2019 mean NDVI was 0.19 and UTC was 0.89% (Figure

S2.1). Both cities' block group NDVI exposures fell significantly with increasing social vulnerability (in Denver, 1st-4thquartile mean NDIVs of 0.386 through 0.316, and more dramatically in Phoenix 0.228 through 0.164); Denver's UTC followed the same trend (quartile means of 15.5% through 11%), but Phoenix saw no significant difference between detected NDVI in the 1st (0.96%) and 4th quartiles (0.93%) of SVI, with slightly higher UTC in intermediate quartiles (detected as around 1%) (Figure S4), a large gap from the city's 9% UTC assessment(Mikulanis 2014), and suggesting a weaker link with NDVI.

If Denver's 20% UTC goal were uniformly achieved "instantaneously", we predicted 200 fewer deaths annually (95% UI: 100, 300), carrying an economic impact of roughly $1.894 billion (95% UI: $0.97, $2.97 billion). We also predicted 4.1 (95% UI: 2.2, 6.7) fewer annual cases of stroke and 2.6 (95% UI: 1.5, 4.1) of dementia, with economic impacts of $49,800 (95% UI: $26,500, $81,200) and $124,000 (95% UI: $70,00, $194,000). The 16.5% UTC 'halfway' scenario would avert 128 (95% UI: 61.9, 197) deaths annually, with economic impacts of roughly $1.24 billion (95% UI: $0.60, $1.91 billion), 2.6 (95% UI: 1.4, 4.2) cases of stroke and 1.6 (95% UI: 0.9, 2.4) of dementia, with economic impacts $31,400 (95% UI: $16,500, $50,300) and $75,100 (95% UI: $42,300, $116,000). Projected health benefits are spatially heterogenous, with UTC benefits concentrated in lower-UTC and -NDVI areas like the South Platte River corridor while other areas already meet one or both policy goals (Figure 1A, 1B). Both policy scenarios saw lower all-cause mortality per 100,000 residents as social vulnerability rose, and significant pairwise differences across all but the 2nd and 3rd SVI quartiles (two-sided Wilcoxon rank test; $p<0.05$); total averted mortality had the same significance relationships, but counts were modified by population densities (e.g., the 2nd-least vulnerable quartile has the highest total averted deaths) (Figure 2.2). Neurological health outcomes closely reflected mortality trends (Figure S5).

We predicted Phoenix's 2030 25% UTC goal would avert 368 (95% UI: 181, 558) deaths annually at 2019 population levels, with an economic impact of $3.57 billion (95% UI: $1.76, $5.42 billion) in addition to 8.7 (95% UI: 4.7, 13.9) stroke and 5.1 (95% UI:2.9, 8.0) dementia cases, with economic impacts of $105,000 (95% UI: $56,600, $168,000) and $245,000 (95% UI: $140,000, $38,000). The 17% UTC "halfway" scenario saw 161 (95% UI: 79.7, 244) fewer annual deaths, with $1.57 billion (95% UI: $0.77, $2.37) in economic impacts, plus 4.2 (95% UI:2.3, 6.7) fewer cases of strokes and 2.5 (95% UI:1.4, 3.9) of dementia, with economic impacts of $50,700 (95% UI: $27,400, $80,800) and $119,000 (95% UI:$67,900, $184,000). Phoenix's projected health outcomes showed less variability than Denver's spatially (Figure 2.1C, 1D) and socioeconomically (Figure 2), with only 1 fewer death per

100,000 between the 1st and 4th quartiles to Denver's 22 per 100,000 (Figure 2). In both scenarios, only the 4th quartile significantly differed from the 1st or 3rd quartiles in averted deaths per 100,000, although differing population modified total averted deaths. As with Denver, averted neurological health outcomes per 100,000 largely mirrored mortality, but total averted stroke cases were more evenly distributed, and dementia case had effectively the opposite trend (fewer cases with increasing vulnerability) for the intermediate scenario, reflecting more 65-and-older Phoenix residents in high-SVI block groups (Figure S5).

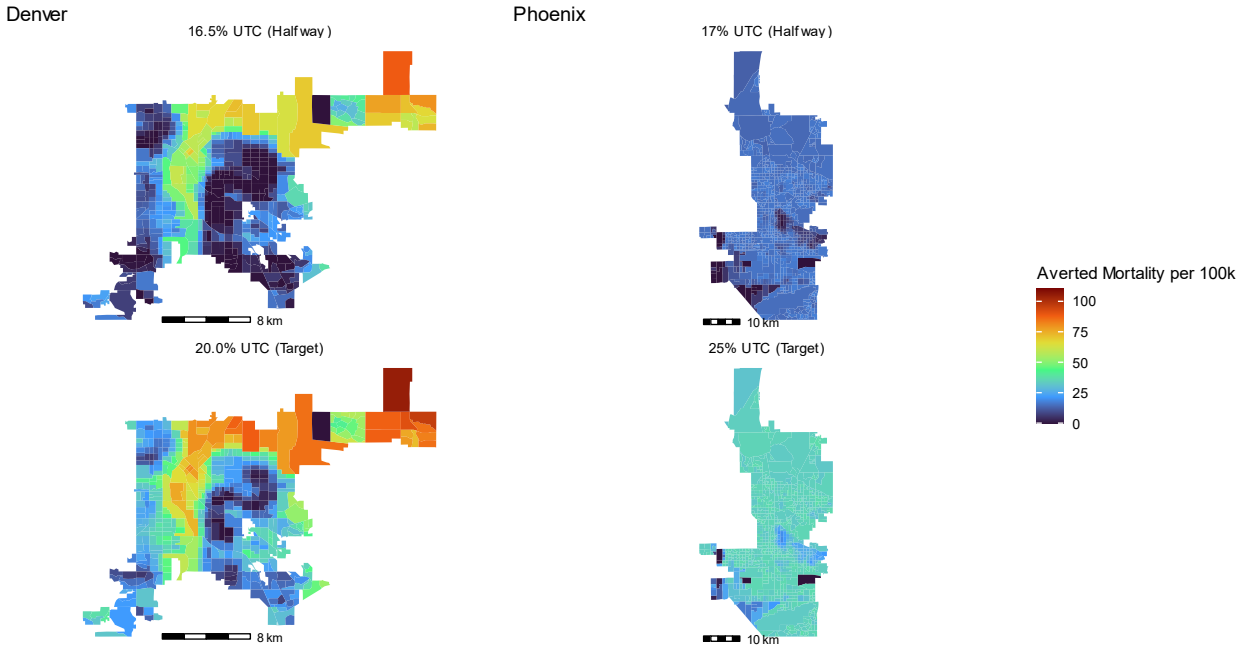

Figure 2.1: Maps of annual averted all-cause mortality (per 100k) in Denver and Phoenix by Scenario; *Maps showing Denver's projected averted all-cause mortality cases at the block group level per 100k residents halfway [16.5%] to (A) or at (C) the 2050 goal of 20% UTC, and Phoenix's 17% UTC goal (B) and 2030 25% UTC goal (D). Estimated from the projected change in NDVI, 2019 ACS5 populations (here, residents 18 and older), and CDC estimates for 2019 mortality rates in Denver County or "Urban" Maricopa County.*

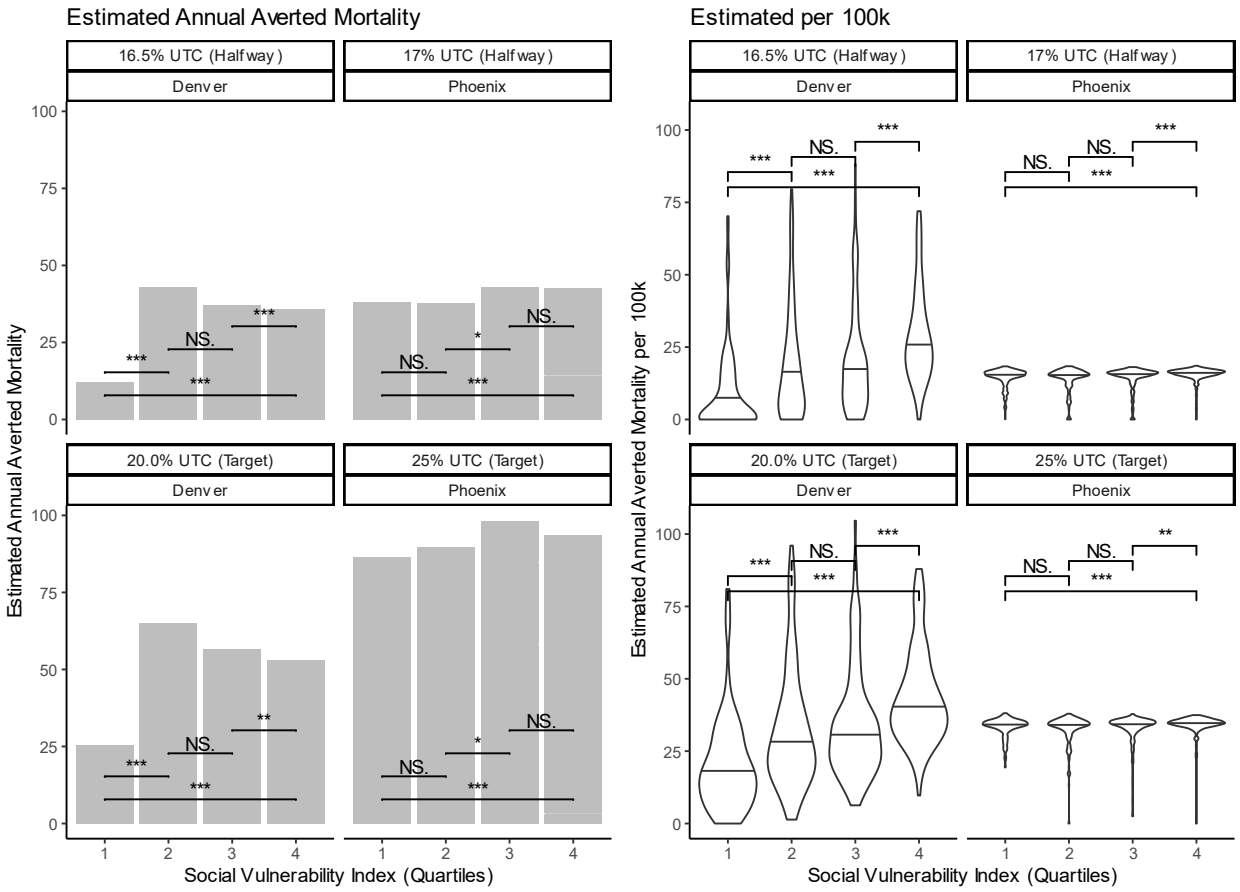

Figure 2.2: Social Vulnerability Quartile Estimated Annual Averted Attributable Mortality by City and Canopy Goal. S*ignificance levels from two-sided Wilcoxon Rank tests between indicated pairs of SVI quartiles ('NS' = p-value >0.05, '*' = p-value <0.05, >0.01], '**' = p-value <0.01, >0.001, '***' = p-value <0.001). Right: Social Vulnerability Quartile Estimated Attributable Mortality per 100k Residents by City and Canopy Goal; significance levels also from two-sided Wilcoxon Signed Rank Test with the same nomenclature.*

City Name	Denver		Phoenix	
Urban Tree Canopy (UTC) Goal	**16.5% [Halfway]**	**20% [2050 Goal]**	**17% [Halfway Goal]**	**25% [2030 Goal]**
Attributable Mortality Results				
Urban population 18+	521,000		1,110,000	
Annual Deaths Averted in at-risk (95% Uncertainty Intervals [UI])	128 (61.9, 197)	200 (99.6, 306)	161 (79.7, 244)	368 (181, 558)
Deaths Averted per 100k at-risk (95% UI)	19.4 (9.4, 29.6)	32.9 (16.2, 50.2)	14.7 (7.2, 22.1)	33.3 (16.4, 50.4)
Estimated Economic Impact (USD; 95% UI)	1.24e+09 (6.01e+08, 1.91e+09)	1.94e+09 (9.67e+08, 2.97e+09)	1.57e+09 (7.73e+08, 2.37e+09)	3.57e+09 (1.76e+09, 5.42e+09)
Stroke Results				
Population 35-84	298,000		689,000	
Annual Strokes Averted in at-risk (95% UI)	2.6 (1.4, 4.2)	4.1 (2.2, 6.7)	4.2 (2.3, 6.7)	8.7 (4.7, 13.9)
Strokes Averted per 100k at-risk (95% UI)	0.7 (0.4, 1.2)	1.2 (0.7, 2)	0.6 (0.3, 1)	1.3 (0.7, 2)
Estimated Economic Impact (USD; 95% UI)	31,600 (16,600, 50,700)	49,800 (26,500, 81,200)	50,700 (27,400, 80,800)	105,000 (56,600, 168,000)
Dementia results				
Population 55-84	104,000		252,000	
Annual Dementia Cases Averted in at-risk (95% UI)	1.6 (0.9, 2.4)	2.6 (1.5, 4.1)	2.5 (1.4, 3.9)	5.1 (2.9, 8)
Dementia Averted per 100k at-risk (95% UI)	1.4 (0.8, 2.2)	2.4 (1.3, 3.7)	1 (0.6, 1.5)	2 (1.2, 3.1)
Estimated Economic Impact (USD; 95% UI)	75,200 (42,400, 117,000)	124,000 (70,000, 194,000)	119,000 (67,900, 184,000)	245,000 (140,000, 380,000)

Table 2.1: Health Impact Assessment Summary for UTC Policy Scenarios. *Quantitative summary of health impact projections; central estimates are provided with 95% uncertainty intervals (UIs) calculated via 200 bootstrapping resampling iterations as described in methods. Values are rounded to three significant digits, truncating at a single decimal place. Table sections are divided by health outcome.*

4. Discussion

Our findings suggest that UTC interventions can have public health benefits for arid cities, as mediated by mechanisms like shade-based cooling and air purification (Rojas-Rueda et al. 2019). Our models projected Denver's 20% UTC goal would, if uniformly attained, avert about 200 deaths, 4.1 stroke, and 2.6 dementia cases annually, with a combined economic impact of $1.94 billion, with about 64% of these benefits captured by half-way attainment (16.5% UTC). Phoenix's 25% UTC target for 2030 would avert roughly 368 annual deaths, 8.7 stroke cases, and 5.1 dementia cases, with a combined economic impact of $3.57 billion, and the 17% "halfway" scenario seeing ~44% of these benefits. More vulnerable communities (by SVI quartile) generally saw greater proportional health benefits from UTC intervention. We modeled each city's current canopy goals as proportional change from nominal citywide means in USFS UTC data to standardize across methodologies.

Our approach has several strengths. Utilizing publicly accessible and geographically standardized data makes it readily transferrable to other US cities, or internationally with minor modifications. By adapting our exposure-response functions' operational definitions of NDVI exposure with population-weighted buffers, our exposure estimates accommodate large and unevenly inhabited block groups, and multiple outcome-specific buffer sizes. Our city-specific predictive UTC-to-NDVI models strengthen transferability by capturing nonlinear trends with splines and reflecting spatial variation in the association within cities. Overfitting would be a concern for exploratory purposes or wider generalization, but is not an issue in this purely predictive, city-specific application. Adjusting for developed open space in the model accounts for varying contributions by vegetation type to mean NDVIs (Martinez and Labib 2023), and potential confounding with trees acting largely as "indicators" of generally well-watered landscapes like golf courses. Using a "greenest month" rather than a single representative day (Kondo et al. 2020) or uniform sampling window between cities (Brochu et al. 2022), helps reflect seasonal variation and provide a locally comparable level of greenness across cities.

Several noteworthy assumptions about our scenarios and health outcome estimates include the assumption of fixed 2020 ACS5 population levels, but we could expect growth and other population changes over the true period of interest, affecting estimated totals. Newly planted trees also need years to reach full efficacy, underscoring the importance of preserving urban forests. While the exposure-response functions adjusted for available demographic, socioeconomic, and environmental factors, residual confounding from positive associations between SES (e.g. quality of healthcare) and greenspace still likely bias apparent NDVI benefits upward (Nardone et al.

2021). Residence-based NDVI exposure metrics also omit factors like personal travel and legal/logistical greenspace accessibility, introducing (likely non-differential) bias (Ekkel and de Vries 2017). We also assume the exposure-response functions are transportable to our specific context of arid or semi-arid US cities despite being drawn from other populations; these also assume a linear association between NDVI and health outcomes. Modeling more precise relationships could be an important avenue for further research, especially in equity and determining priorities in UTC-raising and other greening initiatives. We also assumed USFS UTC estimates were essentially proportional to true canopy cover.

Another assumption is that UTCs are uniformly implemented across all block groups, with a preference for developed areas, where both cities' UTC goals could and realistically be achieved via less-uniform UTC coverage. This represents an idealization of equity goals, but while Denver's Game Plan for a Healthy City explicitly prioritizes low-UTC neighborhoods (Denver Parks and Recreation 2019) and recent official City of Phoenix communications seek "tree equity" for under-served communities (Urban Heat Island and Tree and Shade Subcommittee 2021), Phoenix's Tree and Shade Master Plan lacks this language, even suggesting "tree banking" where areas with poor canopy suitability subsidize planting elsewhere (Phoenix City Council 2010).

Our study also inherits dataset limitations. Notably, aggregating state-level incidence rates obscures within-city variability, especially for the state-level neurological health outcomes; this warrants caution, but could be considered a form of non-directional Berkson's Error (Armstrong 1998). Likewise, EPA economic impacts are nationally aggregated across dimensions like age and income, reducing precision for specific populations (Hammitt and Robinson, 2011). While NDVI is positively correlated in adjacent years, chronological gaps can distort exposure-outcome associations (Helbich 2019). NLCD's overall 2016 thematic accuracy is 90.6±0.6% for the level II category used here (Wickham et al. 2021). Still, the contrast between Phoenix's MLRC-estimated UTC of ~1% and higher 9% city canopy assessment raises concerns about accuracy in this application. Another consideration is that NLCD's "developed" land category includes both residential and non-residential land use (including larger roadways), potentially leading to exposure misclassification in our use case. In single-city analysis, municipal zoning data would provide greater precision, and other approaches may be useful in certain contexts--e.g., Philadelphia's smaller, more uniform block groups reasonably approximated 500-meter buffers in Kondo et al. (2020), or gridded population data like the 2015 Global Human Settlement Layer (as in Iungman et al., 2023), removes the dependence on political boundaries and the potential misclassification of non-residential land use.

For implementation, UTC benefits vary with factors like tree species and age, ambient humidity, and species selection (Rahman et al. 2020). For instance, shade is more effective than evapotranspirative cooling in climates, favoring short, broadleaved trees (Rahman et al. 2020). Upkeep costs could be a barrier in more vulnerable communities if funding is left to residents. Beyond equity concerns, lower-UTC areas may be resource-efficient priorities because UTC-driven cooling sees diminishing returns in already well-forested neighborhoods (Zhou et al. 2021). Both regions' native trees, despite generally lower NDVI, are more drought tolerant, lower upkeep, and biodiversity benefits (Garber et al. 2023). Careful tree selection and ongoing maintenance can mitigate UTC-associated hazards like allergens and physical obstructions (Rojas-Rueda et al. 2019). While these considerations are critical for effective implementation, our projections suggest UTC growth could offer health benefits for both cities, especially in more vulnerable communities.

5. Conclusion

This study supports UTC Denver and Phoenix's policy goals to improve public health and prioritize spatial and socioeconomic equity in their implementation for optimal effectiveness, with more vulnerable communities generally seeing lower initial greenspace exposure and greater proportional modeling benefits under UTC intervention. These results act as general guideposts, with more granular research and information needed to assess the practicality of these targets (especially in Phoenix) and the optimal varieties of trees and vegetation for each city. We also tested a scalable process for performing similar analyses and modified prior research to improve relevance for arid conditions.

References

Agency for Toxic Substances and Disease Registry (ATSDR). 2021. CDC/ATSDR's Social Vulnerability Index (SVI). Available: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html [accessed 30 September 2021].

Armstrong BG. 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational and Environmental Medicine 55:651–656; doi:10.1136/oem.55.10.651.

Brochu P, Jimenez MP, James P, Kinney PL, Lane K. 2022. Benefits of Increasing Greenness on All-Cause Mortality in the Largest Metropolitan Areas of the United States Within the Past Two Decades. Front Public Health 10:841936; doi:10.3389/fpubh.2022.841936.

City and County of Denver Technology Services. 2014. Denver Open Data Catalog: Tree Canopy (2014). Available: https://denvergov.org/opendata/dataset/tree-canopy-2014 [accessed 30 September 2023].

City and County of Denver Technology Services. 2020. Denver Open Data Catalog: Tree Canopy (2020). Available: https://denvergov.org/opendata/dataset/city-and-county-of-denver-tree-canopy-2020 [accessed 30 September 2023].

City of Phoenix. 2021. Phoenix Village Planning Committees. Village Planning Committees. Available: https://www.phoenix.gov/villages [accessed 22 April 2023].

Colorado Information Marketplace. 2023. Tree Inventory of Denver. Available: https://data.colorado.gov/Environment/Tree-Inventory-of-Denver-Map/hzmx-2dfk [accessed 13 September 2023].

Denver Parks and Recreation. 2019. Game Plan for a Healthy City.

Ekkel ED, de Vries S. 2017. Nearby green space and human health: Evaluating accessibility metrics. Landscape and Urban Planning 157:214–220; doi:10.1016/j.landurbplan.2016.06.008.

Flies EJ, Skelly C, Negi SS, Prabhakaran P, Liu Q, Liu K, et al. 2017. Biodiverse green spaces: a prescription for global urban health. Front Ecol Environ 15:510–516; doi:10.1002/fee.1630.

Garber MD, Guidi M, Bousselot J, Benmarhnia T, Dean D, Rojas-Rueda D. 2023. Impact of native-plants policy scenarios on premature mortality in Denver: A quantitative health impact assessment. Environment International 178:108050; doi:10.1016/j.envint.2023.108050.

Google Developers. 2022. Landsat 8 Collection 1 Tier 1 32-Day NDVI Composite. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_32DAY_NDVI [accessed 15 August 2022].

Harris-Roxas B, Viliani F, Bond A, Cave B, Divall M, Furu P, et al. 2012. Health impact assessment: the state of the art. Impact Assessment and Project Appraisal 30:43–52; doi:10.1080/14615517.2012.666035.

Helbich M. 2019. Spatiotemporal Contextual Uncertainties in Green Space Exposure Measures: Exploring a Time Series of the Normalized Difference Vegetation Indices. IJERPH 16:852; doi:10.3390/ijerph16050852.

Hunter RF, Cleland C, Cleary A, Droomers M, Wheeler BW, Sinnett D, et al. 2019. Environmental, health, wellbeing, social and equity effects of urban green space interventions: A meta-narrative evidence synthesis. Environment International 130:104923; doi:10.1016/j.envint.2019.104923.

Institute for Health Metrics and Evaluation. 2020. Global Burden of Disease Study 2019 (GBD 2019) Results.

Iungman T, Cirach M, Marando F, Pereira Barboza E, Khomenko S, Masselot P, et al. 2023. Cooling cities through urban green infrastructure: a health impact assessment of European cities. The Lancet 401:577–589; doi:10.1016/S0140-6736(22)02585-5.

Jin S, Homer C, Yang L, Danielson P, Dewitz J, Li C, et al. 2019. Overall Methodology Design for the United States National Land Cover Database 2016 Products. Remote Sensing 11:2971; doi:10.3390/rs11242971.

Kondo MC, Mueller N, Locke DH, Roman LA, Rojas-Rueda D, Schinasi LH, et al. 2020. Health impact assessment of Philadelphia's 2025 tree canopy cover goals. The Lancet Planetary Health 4:e149–e157; doi:10.1016/S2542-5196(20)30058-9.

Martinez A de la I, Labib SM. 2023. Demystifying normalized difference vegetation index (NDVI) for greenness exposure assessments and policy interventions in urban greening. Environmental Research 220:115155; doi:10.1016/j.envres.2022.115155.

Mikulanis V. 2014. Phoenix, Arizona Project Area-Community Forest Assessment.

Multi-Resolution Land Characteristics Consortium. 2023. National Land Cover Database Class Legend and Description. Available: https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description [accessed 12 September 2023].

Nardone A, Rudolph KE, Morello-Frosch R, Casey JA. 2021. Redlines and Greenspace: The Relationship between Historical Redlining and 2010 Greenspace across the United States. Environ Health Perspect 129:017006; doi:10.1289/EHP7495.

National Center for Health Statistics (NCHS). 2021. Underlying Cause of Death, 1999-2020 Request. Underlying Causes of Death 1999-2020 on CDC WONDER Online Database, released 2021. Available: https://wonder.cdc.gov/Deaths-by-Underlying-Cause.html [accessed 8 March 2021].

Nowak DJ, Hirabayashi S, Doyle M, McGovern M, Pasher J. 2018. Air pollution removal by urban forests in Canada and its effect on air quality and human health. Urban Forestry & Urban Greening 29:40–48; doi:10.1016/j.ufug.2017.10.019.

Phoenix City Council. 2010. City of Phoenix Tree and Shade Master Plan.

Rahman MA, Stratopoulos LMF, Moser-Reischl A, Zölch T, Häberle K-H, Rötzer T, et al. 2020. Traits of trees for cooling urban heat islands: A meta-analysis. Building and Environment 170:106606; doi:10.1016/j.buildenv.2019.106606.

Rojas-Rueda D, Nieuwenhuijsen MJ, Gascon M, Perez-Leon D, Mudu P. 2019. Green spaces and mortality: a systematic review and meta-analysis of cohort studies. The Lancet Planetary Health 3:e469–e477; doi:10.1016/S2542-5196(19)30215-3.

United States Census Bureau. 2023. Glossary. Census.gov. Available: https://www.census.gov/programs-surveys/geography/about/glossary.html [accessed 24 September 2023].

Urban Heat Island and Tree and Shade Subcommittee. 2021. UHITS Heat Equity Memorandum. Available: https://www.phoenix.gov/oepsite/Documents/UHITS%20Heat%20Equity%20Memorandum%2012.8.2021.pdf [accessed 24 September 2023].

U.S. Census Bureau QuickFacts: Denver County, Colorado. 2021. Available: https://www.census.gov/quickfacts/denvercountycolorado [accessed 4 October 2021].

U.S. Census Bureau QuickFacts: Phoenix city, Arizona. 2021. Available: https://www.census.gov/quickfacts/fact/table/phoenixcityarizona/PST045219 [accessed 4 October 2021].

US EPA O. 2014. Mortality Risk Valuation. Available: https://www.epa.gov/environmental-economics/mortality-risk-valuation [accessed 14 October 2022].

Walker K, Herman M. 2023. *tidycensus: Load US Census Boundary and Attribute Data as "tidyverse" and 'sf'-Ready Data Frames*.

Waters H, Graf M. 2018. The Costs of Chronic Disease in the U.S. Milken Institute.

Wickham J, Stehman SV, Sorenson DG, Gass L, Dewitz JA. 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sensing of Environment 257:112357; doi:10.1016/j.rse.2021.112357.

Yang L, Jin S, Danielson P, Homer C, Gass L, Bender SM, et al. 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing 146:108–123; doi:10.1016/j.isprsjprs.2018.09.006.

Zhou W, Huang G, Pickett STA, Wang J, Cadenasso ML, McPhearson T, et al. 2021. Urban tree canopy has greater cooling effects in socially vulnerable communities in the US. One Earth 4:1764–1775; doi:10.1016/j.oneear.2021.11.010.

CHAPTER 4: PROJECT 3: PROJECTING MID-21ST CENTURY EXCESS MORTALITY IMPACTS FROM TEMPERATURE EXTREMES UNDER A SOLAR CLIMATE INTERVENTION SCENARIO.

1. Introduction

There is an overwhelming consensus among researchers that human activity, including the emission of greenhouse gasses which absorb outgoing solar radiation, is changing the climate (Intergovernmental Panel On Climate Change 2023). While there have been mitigation efforts to an often limited and localized extent, researchers project the planet to keep warming without active intervention to remove greenhouse gases or otherwise offset warming (Mechler et al. 2020). While the average global temperature is rising, there is considerable heterogeneity in local effects, including rates of increased temperature extremes, with corresponding increases in heat-associated, and potentially more broadly temperature-associated, excess mortality (Gasparrini et al. 2017a). Because projections for even aggressive greenhouse gas reductions would be insufficient to avoid ongoing warming, organizations like the National Academy of Sciences, Engineering, and Medicine have called for research into the potential risks and benefits of climate intervention measures (NASEM 2021). Where mitigation efforts seek to slow emissions of greenhouse gases, climate interventions actively modify earth systems with the intent of directly counteracting effects of anthropogenic warming (Crimmins et al. 2016).

One prominent proposed climate intervention method is Stratospheric Aerosol Injection (SAI), which achieves a cooling effect by introducing reflective aerosols (commonly sulfur dioxide, or SO_2) to the stratosphere to scatter and reflect incoming solar radiation (Fasullo and Richter 2023). SAI as a climate intervention proposal builds on decades of observed cooling from inadvertent near surface-level anthropogenic sulfate pollution; e.g. stricter air quality regulations on sulfate pollution from fossil fuel production and use led these energy sources to have a stronger net warming effect in terms of radiative forcing (Jiang et al. 2022). SAI is seen as a strong candidate for SAI for reasons including this well-supported cooling mechanism (Richter et al. 2022), technological feasibility, and relatively low cost on the order of several billion USD annually (Tracy et al. 2022), although costs for more extensive, multi-aerosol scenarios, have been estimated up to ~$18 billion annually (Smith 2020).

With SAI being an active area of inquiry, a number of explorations into potential health and wider human society. SAI also carries wider ecological and sociological implications, overall and varying by specific deployment models, with ongoing research into earth system mechanisms including changes in exposure to surface-level sulfates and ozone depletion as a result of SO_2 interactions (Tracy et al. 2022). One relatively underexplored facet of the

total health impacts of SAI is the direct influence of temperature extremes (both heat and cold) on human health. In this study, we present a health impact assessment case study of modeled SAI effects on temperature-related mortality in major urban areas in the United States of America, using the Assessing Responses and Impacts of Solar climate intervention on the Earth system with Stratospheric Aerosol Injection (ARISE-SAI), which models a 2035-2070 SAI intervention, with SO_2 injection increasing year-over-year and directed by an algorithm intended to minimize disturbance to climate systems like north-south temperature gradients, which influence features like atmospheric currents and precipitation patterns (Richter et al. 2022).

ARISE-SAI has been recently developed to reflect our current understanding of factors like technological feasibility and emissions trends (Richter et al. 2022). A complicating factor in modeling health costs and benefits of climate interventions like SAI is that the relevant point of comparison is climate change without the intervention, rather than present-day conditions. This means that both the SAI and climate change "control" scenarios represent distinct changes from the present day and carry their own distinct health impacts and model uncertainty. To this end, ARISE-SAI is built on the widely-used 2nd iteration of the Community Earth System Model (CESM2) (Danabasoglu et al. 2020), and Shared Socioeconomic Pathways (SSP) framework for formalizing scenarios of assumed population growth, technological changes, renewable energy investment, international stability, etc.—specifically SSP2-4.5, a "middle-of-the-road" scenario with fairly conservative assumptions, although including fairly aggressive mitigation efforts (Riahi et al. 2017). These design choices make ARISE-SAI compatible with all SSP2-4.5 features except for the SAI intervention itself, as well as with population and mortality rate estimates under SSP2-4.5, which we use to model demographic components of the scenario (Hauer 2019; Best et al. 2018).

Globally, under scenarios of climate change, an increased number of heat-related deaths is expected to at least slightly outnumber reductions in cold-related deaths, although the effects are projected to be highly variable, with generally less pronounced impacts in temperate regions (Gasparrini et al. 2017a). Temperature-associated mortality is a useful metric in this application, as it is causally "downstream" of a wide range of specific health burdens and mechanisms, some better understood than others. Most directly, extreme temperatures can lead to acute stress and mortality. For instance, perspiration-based cooling no longer functions above a wet bulb air temperature of 35°C, with adverse effects beginning below this threshold even among young, healthy adults (Vecellio et al. 2022). At the other extreme, as external conditions reduce core body temperature, stage 1 hypothermia can set in

around 35°C, and core body temperatures of ~24°C represents an effective lower limit for survival (Rathjen et al. 2019).

While these extremes are perhaps most emblematic of temperature-associated health burdens, deleterious effects can begin with even small departures from "optimal" conditions. For instance, temperature appears to play a regulatory role in immune system function in laboratory mouse studies, with a 5°C colder environment (26 °C, compared to a 31 °C control) associated with an apparent immunosuppressive effect allowing increased tumor formation rates (Evans et al. 2015). Population-level observational studies of humans have suggested analogous temperature-associated cancer incidence trends (Bandyopadhayaya et al. 2020). Higher-temperature conditions can also dampen immune responses, with for instance 36°C conditions associated with reduced adaptive immunity to influenza as a consequence of weakened antibody production and white blood cell activity (Moriyama et al. 2020). Age is another modulating factor for the health impacts of temperatures; for instance, older adults exhibited much higher blood pressure when responding to temperature changes, a change thought to be related to arterial stiffening, which could provide an explanation for increased risk of health outcomes like cardiovascular diseases in cold conditions (Hess et al. 2009). Aging is also a risk factor for conditions like heart failure and diabetes under both hot and cold conditions relative to body temperature (Greaney et al. 2016).

While there are physiological limits to human health with respect to body temperatures, urban populations exhibit dramatic adaptation to local conditions, with a range of ~20°C (roughly 12-32°C) in local "minimum mortality temperatures" (MMTs), or the temperature at which there is the lowest risk of death (Yin et al. 2019). In the context of nonlinear models of temperature-mortality associations, the MMT can be operationally defined as an inflection point where mortality risk reaches its lowest point (Armstrong 2006). The MMT has been shown to be strongly associated with prevalent temperatures in the local climate (Yin et al. 2019). An important factor in population-level adaptations is that a fairly narrow range of preferred indoor temperatures (roughly 22-25°C) are followed regardless of a population's cultural and climatological context, but many strategies are used to regulate indoor temperatures and outdoor body temperatures, including local architecture, active temperature regulation methods like air conditioning (Jay et al. 2021). At a larger scale, cities can implement strategies like promoting urban tree canopies to provide shade and evaporative cooling, which is especially important in urban contexts where the urban heat island

effect—heightened warming as a result of high concentrations of impervious heat-absorbing surfaces—can raise local temperatures appreciably to the detriment of public health (Iungman et al. 2023).

Adaptive measures can also be behavioral in nature, including clothing selection for thermal regulation (Salata et al. 2018), later sleep schedules to avoid metabolically stressful high temperatures (Hendel et al. 2017), and modified or reduced physical activity outside of 'neutral' temperature conditions (Sharifi and Boland 2018). These adaptations can be quite culturally specific—e.g., in the cold-climate city of Harbin, China, a qualitative study found locals generally dressed less warmly than tourists, but were less active in cold early morning hours (Xi et al. 2020). While there is a trend for adaptation to local conditions, it is not strictly geographically determinative; for instance, Austria and Germany have lower excess cold-related mortality rates than their immediate eastern neighbors, even at the scale of neighboring border cities (Masselot et al. 2023). Adaptive measures like these may also help address non-physiological mechanisms to temperature-related health burdens. For instance, the fluid dynamics of cold air are more conducive to the dispersal of some airborne viral particles (Dbouk and Drikakis 2021). In a holistic sense, health risks are also intertwined with seasonal illnesses and associated social and built environment factors like indoor congregation, working schedules, and indoor air recirculation (Moriyama et al. 2020). Temperature adaptations appear relatively dynamic, with for instance a study of heat-mortality associations in major urban centers in the United States estimating an approximate halving of average excess mortality rates between 1987 and 2005 (without a significant correlation to air conditioner adoption rate), if with heterogeneity like more pronounced reductions among 75-and-older adults and in northern cities (Bobb et al. 2014).

Given the complexity of temperature-mortality associations, and the range of specific mechanisms, it is useful to use models allowing for lag time between the immediate exposure and corresponding health outcomes as well as capable of accounting for long-term trends and seasonal effects in "baseline" mortality rates. To this end, we adapt the widely used Distributed NonLinear Model (DLNM) approach (Gasparrini 2011; Vicedo-Cabrera et al. 2019) to capture interactions between temperature exposures and up to 21 days of lag time since exposure, as well as a restricted spline basis to allow nonlinear temperature-mortality associations while keeping relatively conservative extrapolations beyond observed temperatures. We adjust for long-term and seasonal baseline mortality trends (Gasparrini et al. 2015), but still operationally rest our analysis on an assumed causality—i.e. that temperature exposures mechanically *cause* the associated excess mortality, at the possible expense of complexities like diseases showing seasonal trends but not necessarily direct temperature-driven mechanisms.

Our objective in this analysis is to perform a health impact assessment using temperature-linked all-cause mortality (i.e. excess deaths during and after unusually hot or cold temperatures) among 65-and older adults (with subgroups for ages 65-74 and 75+) in eight major United States metropolitan areas (nominally cities, but represented at the nearest county level) in eight major United States metropolitan regions comparing ARISE-SAI against a corresponding "control" climate change scenario (SSP2-4.5) without the SAI intervention. We will use historical (1987-2005) data on daily temperatures and mortality counts in the counties of interests, meaning that the projections may not capture population-level adaptations to temperature in the intervening decades. To project mid-21st century mortality rates and counts, we use scenario-compatible projected 2050-2060 populations and age compositions developed using demographic projections under SSP2-4.5 (Hauer and Center For International Earth Science Information Network 2020), as well as general census-based mortality rate projections. We hypothesize that SAI's cooling influence will reduce heat-attributable mortality in the studied cities, or at a minimum total temperature-associated mortality will not increase in any given city.

2. Methods

2.1 Overview of Strategy for Health Impact Analysis

In this investigation, we used a health impact assessment framework incorporating multiple sources of data, as well as the associated uncertainties. First, we used historical data of temperature, and of daily all-cause mortality, from communities of interest to generate functions describing the association between this exposure and outcome pairing. Because the temperature-mortality relationship is time-dependent, with time elapsed since exposure changing the relative mortality risk, we used a nonlinear distributed lag approach, with a nonlinear basis to account for complexities. Researchers can calculate uncertainty in these multipart models by generating a large pool of estimates and taking a 95% empirical confidence interval of the resulting values. Having obtained a function describing the historical association between temperature and mortality, we could then apply it to future scenarios by substituting in both projected future temperatures and projected future population and demographic data to translate historical trends into corresponding future temperature-attributable mortality estimates (under the assumption of a fixed trend).

2.2 Study System

Our study focuses on eight major United States cities (Figure 1), selected as high population centers that span a range of geographic regions and local climates. These are (in order of latitude) Seattle, WA, Chicago, IL, New York, NY, Philadelphia, PA, Los Angeles, CA, Phoenix, AZ, Houston, TX, and Miami, FL. We based our operational definitions on the 1987-2005 National Mortality, Morbidity, and Air Pollutions Study (NMMAPS) Part II study in defining the cities at the closest county-level resolution (Health Effects Institute 2000). In the NMMAPS dataset, nominal cities are represented to the nearest county level—e.g. as an extreme example, "Phoenix" is operationally Maricopa county, representing a near-quadrupling of effective population (U.S. Census Bureau QuickFacts: Phoenix city, Arizona 2021), with most cases being a closer approximation, with some city boundaries exactly coinciding with their constituent counties (Health Effects Institute 2000). While no longer publicly available, this widely used dataset still provides a useful basis by providing daily tracking of mortality data and methodological standardization across cities, and for its use in a wide range of research into mortality-temperature associations (Barnett et al. 2012).

From the NMMAPS data, we assessed daily mortality counts by age group, and exposure of daily air temperature (represented as the mean of daily minimum and maximum temperatures following the metric used in historical NMMAPS data). We focused on older adults given their greater vulnerability to temperature-associated mortality, using two age groups (65-74 and 75+), intentionally coinciding with NMMAPS categories. However, at this stage of the analysis we focused on the older two age groups, which tend to represent a consistently larger and more homogenous (concerning temperature risk trends) than the youngest cohort.

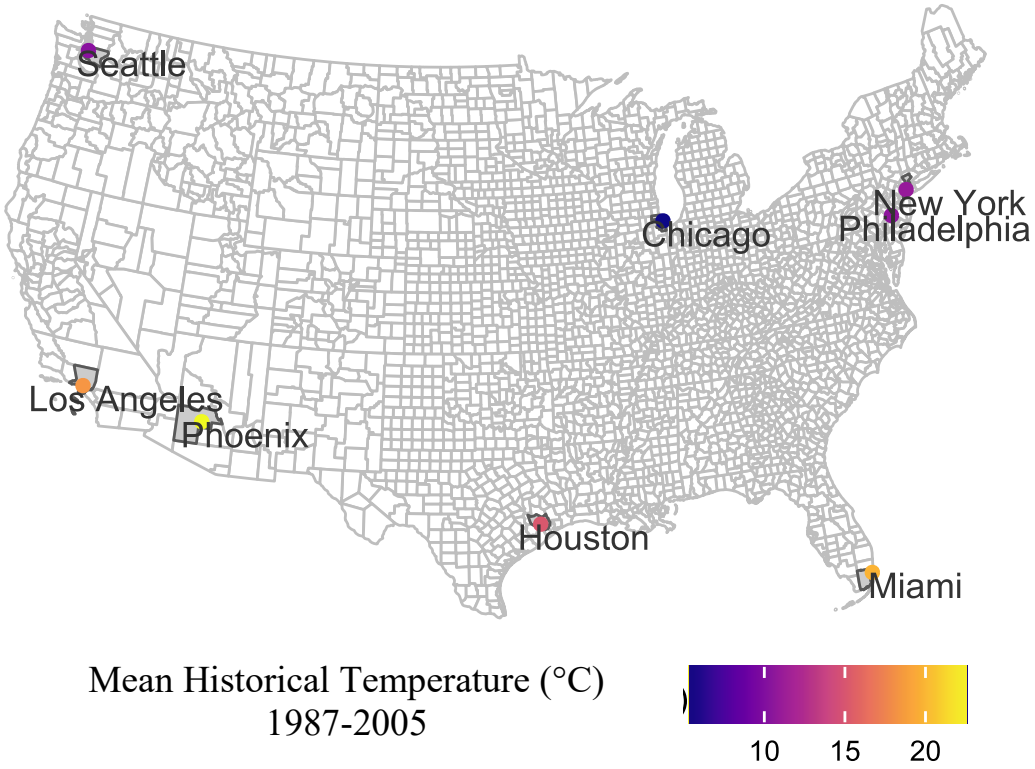

Figure 3.1: Map of (nominal) cities in historical NMMAPS dataset, with county boundaries and points indicating city coordinates, as well as mean temperature over the 1987-2015 observation period indicated on the color scale.

2.3 *Population and Mortality Projections*

Because our analysis focuses specifically on a mid-21st century time, we also wanted to account for projected changes in age demographics and overall populations. To this end, we used population projections from the National Aeronautics and Space Administration's Socioeconomic Data and Applications Center (SEDAC), which estimates county-level populations in 5-year age and temporal increments (i.e., population estimates for ages 0-4, 5-9, etc., in years 2020, 2025, etc.) from 2020 through 2060 (with mortality projections being the limiting step; others extend to 2100). The SEDAC data's county-level population estimates reflect assumptions from the Shared Socioeconomic Pathways (SSP) framework originally codified by the sixth Intergovernmental Panel on Climate Change (Riahi et al. 2017). The SSP framework presents a standardized set of scenarios reflecting a range of paths for socioeconomic factors like demographic growth, energy usage and sources, technological innovations, etc., and their consequent effects on the climate. For this analysis, we chose the SSP2.4-5 scenario, representing a "middle of the road" scenario (relatively conservative assumptions on population growth, technological changes, etc., but notably including estimates on fairly extensive climate change mitigation efforts). In addition to providing a relatively conservative model for societal and climate trends in the coming century, this pathway matches the one chosen for the ARISE-SAI

model's assumptions, providing scenario coherence between projected exposure and population data. While SEDAC estimates are available for a longer period, we focus specifically on the decade of the 2050s for this analysis. In addition to shifting populations, we also wanted to account for anticipated trends in all-cause mortality rates, using United States Census projections for this information (US Census Bureau 2017). We provide this data in single-year increments (chronological and by age) for 2020-2060 nationally.

2.4 Temperature Exposure Projections

For exposure data under both our baseline and SAI intervention scenarios, we used modeled temperatures from the Community Earth System version 2 (CESM2) during the years 2050-2060 (Danabasoglu et al. 2020). The SSP2-4.5 scenario represented our climate change baseline scenario under "middle of the road" climate change conditions, and the Assessing Responses and Impacts of Solar climate intervention on the Earth system with Stratospheric Aerosol Injection (ARISE-SAI) scenario (Richter et al. 2022) representing alternate conditions under the SAI intervention starting in 2035. CESM2 uses an ensemble approach, where the model uses slight variations in initial temperatures in otherwise identical iterations of (ensemble members) to approximate the propagation of uncertainty through climate systems due to stochasticity and boundary conditions. We included ten ensemble members in our analysis, reporting temperatures in terms of distributions among the ten ensembles for each city. To match the metric used in our historical data, we took the mean of mean maxima and minima of daily projected 2-meter air temperatures at a spatial resolution of 1.25 and 0.94241 degrees for latitude and longitude, respectively. NMMAPS city exposures were represented as the grid cell closest to the nominal city's geographic center, or the most-similar adjacent grid cell if the nearest was more than half covered by water, which could result in non-representative temperature ranges and trends, as water is generally more thermally absorbent and stable than land under similar atmospheric temperatures. In our dataset, we made these substitutions for Chicago, where more than half of the closest grid cell was covered by Lake Michigan, and Miami, which was mostly the Atlantic Ocean. Looking for <50% water regions with the most similar climate to the historical data, we picked the area due east of Chicago, and west of Miami, to provide comparable land-based temperature estimates. Projected annual average temperatures, including the 10 ensemble members and model median for average daily temperatures, are shown in Figure 3.2.

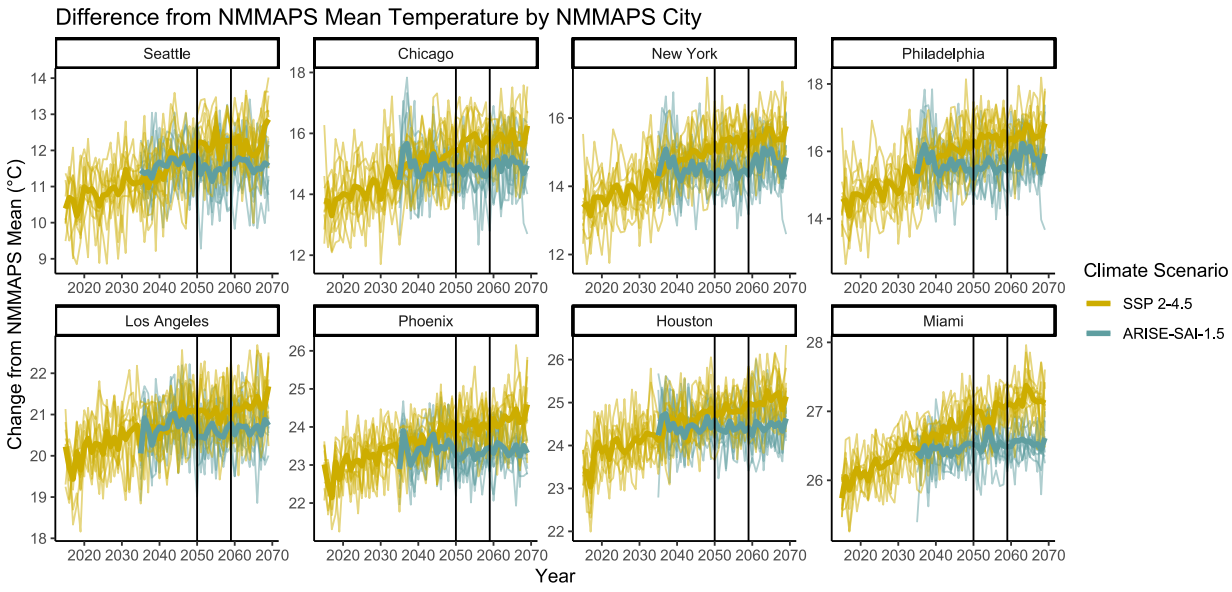

Figure 3.2: CSEM2 Models Projections of Temperatures (averaging across daily averages to annual averages for demonstration purposes) under the unmitigated warming (SSP2-4.5) and ARISE-SAI scenarios (coded gold and blue, respectively). Thin lines indicate ensemble members' contributions to the model, and the heavy line indicates the mean temperature across ensemble members. This shows the entire modeled period (2015-2070, with SAI beginning in 2035), but the vertical lines indicate the 2050-2060 period of interest.

2.3 Future Population and Mortality Count Estimates

We needed to take several steps to harmonize the temporal resolutions and formats of our input data. First, we used natural cubic splines to interpolate SEDAC estimates (five-year chronological intervals) into single-year intervals for better compatibility with the finer temporal resolution of the other datasets, defining age groups in single-year groups. The Census-based mortality needed no adjustments, already being in single-year intervals, and five-year age groups. For each single-year age group, we then multiplied estimated SEDAC populations by the corresponding age-adjusted census-based mortality rates for projected annual mortality counts between 2050 and 2060 under SSP2-4.5. We took sums of both groups of single-year estimates to obtain total population and population-weighted mortality rate estimates for the NMMAPS age groups of 65-74 and 75+.

2.5 Distributed Lag Nonlinear Model, Risk Ratio Estimation, and Minimum Mortality Temperatures

To generate city-specific exposure-response functions based on historical temperature-mortality associations, we adapted a widely used workflow for projecting public health impacts of climate change-related exposures (Vicedo-Cabrera et al. 2019). This method uses distributed lag non-linear models (DLNMs), which account for trends in baseline mortality both on short-term day-to-day and long-term seasonal scales. This approach

accounts not only for mortality on the same day as a temperature exposure, but allows for "lag time", or a method of modeling comprehensive time-dependent risks over a period after the initial exposure. This is important because tracking mortality trends over (in this case) 21 days allows the model to account for the gradual development of for instance cold-related, temperature-associated health burdens without losing sensitivity to typically more immediate heat-related mortality (Vicedo-Cabrera et al. 2019).

In the distributed lag framework, we used a modified penalized spline function basis (Gasparrini et al. 2017b) to accommodate non-linear trends, with initial knots set at the 1st, 30th, 60th, 90th, and 99th percentiles (although these are modified by a penalty function to provide a better fit) to balance the ability to extrapolate beyond observed 1987-2005 temperatures with reasonable uncertainty levels. To generate these functions, we used historical temperatures, all-cause mortality counts, age groups (only 65-74 and 75+), and dates, which were used to track the lag dimension (accounting for 21 days). For projection purposes, the methodology derives a relative risk function for the temperature-mortality association from an initial bidimensional surface which accounts for both temperature and lag time since exposure (Fig S3.2). These functions use an empirically determined minimum mortality temperature (MMT) as a baseline relative risk (RR = 1.0) for excess mortality. By definition, relative risk increases as temperatures deviate from this minimum mortality level, but in some cases (especially at rarely observed extreme temperatures), model uncertainty resulted in negative RRs, representing lower-than MMT mortality risk. We used these values unmodified in analysis. Given the bidirectional nature of this arrangement, moving forward we describe deaths associated with below-MMT temperatures as "cold"-associated, and above-MMT temperatures as "heat"-associated, although given the range of local MMTs, a "cold" temperature in e.g. Phoenix could be a "hot" temperature in New York.

We performed two brief sensitivity analyses to assess the relative stability of the exposure-response functions generated by this process. First, we generated a set of exposure-response functions using only the last ten years of historical temperature exposures (1995-2005, rather than 1987-2005) in consideration of findings (Bobb et al. 2014b) indicating a reduction in heat-attributable excess mortality between 1987 and 2005 in the NMMAPS dataset (if with appreciable regional heterogeneity. To assess the influence of outliers in the form of extreme temperatures, we also generated a set of exposure-response functions using the central 99.8% of temperatures for each city (i.e., excluding the most extreme 0.1% maximum and minimum temperatures). We used these sets of

functions in parallel with the core data to generate corresponding mortality estimates as described in the following steps.

2.5 Mortality Projections

Before we could convert the exposure-response functions, which return relative mortality risk, into concrete mortality count estimates, we needed to obtain estimates for total expected mortality on a given day and year. To accomplish this, we multiplied the combined annual mortality totals (described in section 2.3) by the median fraction of deaths occurring on each day of the 1987-2005 NMMAPS records for each city. For example, if for a given city and age group, a median of 2% of deaths occurred on January 1st across the NMMAPS dataset, and 1,000 total deaths were projected for a year of interest, we would assign 20 deaths as the baseline for expected mortality on that date before incorporating the influence of the temperature-mortality model. From this point, we were able to apply our exposure-response functions to estimate the number of temperature-attributable deaths using the formula:

$$D_{attr} = D \cdot (1 - e^{-(f^*\left(T_{mod};\theta_b^*\right)-s^*(T_{mm};\theta_b^*))}) \quad (1)$$

Where D_{attr} is the number of attributable deaths, D is the total number of deaths, $f^*\left(T_{mod};\theta_b^*\right)$ is our empirically-derived function for the temperature-mortality association (we differ from the Vicedo-Cabrera approach in avoiding downscaling and calibrating to bring estimates more in line with observed data, which can introduce additional uncertainty), with the lag component of the response curve having been "collapsed" into a binary version of cumulative risk, and $s^*(T_{mm};\theta_b^*)$ being a function giving the baseline trend of mortality with respect to seasonal and longer-term trends (again with the binary version of cumulative risk), this time in reference to T_{mm}, or the temperature at which the minimum mortality occurs (Vicedo-Cabrera et al. 2019); the asterisk notation refers to the counterfactual nature of the calculation (Gasparrini 2011).

2.6 Uncertainty Analysis

Our exposure-response function and resulting estimates incorporated multiple forms of uncertainty and alternate distributions, so we estimated uncertainty propagation using a Monte Carlo process, where calculations are repeatedly re-run to allow different combinations of "draws" from the various model probability spaces, and resulting empirical uncertainty intervals (eCIs) can be derived: in this case the 2.5th, 50th, and 97.5th percentiles of the resulting estimates were used to generate a 95% eCI. Specifically, we took 100 "draws" from the exposure-response function's probability space for each of the ten CESM2 ensemble members, yielding a composite 1000

attributable mortality estimates for every simulated day in the CESM2 data (keeping population and demographic assumptions constant). We aggregated both ensemble members and exposure-response function uncertainty to obtain central tendencies for our 95% eCIs, in contrast to approaches keeping ensemble members separate to further explore within-scenario variability (Keys et al. 2022). To obtain cumulative attributable deaths over the modeled 2050-2060 period, we added the sums of all daily estimates within a unique Monte Carlo "draw" before we calculated percentiles for the 95% eCI, although a smaller interval could be used for e.g. year-over-year trends. Despite the operational definition of MMTs as the 'minimum' relative risk, uncertainty intervals especially for rare or highly variable intervals (see figure 3.3) often dropped below zero. To capture the full range of uncertainty, we treated these "protective" estimates normally when taking summary statistics of net or heat/cold-specific temperature-related mortality and uncertainty propagation.

3. Results

3.1: Agreement between Historical and Simulated Temperatures

For context on the following estimates, it is important to consider agreement between the historical NMMAPS data and CESM2 projections, given methodology difference in resolution (point samples compared to a 1°, or roughly 110 km, grid) and temperature (surface or near-surface sensors compared to modeled 2 m air temperature). Comparing temperature distributions between the latest historical year (2005) and earliest CESM2 year (2015) (Figure S3.1), indicated generally good agreement in terms of the absolute range and relative frequency of temperatures. However, there were exceptions in Miami, FL, where historical temperatures were more evenly distributed and extended roughly 8°C below, and 2°C above, their CESM2 counterparts, and Los Angeles, where the CESM2 data extended roughly 4°C below, and 10°C above, the NMMAPS temperatures interval, with a more even distribution of temperatures. We used histograms to compare temperature distributions between the operationally defined cities' latest NMMAPS records (2005) and earliest CESM2 records (2015), with most being fairly comparable, (Figure S3.3). Still, Miami exhibited a narrower distribution of temperatures (i.e. locally "extreme" cold and heat is much rarer than in the NMMAPS data). The CESM2 rendition of Los Angeles was appreciably warmer than its NMMAPS counterpart, seeing a nearly +5°C shift in its relative temperature distribution.

3.2: City Projected Populations and Baseline Mortality Rates

Because the exposure-response functions express attributable mortality in terms of relative risk, we needed to determine underlying populations, as well as baseline mortality rates, using the projected SEDAC population counts, and census-based mortality rate estimates. Summary values (showing averaged annual figures over 2050-2060) are provided in Table 3.1:

City	Average 1987-2005 Temperature (°C)	Age Group	Population in Age Group (annual mean 2050-2060)	Projected Annual Mortality in Age Group (mean 2050-2060)
Chicago	14.03	65-74	593,282	6,326
		75+	595,358	35,488
Houston	24.62	65-74	835,796	8,854
		75+	930,371	61,184
Los Angeles	20.8	65-74	1,366,058	14,460
		75+	1,754,651	122,347
Miami	26.99	65-74	529,732	5,647
		75+	676,420	45,405
New York	14.86	65-74	1,370,910	14,513
		75+	1,397,866	89,268
Philadelphia	15.94	65-74	209,838	2,192
		75+	171,799	10,714
Phoenix	23.64	65-74	817,703	8,727
		75+	1,079,774	73,958
Seattle	11.84	65-74	405,629	4,273
		75+	430,242	28,917

Table 3.1: Summary Statistics by city including average temperature over NMMAPS data, projected age group-specific populations and baseline mortality counts.

3.3: Distributed Lag Nonlinear Model exposure-response functions and Minimum Mortality Temperatures

The DLNM exposure-response functions (Figure 3.3) correlating exposures yielded a variety of temperature-mortality associations (full set for ages 65-74 and 75+ in Figure S1). There was a range of general exposure-response function trends, with most northern cities (and Los Angeles' 75-and-older group) having steeper increases in relative mortality risk for over-MMT ('hot') temperatures, in some cases (New York, Seattle) exceeding the maximum below-MMT relative risk. In southern cities, relative risk above the MMT showed much smaller changes, with age groups in several city/age group combinations (likely erroneously) having the lowest predicted mortality at the maximum temperature. By contrast, most MMTs were otherwise within several degrees of 20°C. Central estimates of relative risk (RR) for temperature-associated excess all-cause mortality ranged from 1.0 (for MMTs, by definition), to a maximum of 4.47 at Miami's lowest observed temperature of 3.1°C for the 65-74 age group, although this was an

outlier, with the next-highest being an RR of 2.84 for Miami's 75+ age group, and 2.23 for the highest RR outside of Miami (Seattle's 65-74 age group) outside of Miami, which may be somewhat unreliable (especially at the cold extreme) given the disparity between CESM2 and historical NMMAPS exposure characterizations in the former. There were wide uncertainty intervals in some cases—generally associated with relatively extreme conditions—leading to apparent "protective" effects and, more rarely, extreme (and likely unrealistic) relative risks—reaching almost 40.0 as an extreme in Miami's 65-74 age group. For summary purposes, we divided temperature-associated mortality into heat-associated (above-MMT) and cold-associated (below-MMT) temperatures, as well as net temperature-related mortality by taking the sum of the two.

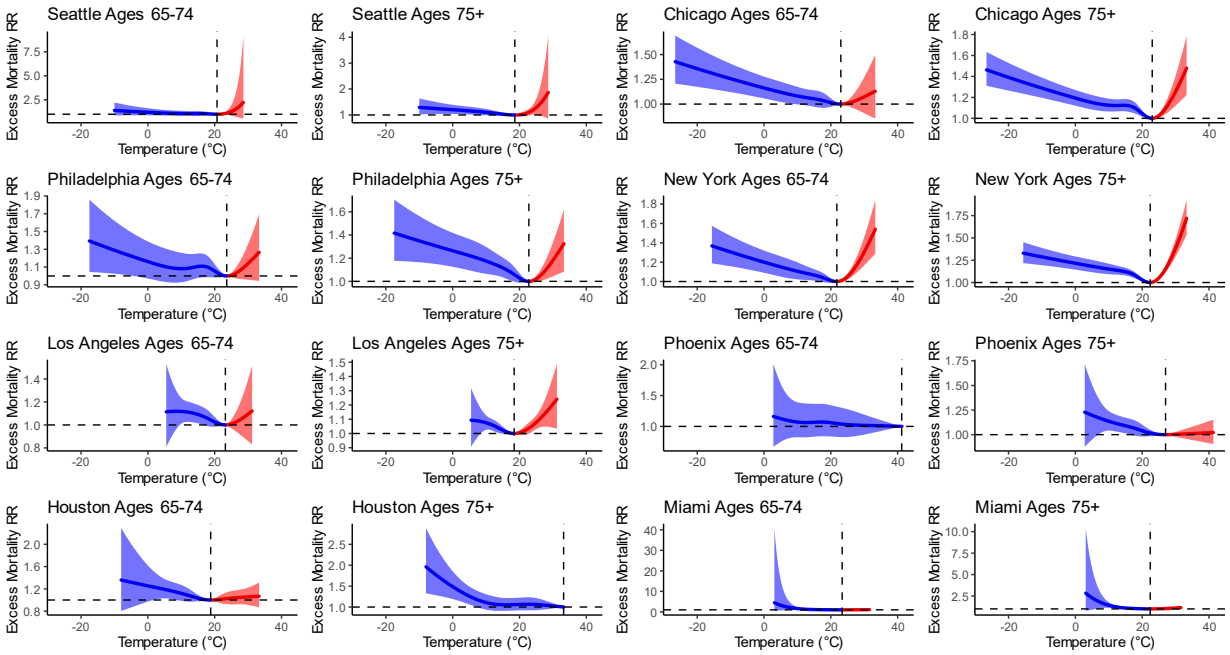

Figure 3.3: Exposure-Response Functions (exposure-response functions) for the association between temperature and excess mortality relative risk in 1987-2005 NMMAPS cities and daily average temperatures (mean of daily maximum and minimum) Lines and shading indicate central estimates and a 95% confidence interval, respectively, and colors are used to indicate RRs associated with "cold" (under-optimal) temperatures, and "hot" (above-optimal) temperatures relative to the minimum mortality temperature (represented by a vertical dotted line. Y axis scales vary freely to exhibit RRs observed between exposure-response functions, while x axis values are fixed to the total temperature range observed across the NMMAPS dataset.

Our sensitivity analyses using only temperatures in the final decade of historical data (Figure S3.4A) and after excluding the most-extreme observations (retaining the central 99.8% of temperatures about the median; Figure S3.4B) showed broadly similar results especially for higher-populations cities like New York. However, some dramatic differences emerged—notably in Miami, where both sets of alternate exposure-response functions showed much lower uncertainty intervals, and the 65-74 age group having the fewest predicted deaths at the (relative)

minimum. We observed other changes in Houston and Phoenix, where the relative placement of the MMT varied (moving from the maximum in the full dataset to a range more in line with the other city-age group observations). Despite the nominal positions of the MMTs shifting in these cases, the overall trajectory of the temperature-mortality associations tended to show less variation, pointing to possible limitations of this dynamically determined "hot" and "cold" designation system. However, the trend was not strictly population-dependent; for instance, Philadelphia, as one of the smaller cities, showed relatively stable trend across all three versions, while Los Angeles, despite having one of the largest populations, showed appreciable variation at the edges of the exposure response function. Overall, these results suggest that, as conceived, our protocol is robust provided large populations and relatively normally distributed temperatures, but that the predicted functions become less stable with one or both of smaller populations and extreme outlier temperatures.

3.3: Projected Mortality

We compared corresponding daily temperature-associated excess mortality estimates over the projected periods by applying the above exposure-response functions to the projected ARISE-SAI or unmodified SSP2-4.5 scenarios in the CEMS2 models. We provide summary all-cause mortality totals over 2050-2060 (across ensemble members) for attributable minimum mortality by climate scenario in Table S3.2, and visually summarized these values in figure 3.4 as temperature-attributable mortality rates per 100,00 residents between 2050 and 2060 and total counts in Figure S3.5.

Overall, the resulting morality estimates for the SSP2-4.5 and ARISE-SAI scenarios showed a high degree of similarity, with extensive overlap in the 95% empirical confidence intervals at our analysis' level of certainty and no decisive separations. Net temperature-associated median mortality rates between 2050 and 2060 ranged from a maximum of 779 and 772 per 100,000 residents (ARISE-SAI and SSP2-45, respectively) in New York's 75+ age group through a minimum of 38 and 46 per 100,000 (ARISE-SAI and SSP2-4.5, respectively) for Miami's 65-74 age group. For the former (New York's 75-and-older age group), the median cold-associated mortality rates were 599 and 565 per 100,000, respectively, while the median heat-associated mortality rates were 183 and 213, respectively. For Miami's 65-74 age group, the model categorized all temperature-associated mortality as heat-associated (that is, warmer than the minimum mortality temperature), in a probable artifact of our classification system, and illustration of limitations of dynamically determined MMTs. The relatively magnitude of differences between medians in New York is fairly indicative of wider trends; the largest modeled median difference between

scenarios was a 38 per 100,000 increase in cold-associated mortality under ARISE-SAI for Seattle's 75+ age group, followed by a 37 per 100,000 decrease for heat-associated and total temperature-associated mortality for Miami's 65-74 age group, with an overall median difference (in median mortality rates) of 5.3 per 100,000 and mean of 11.4 per 100,000; the smallest differences were less than 1 per 100,000 over the 2050-2060 period, with e.g. 53.3 net temperature-associated deaths per 100,000 in Houston under SSP2-4.5 and 53.5 under ARISE-SAI. Somewhat counterintuitively, multiple lower 95% eCI boundaries (that is, the 2.5th percentile of Monte Carloe estimates) extended below zero, implying that extreme (relative to MMT) temperatures would have a protective effect on mortality. This is a result of the wide uncertainty intervals, especially when extrapolating to generally warmer temperatures unobserved in the historical data.

However, some quantitative trends in differences between scenarios did emerge. Cold-related mortality was generally lower in the SSP2-4.5 scenario than the ARISE-SAI scenario, as could be expected from the generally warmer temperatures in the former; this difference was more pronounced in the 75-and-older population in most cases. Intuitively, heat-related mortality was lower in the ARISE-SAI scenario given the cooling effect of this intervention. There were some exceptions, notably in Phoenix and Houston, where warm temperatures tended to have a weaker influence on mortality. Because there were generally a larger number of cold-associated deaths overall, the net effect tended to be a slight increase in net-temperature-associated mortality under ARISE-SAI (true in 12 of the 16 city/age group combinations), if generally on the order of several hundred, if not dozens, between 2050 and 2060. As mentioned, the confidence intervals extensively coincided in each case. In keeping with their higher baseline mortality rates, the 75-and-older age group tended to have higher temperature-attributable excess mortality rates and counts overall; this group also showed more pronounced differences between cities, showing higher temperature-associated mortality (especially cold-related) in more northern cities.

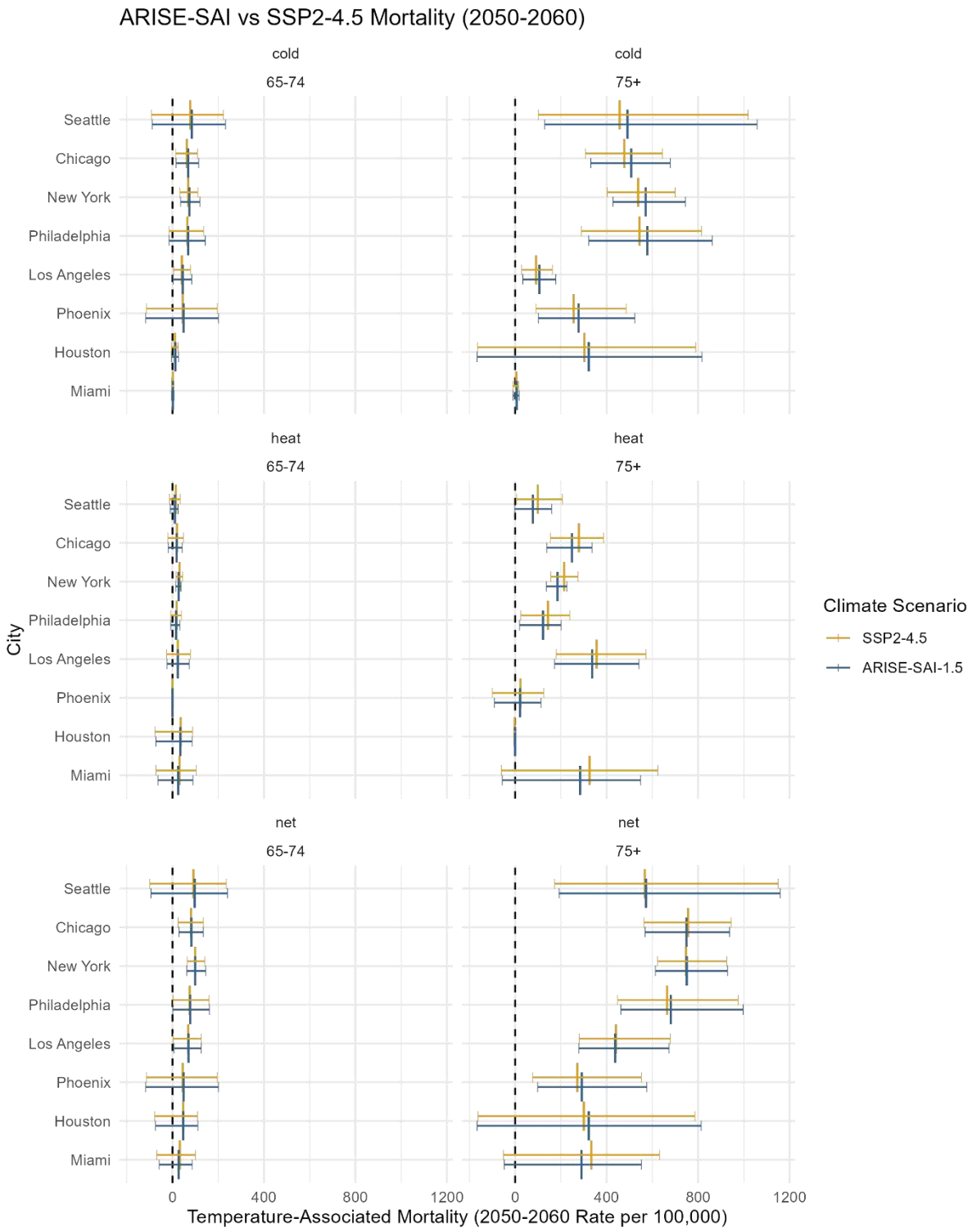

Figure 3.4: Comparison of Modeled Temperature-Associated Mortality (2050-2060 rate per 100k residents). We present mortality rates for SSP2.4-5 (gold) and corresponding ARISE-SAI climate scenarios (blue). The vertical lines indicate (left to right), the 2.5% percentile, 50% percentile (i.e. median), and 97.5% percentile of the Monte Carlo estimates, creating a 95% empirical confidence interval across the various sources of uncertainty in our analysis. Predictions are arranged vertically by temperature category (cold-associated, heat-associated, net temperature-associated) in panes, and by city within panes, with age groups (65-74, 75+) arranged in side-by-side panes.

The sensitivity analyses (Figures S3.5 and S3.6 for the 1995-2005 and central 99.8% versions, respectively) showed broadly similar trends despite the differences in some MMT locations and exposure-response function curves in most cities, with a notable exception in Miami's 65-74 age group, which has much wider confidence intervals than in the full data in both cases, implying an outsize effect from the excluded observations. There was a slightly stronger tendency towards higher mortality under ARISE-SAI than SSP2-4.5 compared to the full dataset. As with the full dataset, the magnitude of inter-scenario differences in net mortality tended to be around 3-5% of the SSP2-4.5 baseline total mortality estimates but varied appreciably—for instance with the central 99.8% of temperatures and just within Chicago, the 65-74 age group's net mortality was predicted to increase by more than 8% under ARISE-SAI, but less than 0.4% for the 75-and-older group. Phoenix and Houston had conspicuously small net and heat-specific effects, with each seeing a small net increase under ARISE-SAI due to more frequent and extreme below-optimal temperature-associated deaths, and little to no difference in above-optimal temperature-associated deaths due to the smaller relative risk of above-MMT exposures.

In summary, we projected that SAI would reduce heat-related all-cause mortality (especially in northern cities), but that higher cold-related mortality under a non-intervention SSP2-4.5 scenario (and the greater prevalence of this kind of mortality) would tend to lead to greater overall mortality under the ARISE-SAI scenario, at least from the mechanism modeled in this analysis. It is important to note that this comparison is between our two mid-21st century climate scenarios and specifically for direct temperature-related mortality, rather than dealing with the present day, or attempting to account for less direct temperature-linked effects on agriculture, political stability, etc.

4. Discussion

Overall, we projected a high degree of agreement between the scenarios in terms of mortality with the available confidence intervals, with the scenarios showing more similarities than differences. However, there were tendencies towards lower cold-associated mortality under SSP2-4.5, and lower heat-related mortality under ARISE-SAI. Net temperature-related mortality was quite similar, with near-complete overlap in 95% eCIs, and a median difference of around 5 deaths per 100,000 between the scenarios, although some cities and age groups saw central differences above 30 per 100,000. There was regional and city-specific heterogeneity in the temperature-mortality associations, with e.g. Houston and Phoenix having notably weaker heat-mortality, and overall temperature-mortality, associations than northern cities; this trend is reminiscent MMTs being closely associated with prevalent

local temperature conditions in a global meta-analysis (Yin et al. 2019). These results would be consistent with SAI reducing heat-associated mortality, but perhaps not net temperature-associated mortality for an urban United States population, which would be consistent with estimates (Gasparrini et al. 2017a), although the wide uncertainty intervals complicate a definitive statement.

Our results, and corresponding sensitivity analyses, also highlighted some potential limitations of the current methodology, and directions for future directions. Notably, the sensitivity analysis shows that while the population and exposure data from some large cities, like New York and Chicago, is able to generate fairly robust exposure-response functions that remained essentially unchanged, other cases like Miami suggest that the method is fairly sensitive to sample size (both for population and the distribution of deaths at a given temperature), with even small numbers of outlier observations being capable of dramatically shifting the calculated MMT and model uncertainty. Combined with the noted discrepancies between remote sensing-based CESM2, and ground sensor-based NMMAPS temperatures, we have limited confidence in predictions for some cities. While some limitations, like the wide confidence intervals, may be difficult to circumvent (although, for instance, larger Monte Carlo estimate pools, pooling similar climates/populations for larger effective sample sizes, or a different handling of year-over-year variation within the sampling period, could alleviate some of the variability), others are more immediately addressable in future research, as will be discussed below.

Our investigation represents a combination of well-established methodology for distributed lag nonlinear model-based health impact assessments (Gasparrini et al. 2010), with a recently revised SAI intervention scenario, while leveraging widely used datasets available over at least the United States, this provides a readily extensible platform to offer insight into temperature-associated mortality as a dimension of potential SAI public health impacts. We also modeled city-specific temperature-mortality associations, a valuable feature for capturing population-specific trends, whereas a uniform (even e.g. North America-based) equivalent would sacrifice the kind of regional granularity critical for understanding the costs and benefits of SAI or a similar intervention. We allowed compatibility with existing or future research by building on the widely used CESM2 and SSP frameworks; these design choices would also facilitate follow-up research using e.g. alternate climate scenarios, interventions, and health impacts.

Our methodology also included separate analysis for 65-74 and 75+ age groups, although we did not observe particularly dramatic differences in temperature-mortality associations between age groups (while there

were several-degree MMT differences the magnitude was unlikely to inform e.g. age-specific risk management plans) , lending credence (at least among the tested populations) to approaches that operationally assume a fixed MMT across age groups (Masselot et al. 2023). In two cases, the MMT was equivalent to the maximum observed temperature; while the true MMTs are likely lower, this was an artifact of an observed 'flat' temperature-mortality trend for warmer conditions, which seems plausible given the protective role of shelter and services like air conditioning during hot conditions.

At the same time, this investigation has notable limitations and assumptions. First, our underlying approach relies on projected datasets with specific narrative assumptions. Health Impact Assessments also carry an assumption of causality, here that temperature mechanistically causes the associated change in mortality. It is likely that despite our adjustments for long-term and seasonal trends in mortality, our methodology did not fully disentangle mechanisms like seasonal illnesses from the proximate effects of temperature. Mid-21st century temperature-mortality associations would also likely differ from those in the 1988-2005 historical period. Extrapolating from a previous analysis of the NMMAPS dataset (Bobb et al. 2014), heat-associated mortality could be expected to be lower than modeled here, although the study did not study cold-associated mortality. Conversely, greater heat-associated health burdens than historically observed may strain public health system capacity, potentially yielding higher than modeled excess mortality at least in the short term. Our choice of spline basis for the distributed lag nonlinear models will also influence the observed trends. Here, we used a penalized spline bases for the DLNM to provide conservative extrapolations beyond observed exposure (usually warmer-than-historical temperatures), but there may be steeper increases in relative risk in practice. A more pressing immediate limitation we intend to address in a future iteration is that differences between CESM2 and NMMAPS temperatures in Los Angeles and Miami require caution in interpretation, with even the "baseline" period with marginal projected warming would see apparent climate shifts and consequent mortality trends.

Moving beyond specific methodology, our conclusions on the relative efficacy of SAI are also specific to the modeled populations and intervention scenario. Most fundamentally, while the SSP2-4.5 scenario used in both the CESM2 and ARISE-SAI scenarios, as well as demographic modeling, is widely used and follow well-grounded assumptions, it represents one narrative among several (which include more- or less-optimistic scenarios with respect to emissions mitigation, etc.), meaning that our results are by no means conclusive or inevitable even within the assumptions of the larger SSP2 framework. As an exposure scenario, ARISE-SAI reflects well-justified

assumptions on details like technical feasibility, atmospheric features like the north-south temperature gradient, and deployment locations (Richter et al. 2022). But like its SSP2-4.5 basis, this deployment strategy and outcome is also not inevitable. Even within ARISE-SAI outcomes, an analysis of separate ensemble member outcomes showed dramatic variation—for instance, in some cases regions like northern India and parts of North America could sustain ongoing warming even as global average temperatures fall (Keys et al. 2022). Naturally, other SAI intervention scenarios would carry their own variations in outcomes in terms of overall cooling effects, regional heterogeneity, etc. (Lee et al. 2021).

We identified several potential new lines of inquiry from this preliminary investigation and proof of concept, both to address immediate methodological limitations, and to expand the wider field of research into SAI health impacts. In the first category, an immediate methodological fix to concerns about exposure mischaracterization in especially Miami and Los Angeles would be to substitute NMMAPS exposure temperatures with historical remote sensing temperature data—e.g. ERA5 (European Centre for Medium-Range Weather Forecasts ReAnalysis version 5)—which would provide better a better methodological match with the CESM2 temperature metric, and could be resampled to match the coarser 1° spatial grid of the CESM2 output. This would also enable sensitivity analyses using e.g. daily maxima and minima in addition to NMMAPS-style daily means. Another step could be to attempt to explicitly model adaptations in temperature-mortality associations, ideally based on more recent historical data if available, in order to reflect the kind of plasticity observed in heat-mortality associations in Bobb et al (2014); for instance, Yin et al. (2019) extrapolate from current-day associations between city-level MMTs and prevalent local conditions to estimate midcentury MMTs in a number of global cities and regions. Given the similarity between at least 65-74 and 75+ age group MMTs in our observations, a similar approach for comparing SAI costs and benefits could be valuable.

More ambitiously, some issues with effective sample size, especially in extreme temperatures, could be alleviated using an expanded exposure-response model built on an integrated Bayesian framework to make better use of the available sample population. Specifically, estimates from larger causal-inference model, trained on composite populations of similar cities, being fed into city-specific predictive models accounting for local characteristics and demographics (following the methodology of e.g. Nethery et al. 2020). This could increase the effective sample size without losing city-level specificity, potentially reducing the high sensitivity to outliers seen in the sensitivity analyses for exposure-response functions, and potentially narrowing the uncertainty intervals to

provide a clearer picture of relative differences between scenarios for 2050-2060 temperature-associated mortality. Especially for city- or region-specific implementations of a similar methodology, stakeholder engagement and other forms of more granular local data gathering could also provide a useful perspective for researchers. For instance, warmer cities like Phoenix and Houston had notably mild heat-mortality associations for the 75-and-older populations, which we assumed may be a factor of access to air conditioning and time spent indoors, but follow-up analysis drawing on local expertise, surveys, or field research could provide a more definitive answer, and potentially highlight transferrable strategies that could be applied to other populations which may not be discernible at the county-level granularity used in the current methodology.

In terms of practical implications of the analysis (with the noted limitations and low confidence for lower-population or more variable cities like Miami), health impacts of even baseline climate change showed appreciable regional heterogeneity. For example, in a DLNM-based health projection, North America was predicted to see higher net temperature-associated mortality only under the least optimistic (with respect to mitigation efforts) of the modeled climate change scenario and late in the 21st century, while even the most optimistic scenario had Southeast Asia seeing heat-driven excess mortality growth by the 2030s (Gasparrini et al. 2017a). Accordingly, projected ARISE-SAI influence on more heat-vulnerable populations may well show decisive reductions in net temperature-associated mortality. An additional consideration is that cold-associated mortality has some fairly context-dependent mechanisms like airborne viral particulates (Dbouk and Drikakis 2021) or social interactions (Moriyama et al. 2020). Given this, it is possible that heat stress may have a more robust and "mechanistic" link to mortality rates, while cold-related factors like seasonal illnesses and indoor congregation may be less response to warming than historical modeling may imply. Our outcome metric, excess temperature-associated mortality, also represents a single facet of comprehensive SAI-associated health costs and benefits. Beyond proximate temperature reduction effects, SAI could influence earth systems like precipitation and carbon cycling (Lee et al. 2021), with potential political ramifications (Halstead 2018), although climate change without intervention would also be associated with instability in these respects (Intergovernmental Panel On Climate Change 2023). There have been proximate (if model-dependent) SAI-driven health risks identified, including wider exposure to surface-level sulfates as injected compounds settle out of the atmospheric, or increased UV exposure due to ozone depletion from interaction with increased stratospheric SO_2 (Eastham et al. 2018; Tracy et al. 2022). As mentioned in the introduction, however, the point of reference for the health efficacy of SAI is not present-day conditions, but a corresponding non-intervention

climate change scenario, which comes with its own distinct health costs and modeling uncertainties. For instance, in the latter case of UV-B radiation exposure, some greenhouse gases like methane and nitrous oxide are also capable of reducing ozone levels and consequential UV-B radiation exposure (Barnes et al. 2019).

In terms of practical implications, our findings suggest ARISE-SAI, relative to a non-intervention "control" scenario of warming under SSP2-4.5, would have fairly small impacts on direct temperature-attributable mortality with mixed-to-slightly increased tendencies in median net temperature-associated mortality in eight major United States cities between 2050 and 2060 (although with extensive overlap in 95% empirical confidence intervals).This tendency in the net effect represents a combination of generally lower heat-associated mortality under SAI, and lower cold-associated mortality under warmer SSP2-4.5 conditions.

5. Conclusion

In this framework, our results, which show that successful SAI implementation as modeled by ARISE-SAI would see at most marginal changes in net temperature-associated mortality overall, but a more robust shift from heat- to cold-related mortality under SAI (or more accurately, that SAI is better able to preserve the current cold-dominated distribution of temperature-related mortality), poses something of a philosophical question in a public health context, which is whether heat- or cold-related deaths are somehow preferable. To focus on pragmatic implications (compared to e.g. subject quality of life during either outcome), one consideration could be whether one cause of mortality is easier to manage or mitigate than the other—for instance, if one or the other not only in a specific locale but globally could be managed with less reliance on tools like air conditioning units (for heat-related deaths), or is easier to apply preventative measures (e.g. setting municipal policies for school and relevant work closures heat- or cold-related closures under extreme conditions, making warming or cooling shelters available). While the present iteration of this project unfortunately still requires some improvement to capture baseline uncertainty, and could see improvements across future iterations using methods like the most frequent temperature method to predict future MMTs and adjust ERFs to account for climate adaptation, this presents a strong initial starting point for demonstrating the viability of this method, and providing a quantitative look into potential temperature-associated all-cause mortality implications of SAI.

References

Armstrong B. 2006. Models for the Relationship between Ambient Temperature and Daily Mortality. Epidemiology 17: 624–631.

Bandyopadhayaya S, Ford B, Mandal CC. 2020. Cold-hearted: A case for cold stress in cancer risk. Journal of Thermal Biology 91:102608; doi:10.1016/j.jtherbio.2020.102608.

Barnes PW, Williamson CE, Lucas RM, Robinson SA, Madronich S, Paul ND, et al. 2019. Ozone depletion, ultraviolet radiation, climate change and prospects for a sustainable future. Nat Sustain 2:569–579; doi:10.1038/s41893-019-0314-2.

Barnett AG, Huang C, Turner L. 2012. Benefits of Publicly Available Data. Epidemiology 23:500; doi:10.1097/EDE.0b013e31824d9ef7.

Best AF, Haozous EA, Berrington de Gonzalez A, Chernyavskiy P, Freedman ND, Hartge P, et al. 2018. Premature mortality projections in the USA through 2030: a modelling study. The Lancet Public Health 3:e374–e384; doi:10.1016/S2468-2667(18)30114-2.

Bobb JF, Peng RD, Bell ML, Dominici F. 2014. Heat-Related Mortality and Adaptation to Heat in the United States. Environmental Health Perspectives 122:811–816; doi:10.1289/ehp.1307392.

Danabasoglu G, Lamarque J -F., Bacmeister J, Bailey DA, DuVivier AK, Edwards J, et al. 2020. The Community Earth System Model Version 2 (CESM2). J Adv Model Earth Syst 12:e2019MS001916; doi:10.1029/2019MS001916.

Dbouk T, Drikakis D. 2021. Fluid dynamics and epidemiology: Seasonality and transmission dynamics. Physics of Fluids 33:021901; doi:10.1063/5.0037640.

Eastham SD, Weisenstein DK, Keith DW, Barrett SRH. 2018. Quantifying the impact of sulfate geoengineering on mortality from air quality and UV-B exposure. Atmospheric Environment 187:424–434; doi:10.1016/j.atmosenv.2018.05.047.

Evans SS, Repasky EA, Fisher DT. 2015. Fever and the thermal regulation of immunity: the immune system feels the heat. Nat Rev Immunol 15:335–349; doi:10.1038/nri3843.

Gasparrini A. 2011. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. J Stat Softw 43: 1–20.

Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, et al. 2015. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. The Lancet 386:369–375; doi:10.1016/S0140-6736(14)62114-0.

Gasparrini A, Guo Y, Sera F, Vicedo-Cabrera AM, Huber V, Tong S, et al. 2017a. Projections of temperature-related excess mortality under climate change scenarios. The Lancet Planetary Health 1:e360–e367; doi:10.1016/S2542-5196(17)30156-0.

Gasparrini A, Scheipl F, Armstrong B, Kenward MG. 2017b. A Penalized Framework for Distributed Lag Non-Linear Models. Biometrics 73:938–948; doi:10.1111/biom.12645.

Greaney JL, Kenney WL, Alexander LM. 2016. Sympathetic regulation during thermal stress in human aging and disease. Autonomic Neuroscience 196:81–90; doi:10.1016/j.autneu.2015.11.002.

Halstead J. 2018. Stratospheric aerosol injection research and existential risk. Futures 102:63–77; doi:10.1016/j.futures.2018.03.004.

Hauer M, Center For International Earth Science Information Network. 2020. Georeferenced U.S. County-Level Population Projections, Total and by Sex, Race and Age, Based on the SSPs, 2020-2100.; doi:10.7927/DV72-S254.

Hauer ME. 2019. Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. Sci Data 6:190005; doi:10.1038/sdata.2019.5.

Health Effects Institute. 2000. National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and Mortality from Air Pollution in the United States. Health Effects Institute. Available: https://www.healtheffects.org/publication/national-morbidity-mortality-and-air-pollution-study-part-ii-morbidity-and-mortality-air [accessed 22 February 2023].

Hendel M, Azos-Diaz K, Tremeac B. 2017. Behavioral adaptation to heat-related health risks in cities. Energy and Buildings 152:823–829; doi:10.1016/j.enbuild.2016.11.063.

Hess KL, Wilson TE, Sauder CL, Gao Z, Ray CA, Monahan KD. 2009. Aging affects the cardiovascular responses to cold stress in humans. Journal of Applied Physiology 107:1076–1082; doi:10.1152/japplphysiol.00605.2009.

Intergovernmental Panel On Climate Change. 2023. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st ed. Cambridge University Press.

Iungman T, Cirach M, Marando F, Pereira Barboza E, Khomenko S, Masselot P, et al. 2023. Cooling cities through urban green infrastructure: a health impact assessment of European cities. The Lancet 401:577–589; doi:10.1016/S0140-6736(22)02585-5.

Jay O, Capon A, Berry P, Broderick C, De Dear R, Havenith G, et al. 2021. Reducing the health effects of hot weather and heat extremes: from personal cooling strategies to green cities. The Lancet 398:709–724; doi:10.1016/S0140-6736(21)01209-5.

Jiang K, Fu B, Luo Z, Xiong R, Men Y, Shen H, et al. 2022. Attributed radiative forcing of air pollutants from biomass and fossil burning emissions. Environmental Pollution 306:119378; doi:10.1016/j.envpol.2022.119378.

Keys PW, Barnes EA, Diffenbaugh NS, Hurrell JW, Bell CM. 2022. Potential for perceived failure of stratospheric aerosol injection deployment. Proc Natl Acad Sci USA 119:e2210036119; doi:10.1073/pnas.2210036119.

Lee H, Muri H, Ekici A, Tjiputra J, Schwinger J. 2021. The response of terrestrial ecosystem carbon cycling under different aerosol-based radiation management geoengineering. Earth Syst Dynam 12:313–326; doi:10.5194/esd-12-313-2021.

Masselot P, Mistry M, Vanoli J, Schneider R, Iungman T, Garcia-Leon D, et al. 2023. Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in Europe. The Lancet Planetary Health 7:e271–e281; doi:10.1016/S2542-5196(23)00023-2.

Moriyama M, Hugentobler WJ, Iwasaki A. 2020. Seasonality of Respiratory Viral Infections. Annual Review of Virology 7:83–101; doi:10.1146/annurev-virology-012420-022445.

Nethery RC, Katz-Christy N, Kioumourtzoglou M-A, Parks RM, Schumacher A, Anderson GB. 2020. Integrated causal-predictive machine learning models for tropical cyclone epidemiology. arXiv:201011330 [stat].

Rathjen NA, Shahbodaghi SD, Brown JA. 2019. Hypothermia and Cold Weather Injuries. Am Fam Physician 100: 680–686.

Riahi K, van Vuuren DP, Kriegler E, Edmonds J, O'Neill BC, Fujimori S, et al. 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change 42:153–168; doi:10.1016/j.gloenvcha.2016.05.009.

Richter J, Visioni D, MacMartin D, Bailey D, Rosenbloom N, Lee W, et al. 2022. Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI).; doi:10.5194/egusphere-2022-125.

Salata F, Golasi I, Ciancio V, Rosso F. 2018. Dressed for the season: Clothing and outdoor thermal comfort in the Mediterranean population. Building and Environment 146:50–63; doi:10.1016/j.buildenv.2018.09.041.

Sharifi E, Boland J. 2018. Limits of thermal adaptation in cities: outdoor heat-activity dynamics in Sydney, Melbourne and Adelaide. Architectural Science Review 61:191–201; doi:10.1080/00038628.2018.1482824.

Smith W. 2020. The cost of stratospheric aerosol injection through 2100. Environ Res Lett 15:114004; doi:10.1088/1748-9326/aba7e7.

Tracy SM, Moch JM, Eastham SD, Buonocore JJ. 2022. Stratospheric aerosol injection may impact global systems and human health outcomes. Elementa: Science of the Anthropocene 10:00047; doi:10.1525/elementa.2022.00047.

US Census Bureau. 2017. 2017 National Population Projections Datasets. Census.gov. Available: https://www.census.gov/data/datasets/2017/demo/popproj/2017-popproj.html [accessed 29 October 2023].

U.S. Census Bureau QuickFacts: Phoenix city, Arizona. 2021. Available: https://www.census.gov/quickfacts/fact/table/phoenixcityarizona/PST045219 [accessed 4 October 2021].

Vicedo-Cabrera AM, Sera F, Gasparrini A. 2019. Hands-on Tutorial on a Modeling Framework for Projections of Climate Change Impacts on Health. Epidemiology 30:321–329; doi:10.1097/EDE.0000000000000982.

Xi T, Wang Q, Qin H, Jin H. 2020. Influence of outdoor thermal environment on clothing and activity of tourists and local people in a severely cold climate city. Building and Environment 173:106757; doi:10.1016/j.buildenv.2020.106757.

Yin Q, Wang J, Ren Z, Li J, Guo Y. 2019. Mapping the increased minimum mortality temperatures in the context of global climate change. Nat Commun 10:4640; doi:10.1038/s41467-019-12663-y.

CHAPTER 5: CONCLUSION

1: Summary of Key Findings

In this body of work, we have applied the Qualitative Health Impact Assessment (quantitative health impact assessment) framework to exploring variety of environmental health systems and different ecological systems and scales with excess mortality as a common reference point. We also explored system relevant to different policy priorities and urban resilience in the "triad" of exposure reduction (in this case arguably Project 1, with its focus on identifying potentially overlook high-risk areas), vulnerability reduction (Project 2, with urban tree canopy acting as a buffer against 'ambient' adverse environmental exposures like heat, among other benefits), and hazard reduction (Project 3, in exploring a scenario with actively diminished anthropogenic warming) (Hoegh-Guldberg et al. 2018). At a methodological level, this has underscored the versatility of the quantitative health impact assessment framework as far as its ability to integrate and adapt existing data sources and models, while demonstrating some limitations and tradeoffs with, for instance, temporal resolution or coherence of available data sets. We also identified specific findings relevant to public health in each case.

Some common themes for strengths and weaknesses emerged on the methodological side across the projects. By opting to use uniform, and if possible publicly accessible datasets, our work could, in general, be readily expanded to other populations, health outcomes, and to some extent exposures with minimal adaptation (e.g. Project 2 could be extended to any US population represented in the census, if with some manual steps at its current state like cropping LANDSAT8 data to fit city boundaries; this could also be automated or run through an API in a future expansion or iteration). This builds both on the intrinsic modularity of quantitative health impact assessment methodology, as well as on the availability of methodologically-uniform datasets like the US census, raw or derived remote sensing data, standardized models like the CESM2, and in cases where privacy and confidentiality can be asserted, more restricted-access datasets like NMMAPS or the Medicare claims data used in Project 1. The causal inference approach used here allows us to explore hypothetical or future scenarios using existing population, health, and exposure data, improving the "return on investment" for these existing resources. However, our projects carry some quantitative health impact assessment-intrinsic and implementation-specific limitations. In terms of overall limitations of quantitative health impact assessments, they are strongly dependent on their assumptions and scenario modeling; while researchers should seek to accurately represent as many relevant dimensions as possible of the chosen population and exposure scenario, there an unavoidable level of abstraction, which could range from the

precision of available data (e.g. county-level health outcomes may be the smallest available geographic resolution, but meaningful health differences likely play out on a block-by-block, or e.g. family-by-family basis), although ideally this de facto selection bias would be a case of Berkson bias, where the observed trend would deviation towards the null. There are also often-unavoidable concessions in operational exposure definitions; for instance, in Projects 2 and 3, we use NDVI around a fixed address, and mean countywide outdoor temperatures, where in practice personal exposures likely vary appreciably over and between days based on work, travel, and other activities, and especially in the latter case would be heavily mediated by access to indoor conditions. In this sense, our quantitative health impact assessment-based approaches as implemented here could be seen as providing more of "guideposts" pointing towards magnitudes of impact than precise effects, although greater specificity could be achieved in some cases, especially for short-term, limited-scope interventions (e.g. anticipated impacts of a change to municipal infrastructure) where the spatial and temporal scope allows more granular data collection and less temporal projection is needed. Our projects tended to be somewhat chronologically ambiguous given the nature of the research questions, e.g. Project 2 was based on a concrete scenario, we assumed no change in the populations, environments, or relevant characteristics during the intervening years to decades realistically needed to reach target canopy goals; while we used demographic projections for 2050-2060 temperature-associated mortality in Project 3, the 1980-2005 period of the NMMAPS data, and no modeling of intervening population adaptation, limits applicability to even the theoretically defined future scenario.

In Project 1, we determined that STORM, as a means of extending the limited historical tropical cyclone exposure dataset with machine learning-generated plausible tracks, provided a reasonably close facsimile of tropical storms and hurricanes, if not other types of storms (e.g. sub- or extratropical depressions), and with a lower frequency of total storms and landfall events in our operational area (within 250 km of United States coasts). Within these caveats, we were able to generate a spatially and temporally contiguous map of tropical-cyclone associated excess all-cause mortality among 65-and-older US adults; these intensity-based findings have fairly intuitive policy implications, with at-risk areas in the synthetic STORM model likely being at greater long-term risk of tropical cyclone exposure even if they have not experienced this in the previous 40 years; also different temporal patterns may of some utility for expecting early versus late-season storms in a given region]. An interesting implication of our exposure-response function was that low-intensity wind exposures provide mixed-to-protective effects, with a net increase in mortality only when we limited analysis to hurricane-strength exposures. Naturally, this should not

be taken to suggest that hurricane seasons are a net positive for public health (even within the same model (Nethery et al. 2020), other health outcomes like hospitalization from respiratory illnesses increase more linearly with windspeed), but rather points to the relative success of current preparedness and response measures, especially in contrast to nearby regions which see higher associations between tropical cyclone exposure and excess mortality (Huang et al. 2023).

In Project 2, Our models projected Denver's 20% UTC goal would, if uniformly attained, avert about 200 deaths, 4.1 stroke, and 2.6 dementia cases annually, with a combined economic impact of $1.94 billion, with about 64% of these benefits captured by half-way attainment (16.5% UTC). Phoenix's 25% UTC target for 2030 would avert roughly 368 annual deaths, 8.7 stroke cases, and 5.1 dementia cases, with a combined economic impact of $3.57 billion, and the 17% "halfway" scenario seeing ~44% of these benefits. More vulnerable communities (by SVI quartile) generally saw greater proportional health benefits from UTC intervention. This provides supporting evidence for the pursuit of these goals, even if full attainment is not practical under real-world logistical considerations (e.g. funding, water resources). These findings were qualified by several abstractions (notably assuming uniform policy goal attainment in each population-weighted block group, and framing the intervention as occurring "instantaneously," rather than attempting to model population growth, tree maturation rates, etc.), but we would still expect a broadly beneficial impact on public health. On the socioeconomic dimension, we observed that regions higher on the Socioeconomic Vulnerability Index (SVI) scale (i.e. more vulnerable) were more likely to see lower current exposure to UTC ang general vegetation (on the normalized difference vegetation index), and consequently would stand to benefit more from local attainment of city canopy goals; this finding lends weight to both cities' expressed desire to prioritize underserved communities or otherwise reach equity in this respect.

Project 3 saw mixed results for the overall efficacy of stratospheric aerosol injection (SAI) as represented by the ARISE-SAI scenario (Richter et al. 2022) with respect to temperature-associated excess mortality. There was appreciable local variation in trends between cities. In some cases, age groups, but there were generally relatively muted changes in overall temperature mortality, albeit with the net trend representing a "transfer" of cold-related mortality to increased heat-related mortality under unmitigated warming. There was considerable variation in local associations between temperature and mortality. Generally, warmer cities tended to see weaker heat-temperature associations (with the reverse being true for cooler cities), representing local adaptation to prevalent conditions.

2: Policy Implications

Our investigations had project-specific policy implications, and provided examples supporting the utility of making public health-relevant datasets available to the general public or researchers where confidentiality is required. In terms of direct implications, project 1 points to a potential use case for simulated hurricane seasons as part of a long-term risk assessment process (if potentially using an algorithm derivative tuned to represent post-landfall behavior more comprehensively), and the specific spatial and temporal distributions of impacts may be of interest to stakeholders and policymakers in a preparedness context, especially for regions with a higher predicted tropical cyclone exposure frequency than reflected in the historical record. Project 2, by building on preexisting municipal policy goals, has fairly direct and intuitive policy implications; as described above, our findings provide support for the ongoing growth and maintenance of urban tree canopies, with even partial attainment projected to see appreciable public health benefits, especially concerning excess all-cause mortality, as well as the importance of prioritizing socially-vulnerable communities, and striving to ensure equitable access to tree canopies and other vegetation, with these populations having the largest capacity for UTC-based health benefits, especially in Denver (Phoenix, by its lower current UTC levels, was modeled to see more uniform benefits) . Of course, our model abstracted some important logistical considerations like the selection of tree species (e.g. the non-native species prevalent in especially Denver provide more shade and greenery, but tend to be require more water and are less able to endure drought stress), and constraints imposed by land access, economic and natural resources so that the specific model results can be interpreted as representative and potentially somewhat optimistic.

With the prospect of stratospheric aerosol injection subject to intensive interdisciplinary exploration, Project 3 explores potential effects of SAI (relative to unmitigated warming) on direct temperature effects on excess mortality in a handful of major American metropolitan areas. As noted above, there were mixed results, with the shift being more in the type of mortality than an overwhelming net reduction. In some cases, there was even a comparative increase (relative to unmitigated warming, rather than the present day) especially in heat-adapted, urban populations. As such, the project offers a small contribution to a larger ongoing discussion among researchers and policymakers. However, the framework could apply to other American cities, and internationally (provided equivalent daily mortality data, or suitable interpolation) with little effort. Given the high local variability of SAI effects, it is critical to perform compatible research broadly, especially in the Global South, to help understand the

range of implications globally to help the international community to work towards a, equitable consensus between all stakeholders as far as the baseline decision to intervene, and specific implementation choices.

By drawing on widely (if not, as in the case of the NMMAPS dataset or Medicare claims data undergirding Project 2's exposure-response function), publicly available data, our set of investigations also demonstrate the value of these resources, with the initial investment into these resources supporting ongoing research and public health-relevant findings. Methodologically uniform datasets like the LANDSAT8 measurements in Project 2, the CESM2 modeling framework used in Project 3, or census data used throughout are especially valuable by enabling researchers to develop and potentially expand flexible platforms and facilitating compatibility between research efforts. At the same time, these can lose some of the specificity of more localized data—for instance, while Denver's USFS remote sensing estimates for canopy cover were largely in line with contemporary municipal tree surveys, it reported much lower canopy cover than Phoenix's equivalent surveys, raising questions about the ability of this dataset to capture the characteristic vegetation of the area; similarly, we used estimated land use data, rather than likely more reliable zoning maps, for estimation population density in the interest of framework expansibility, due to the latter being stored in different locations and formats. This points to some limitations of current data, and the value of ongoing improvement and standardization of public efforts.

3: Future Directions for Research

In developing these projects and evaluating findings, we identified short-term, methodological improvements that could improve these or derived efforts, as well as potentially valuable larger-scale directions for inquiry. In project 1, our existing framework could be seamlessly used to explore alternative climate scenarios, either from preexisting datasets (e.g. the STORM team's simulations of a mid-21st century conditions under a range of climate models (Bloemendaal et al. 2023), or user-generated explorations like the health impacts of specific phases of the El Niño Southern Oscillation, or interactions with other recurrent climate patterns; because our approach uses a standardized IBTrACS-compatible format, the explorations would also not necessarily be tied to any single synthetic cyclone model; with sufficient computing resources, an ensemble-style approach using multiple models could help mitigate biases and limitations of any single model. Our quantitative health impact assessment could accommodate new functions in place of the current Bayesian exposure-response function for Medicare-aged Americans—for instance, a newer county-level model incorporating all age groups and additional health outcomes on a monthly basis (Parks et al. 2022). The code for the Nethery et al (2020) model is publicly available, so

researchers could adapt the overall process to new populations (especially given that the globally standardized tropical cyclone parameters, and specifically maximum sustained windspeed, had more influence in the model than the more US-specific census traits). A less direct route of explorations could be to modify STORM, or develop a complementary algorithm, tuned more for post-landfall behavior and better represent events like extra- and subtropical depressions.

As mentioned, Project 2 saw some abstraction of potentially important logistical features like selection of native versus introduced species (a question explored in (Garber et al. 2023), as well as with more spatially-explicit implementation scenarios), and constraints like water and economic resources; these could be explored in future studies, and would likely give a more realistic portrayal of likely outcomes (assuming uniform UTC attainment in every block group is admittedly optimistic, if an operationalization of a stated desire for equitable access). Within the quantitative health impact assessment as designed, future iterations could take advantage of more recently-released datasets; notably, NLCD-assessed "development" land use as a proxy for population centers could be replaced with a recent global dataset directly modeling populations at a 30m resolution (Meta 2023), which would have the advantage of avoiding confusion between e.g. industrial parks or even large roads and population centers, as well as accounting for higher predicted densities in e.g. apartment complexes compared to single family housing. Another example would be to update NLCD and USFS land cover data (2016 was the latest uniformly available timepoint while developing Project 2), which could not only represent more up-to-date conditions and improve temporal coherence. Ongoing improvements in machine vision and learning could also improve the accuracy of future equivalent datasets, which could apply even retroactively to archived remote sensing data.

Of the three projects, Project 3 has the most immediate room for improvement, as we've identified several short-term areas for improvement before moving from the current draft to publication. The most immediate planned improvement is to A smaller, but also immediate change could be to replace NMMAPS temperatures (pooled from specific sensors located in the cities of interest) with temporally equivalent 2m air temperature remote sensing data (e.g. ERA5, or the 5th generation EXMRWF [European Centre for Medium-Range Weather Forecasts, or a similar reanalysis dataset] Atmospheric Reanalysis data (Hersbach et al. 2020) resampled to match the spatial resolution of CESM2 gridpoints; this would improve methodological and spatial consistency, alleviation some of the discrepancies noticed with Miami and Los Angeles, where there was a palpable gap between the surface-level point samples and 2m spatially averaged values in even contemporary CESM2-based estimates. While the results

presented here are consistent with our methodology as designed, we also intend to take further exploratory steps to improve the robustness of the generated exposure-response functions to smaller sample sizes and/or extreme temperatures. One prominent, if involved, immediate improvement that could be implemented in Project 3 would be to augment the exposure-response function with an integrated Bayesian approach as in Project 1, where a causal-predictive model provides a national or pooled regional temperature-mortality trend that is modified by county-specific predictive model; this could help offset sample size limitations, allowing generalization to smaller communities, allow subsampling late in NMMAPS to provide a closer-to-contemporary association, and provide additional sampling power near the limits of currently-observed temperatures, which would improve extrapolation behavior. In the longer term, it would also be important to explore and incorporate populations' ability to adapt to temperatures over time; e.g. while the within-NMMAPS study in Bobb et al (2014) did not concern the same extent of warming as especially the range expected in unmitigated climate change, it does hint at an adaptive capacity in the range of years-to-decades; depending on the extent of this plasticity, the ability of SAI to slow temperature changes could be another important dynamic beyond what can be extrapolated from a fixed 1980-2005 trend.

4: Conclusion

This body of work explores the versatility of the quantitative health impact assessment framework by applying it to a variety of environmental exposures at different spatial and temporal scales and periods (past, near-present, and future), as well as different populations and exposure-response functions. It also demonstrates applicability to both 'ambient' risk characterization using historical data (Project 1) and different outcomes of specific policy decisions (Projects 2 and 3). By leveraging widely available and methodologically uniform datasets where available, our overall approach lays the foundation for future analyses exploring different populations, exposures, or outcomes, within the same frameworks. This complements the modularity of the quantitative health impact assessment framework. At the same time, this design choice, as well as the necessary degree of abstraction in using e.g. county-level census data or latest available tree canopy data several years behind the period of interest. This shows some of tradeoffs of our general approach, where complementary studies could obtain greater specificity that could be obtained by e.g. municipal survey data or on-the-ground fieldwork at the cost of reduced generalizability and added cost and complexity. Overall, this demonstrates the value and use cases of quantitative health impact assessments as an epidemiological analysis tool.

References

Bloemendaal N, de Moel H (Hans), Martinez AB, Muis S (Sanne), Haigh ID (Ivan), van der Wiel K, et al. 2023. STORM Climate Change synthetic tropical cyclone tracks.; doi:10.4121/14237678.V2.

Garber MD, Guidi M, Bousselot J, Benmarhnia T, Dean D, Rojas-Rueda D. 2023. Impact of native-plants policy scenarios on premature mortality in Denver: A quantitative health impact assessment. Environment International 178:108050; doi:10.1016/j.envint.2023.108050.

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, et al. 2020. The ERA5 global reanalysis. Quart J Royal Meteoro Soc 146:1999–2049; doi:10.1002/qj.3803.

Hoegh-Guldberg O, Jacob D, Bindi M, Brown S, Camilloni I, Diedhiou A, et al. 2018. Impacts of 1.5°C Global Warming on Natural and Human Systems. In: *Global warming of 1.5°C.* (V. Masson-Delmotte, P. Zhai, H.O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, et al., eds). IPCC Secretariat. 175–311.

Huang W, Li S, Vogt T, Xu R, Tong S, Molina T, et al. 2023. Global short-term mortality risk and burden associated with tropical cyclones from 1980 to 2019: a multi-country time-series study. The Lancet Planetary Health 7:e694–e705; doi:10.1016/S2542-5196(23)00143-2.

Meta. 2023. Data For Good at Meta High Resolution Population Density Maps. Available: https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps [accessed 30 October 2023].

Nethery RC, Katz-Christy N, Kioumourtzoglou M-A, Parks RM, Schumacher A, Anderson GB. 2020. Integrated causal-predictive machine learning models for tropical cyclone epidemiology. arXiv:201011330 [stat].

Parks RM, Benavides J, Anderson GB, Nethery RC, Navas-Acien A, Dominici F, et al. 2022. Association of Tropical Cyclones With County-Level Mortality in the US. JAMA 327:946; doi:10.1001/jama.2022.1682.

Richter J, Visioni D, MacMartin D, Bailey D, Rosenbloom N, Lee W, et al. 2022. Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI).; doi:10.5194/egusphere-2022-125.

SUPPLEMENTAL SECTION 1:

1: Supplemental Methods:

S1.1: Exposure Response Function and Health Impact Predictions

To model the association between our exposure (TC-driven county-level maximum sustained windspeeds) and outcome (all-cause storm-associated mortality among older adults), we used a previously developed epidemiological model that characterizes the association between tropical cyclone exposure (focusing on exposure to storm-associated winds) and all-cause mortality among older adults. The model incorporates two elements: a causal inference component about county-level health impacts after tropical cyclone exposure events, and a predictive component which modifies the output of the causal inference element using meteorological features and socioeconomic/demographic features of the counties (Nethery et al. 2020). The model operationalizes "exposed" counties as those experiencing a storm track passing within 150 km of the population-weighted county center, and consequentially sustaining Willoughby model-predicted sustained winds of at least 17.4 m/s (gale force). The causal model defined "controls" as unexposed neighboring counties within 150 miles of the exposed county's population-weighted centered, and for the predictive model, also considered mortality trends in the exposed county outside of tropical cyclone-exposed periods.

Because tropical cyclone public health impacts are not fully realized on initial exposure, the model's framework allows lag times to account for these post-exposure health outcomes, modeling the exposure window as a 14-day window beginning 2 days before, and 11 days after, the time of tropical cyclone to capture delayed excess mortality. The model also considers underlying trends based on day-of-year in the relative "control" data per county (Nethery et al. 2020). The causal model can obtain an effective "control group" both for individual counties by using the previous 18 weeks of health outcome incidence from impacted counties, split into 2-week intervals to give a better picture of baseline variance and any longitudinal trends, and by comparing comparable unexposed counties for a given event (Nethery et al. 2020).

The predictive model considers tropical cyclone exposure traits like the number of exposure events over the training period (1999-2015)—intended to reflect community-level adaptations to tropical cyclone events—location along a coast (using Census designations), year of tropical cyclone impact, and duration of sustained ≥20 m/s winds at the county center. Because the synthetic STORM algorithm designates seasons with arbitrary year identifiers (with no

chronological continuity) we assigned each tropical cyclone the year 2015 for modelling purposes. We assumed the number of prior tropical cyclone exposures in the historical IBTrACS training data would represent relevant factors. Of four modeled health outcomes (all-cause mortality, and hospitalizations associated with chronic obstructive pulmonary disease, respiratory illnesses, and cardiovascular disease), we selected all-cause mortality for this investigation. The model generates a metric of excess events (deaths in this case) per 100,000 county residents, with factors like underlying mortality rates captured implicitly in the training data using comparable control counties, and comparisons to the county's past trends outside of exposure periods.

The predictive model also considers demographic and socioeconomically-relevant data drawn from Census data (here, we use bicentennial 2020 Census data) including proportions of residents below the poverty line, identifying as white, aged at 65 or older, without at least a high school diploma, homes that are owner-occupied (as a proxy for wider socioeconomic factors), and county-level medians for age, population density, and home values; these were accessed from the 2020 decennial census using the *tidycensus* package (US Census Bureau 2022; Walker and Herman 2022). Of the model parameters used in the model, only sustained maximum windspeed had consistently positive estimated posterior mean coefficients for the influence on attributable all-cause mortality (i.e. the 95% confidence interval does not cross 0). However, some uncertainty distributions were more generally positive or negative (e.g. previous exposures, proportion living below poverty, and median house value), while others like percent owner-occupied housing or population density had fairly "centered" distributions, indicating less-consistent influences on health outcomes (Nethery et al. 2020).

The resulting exposure-response relationship (Figure 2) reveals a nonlinear relationship between tropical cyclone maximum sustained windspeed and excess all-cause mortality events (deaths) per 100,000 residents. The model had a lower threshold for maximum sustained windspeeds of 17.4 m/s, which was associated with a near-zero impact on excess mortality rates, which trended downward to a nadir of around 12 averted deaths per 100,000 in the mid-20 m/s range, before rebounding to a net hazardous effect by around 30 m/s, and increasing essentially linearly with windspeed for the rest of observed exposures, with widening confidence intervals in light of the increasing rarity of high-intensity exposures.

Figure S1.1: Example Conversion Output from Synthetic tropical cyclone Tracks to Population-Weighted County Centers

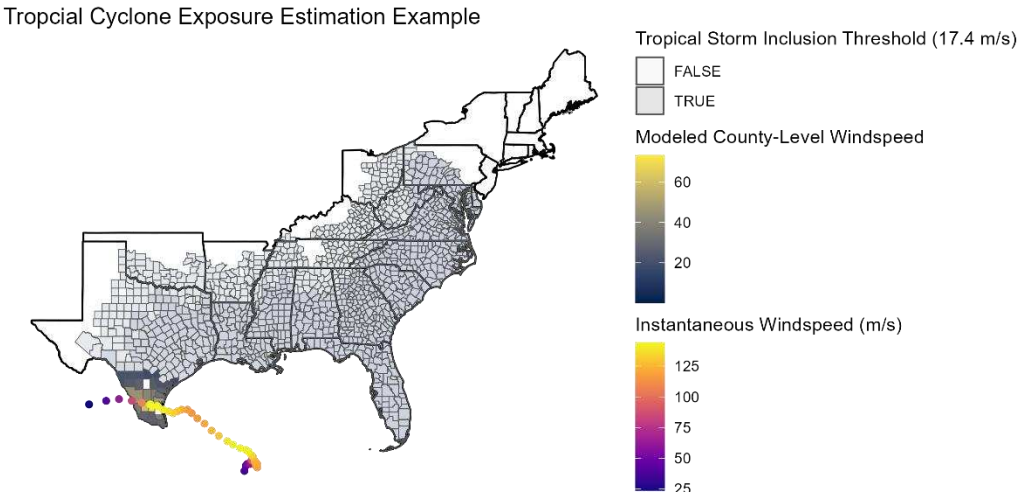

Figure S1.1: Example synthetic STORM-generated IBTrACS-compatible tropical cyclone track and corresponding county-level wind exposures after conversion with the stormwindsmodel *r package's implementation of the Willoughby wind field model. Opaque counties represent those with a maximum sustained windspeed of at least 17.4 m/s (minimum inclusions threshold for the exposure-response function), while lightly shaded counties represent counties with sub-threshold exposures. Color scales (m/s) represent the 3-hour maximum windspeeds generated by the STORM algorithm timesteps (represented as points), and the county-level maximum sustained windspeed in the interpolated 15-minuted wind fields output.*

Figure S1.2: Comparisons of Windspeed, Storms and Exposure Frequency by Minimum Windspeed

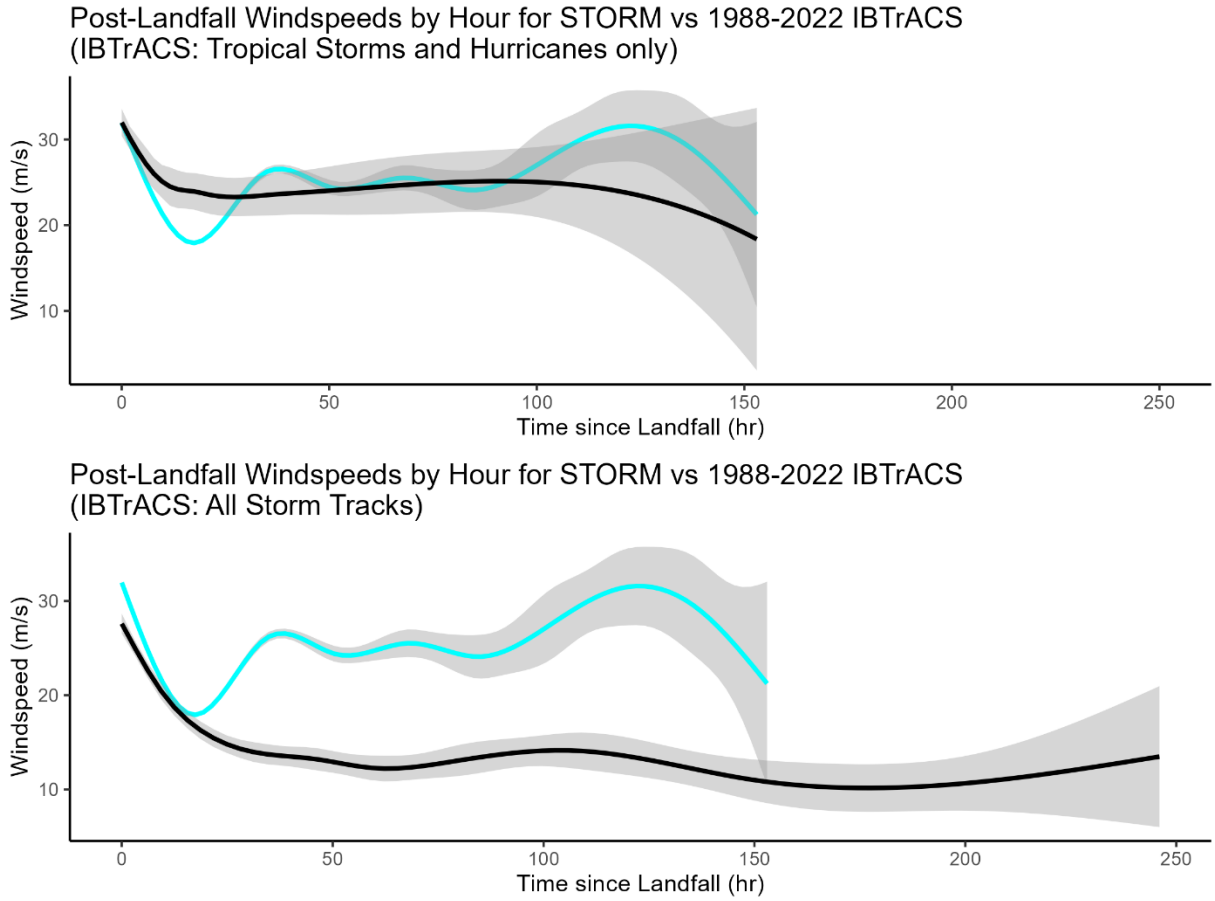

Figure S1.2: Comparisons of post-landfall windspeed decay in 10,000 simulated season in STORM (cyan) compared to 1980-2022 IBTrACS data (black) when filtered only to hurricanes and tropical storms (above) versus including all historical events (below); lines show central tendency for tropical cyclone track windspeeds in the hours after landfall, with grey indicating an uncertainty interval.

Figure S1.3: Extents of tropical cyclone Exposures by Dataset (STORM, Historical, and filtered Historical)

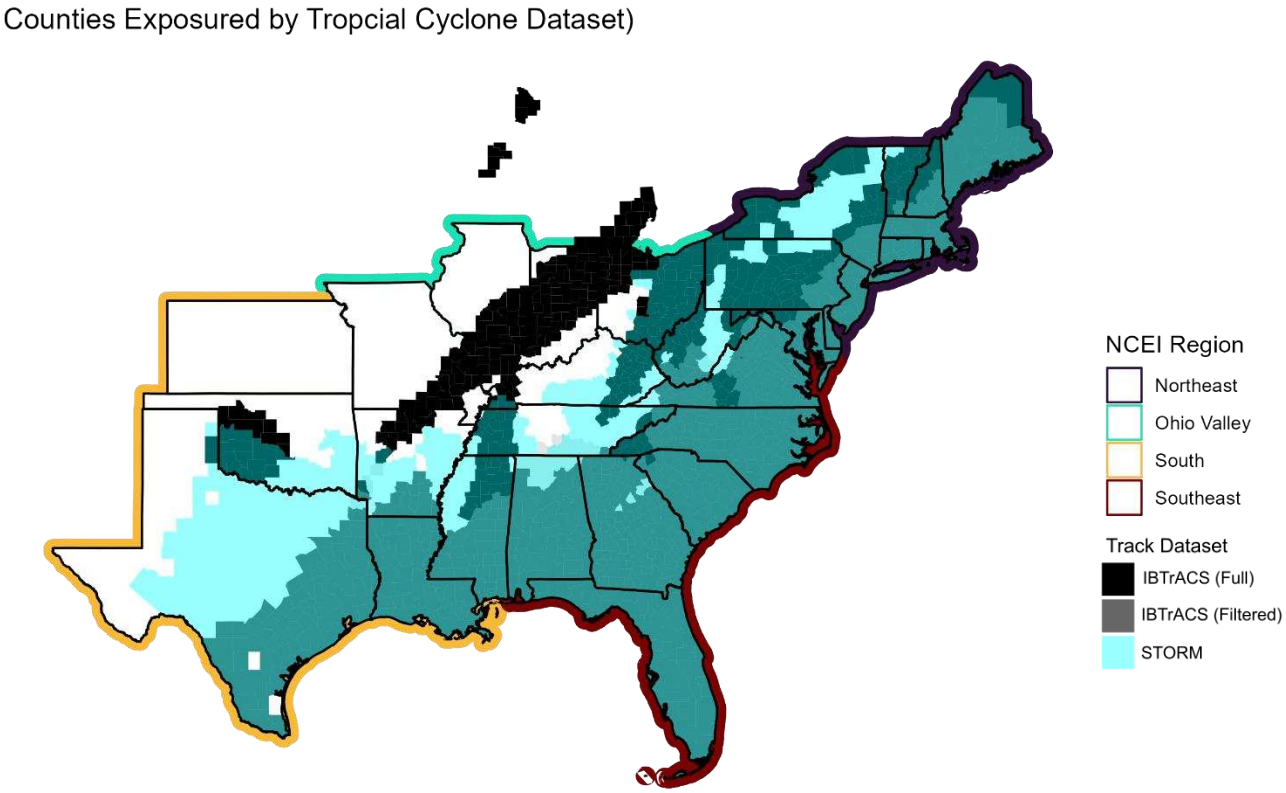

Figure S1.3: Map showing distributions of counties with at least one gale-force exposure in STORM (cyan), full 1980-2022 IBTRACS (black) and the same data filtered only to tropical storms and hurricanes (dark cyan) and 10,000 simulated seasons of STORM data (cyan)

Figure S1.4: Median Windspeeds in STORM and 1980-2022 Historical Data

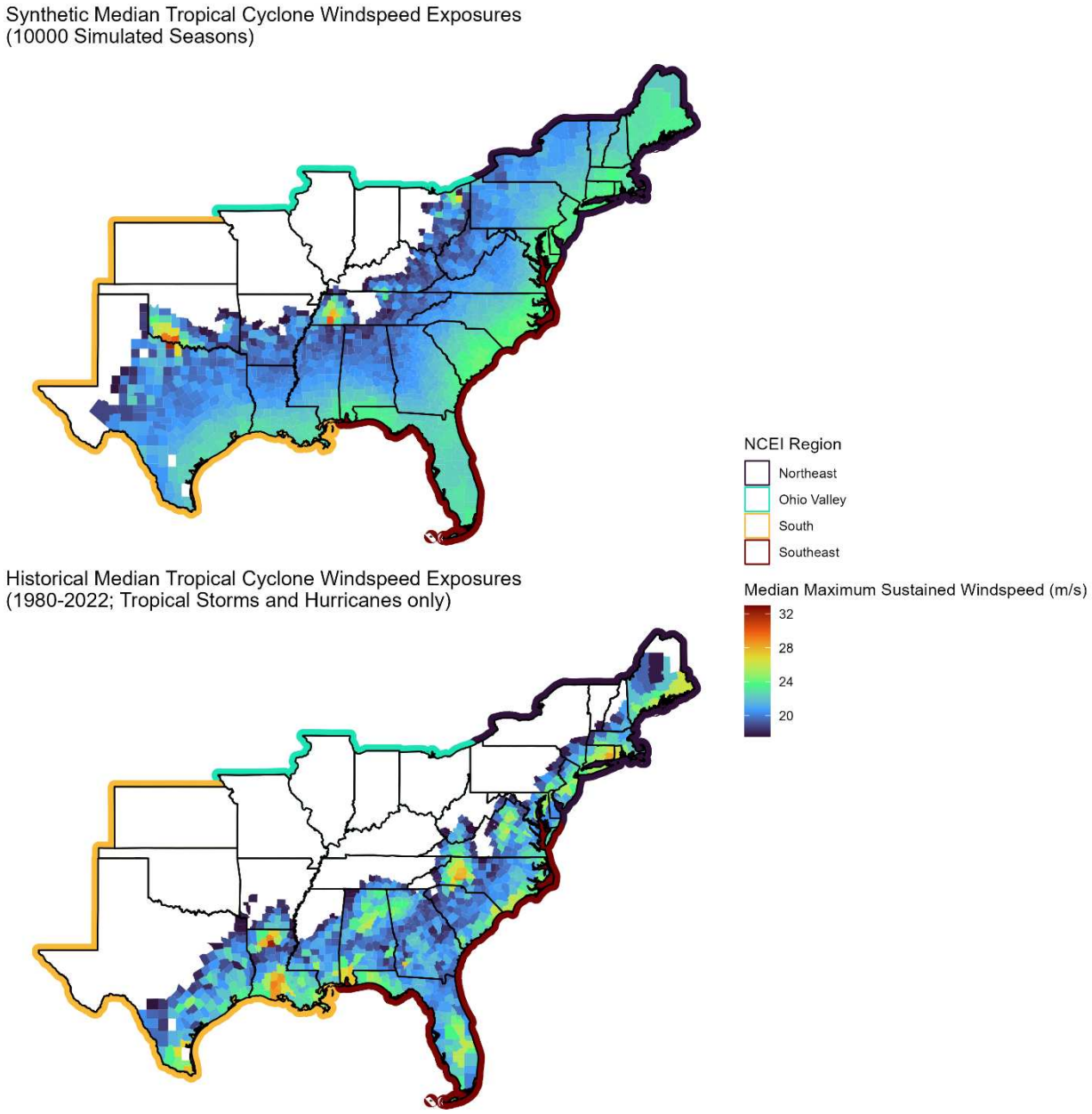

Figure S1.4: Map of median county-level maximum sustained windspeeds (m/s) under 10,000 synthetic STORM seasons (top) compared to median county-level maximum sustained windspeeds in historical IBTrACS data (1980-2022) . Shading indicates median maximum sustained windspeeds modeled at population-weighted county centers..

Figure S1.5: Tropical Cyclone Days by Month and Storm Category

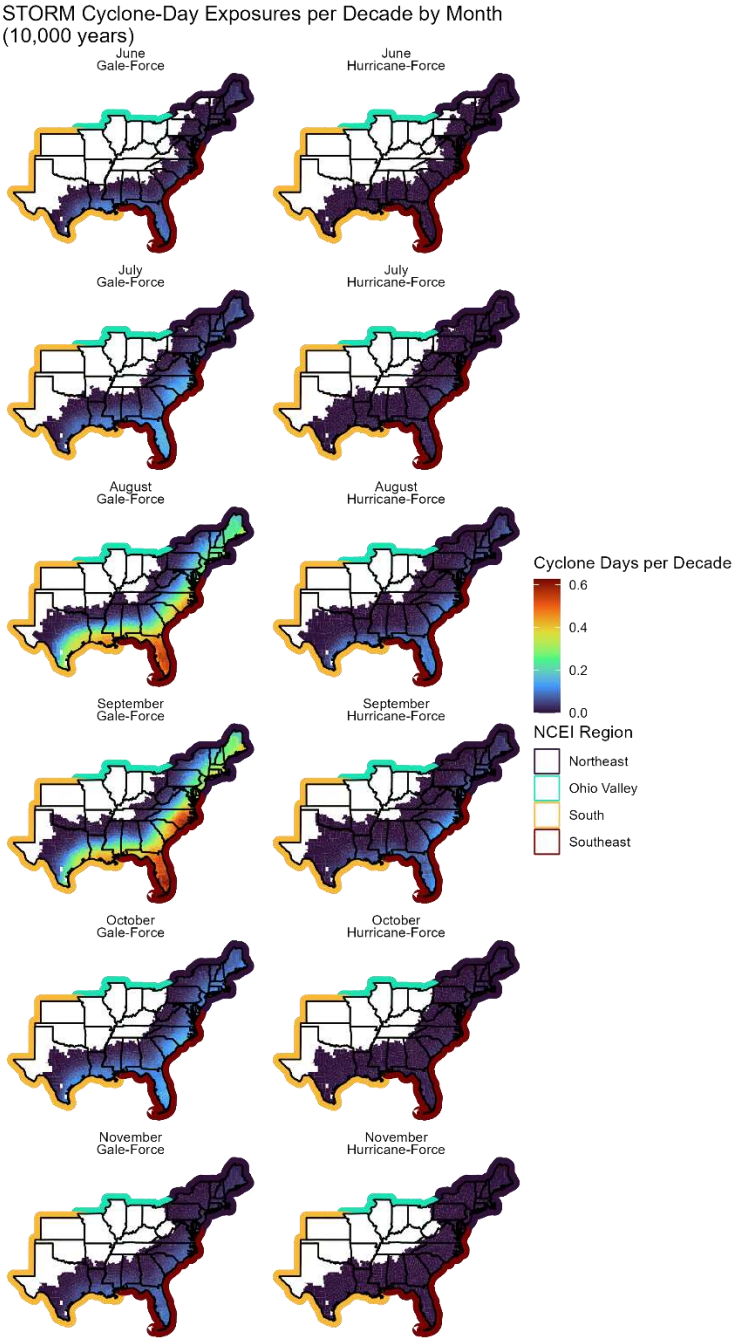

Figure S1.5: Map of cyclone-day exposures by and tropical cyclone exposure category (hurricane-force local exposures, sub-hurricane force), showing average yearly exposures using county shading (linear scale).

Table S1.1: Summary of County-Level Traits used in Exposure Response Function (Summarized by State)

State	Region	Under Poverty Line (%)	White (%)	Owner-Occupied Housing (%)	% ≥65	Median Age	Median Home Value ($)	No HS diploma (%)	Coastal (%)	Exposures (1988-2018)	Population Density
Maine	Northeast	0.1 (0.1, 0.2)	1 (0.9, 1)	0.8 (0.7, 0.8)	0.2 (0.2, 0.3)	47 (41.2, 51.4)	165300 (105412.5, 278300)	0.1 (0, 0.1)	0.5 (0, 1)	1 (0.4, 2)	47.4 (6.1, 275.8)
Delaware	Northeast	0.1 (0.1, 0.1)	0.7 (0.6, 0.8)	0.7 (0.7, 0.8)	0.2 (0.2, 0.3)	38.7 (38.1, 49.7)	266500 (228595, 269540)	0.1 (0.1, 0.1)	1 (1, 1)	3 (3, 4)	300.2 (238.7, 1236)
District of Columbia	Northeast	0.2 (0.2, 0.2)	0.4 (0.4, 0.4)	0.5 (0.5, 0.5)	0.1 (0.1, 0.1)	34.1 (34.1, 34.1)	618100 (618100, 618100)	0.1 (0.1, 0.1)	0 (0, 0)	3 (3, 3)	10284.2 (10284.2, 10284.2)
Maryland	Northeast	0.1 (0, 0.2)	0.8 (0.2, 0.9)	0.7 (0.6, 0.9)	0.2 (0.1, 0.3)	40 (35.8, 50.7)	285150 (129375, 476060)	0.1 (0.1, 0.2)	1 (0, 1)	2 (1, 5)	302.1 (49.6, 4154.2)
Massachusetts	Northeast	0.1 (0.1, 0.2)	0.8 (0.6, 0.9)	0.7 (0.5, 0.8)	0.2 (0.1, 0.3)	41 (34.2, 52.2)	400300 (217692.5, 1012227.5)	0.1 (0, 0.1)	1 (0, 1)	2 (1.3, 4)	630.6 (109.1, 9232.4)
New Hampshire	Northeast	0.1 (0.1, 0.1)	0.9 (0.9, 1)	0.8 (0.7, 0.8)	0.2 (0.2, 0.3)	44 (37.7, 52.3)	237750 (141927.5, 331377.5)	0.1 (0, 0.1)	0 (0, 0.8)	2 (1, 2)	117.4 (24.4, 456.9)
Vermont	Northeast	0.1 (0.1, 0.1)	1 (0.9, 1)	0.8 (0.7, 0.9)	0.2 (0.2, 0.3)	46.2 (37.8, 50.5)	218400 (147935, 305425)	0.1 (0.1, 0.1)	0 (0, 0)	1 (0, 1)	53.1 (18, 205.6)
Connecticut	Northeast	0.1 (0.1, 0.1)	0.8 (0.7, 0.9)	0.7 (0.7, 0.8)	0.2 (0.2, 0.2)	41 (38.1, 47.2)	255300 (216005, 408552.5)	0.1 (0.1, 0.1)	0.5 (0, 1)	3 (2.2, 3.8)	402.8 (197.7, 1436.5)
New Jersey	Northeast	0.1 (0, 0.2)	0.7 (0.5, 0.9)	0.7 (0.4, 0.9)	0.2 (0.1, 0.2)	41.8 (36.3, 48.5)	306200 (176050, 469750)	0.1 (0, 0.2)	0 (0, 1)	4 (2, 4)	1024.6 (221.9, 9535)
New York	Northeast	0.1 (0.1, 0.2)	0.9 (0.4, 1)	0.8 (0.3, 0.8)	0.2 (0.1, 0.2)	42 (33.7, 48.4)	146050 (91617.5, 651220)	0.1 (0.1, 0.2)	0 (0, 1)	0 (0, 4)	111.6 (20, 31227.8)
Pennsylvania	Northeast	0.1 (0.1, 0.2)	0.9 (0.7, 1)	0.8 (0.7, 0.8)	0.2 (0.2, 0.3)	43.6 (37.3, 49.7)	154200 (88410, 331205)	0.1 (0.1, 0.2)	0 (0, 0)	1 (0, 2.3)	141.3 (14.7, 2142.5)
Rhode Island	Northeast	0.1 (0.1, 0.1)	0.9 (0.7, 0.9)	0.7 (0.6, 0.8)	0.2 (0.2, 0.2)	44.5 (38, 45.7)	359300 (247780, 396000)	0.1 (0.1, 0.1)	1 (1, 1)	3 (2.1, 3.9)	935.7 (398.9, 1697.1)
West Virginia	Ohio Valley	0.2 (0.1, 0.3)	1 (0.9, 1)	0.8 (0.7, 0.9)	0.2 (0.2, 0.3)	44.4 (38.9, 49.7)	110600 (73355, 204425)	0.1 (0.1, 0.2)	0 (0, 0)	0 (0, 2)	48.3 (11.7, 353.8)
Tennessee	Ohio Valley	0.2 (0.1, 0.2)	0.9 (0.5, 1)	0.8 (0.6, 0.8)	0.2 (0.1, 0.3)	43 (36.6, 48)	125700 (90677.5, 201857.5)	0.2 (0.1, 0.2)	0 (0, 0)	0 (0, 0)	66 (22.5, 652.5)
Ohio	Ohio Valley	0.1 (0.1, 0.2)	0.9 (0.8, 1)	0.7 (0.6, 0.8)	0.2 (0.1, 0.2)	41.9 (33.3, 46.8)	122800 (93680, 220680)	0.1 (0.1, 0.2)	0 (0, 0)	0 (0, 1)	122.5 (32.1, 1230.9)
Kentucky	Ohio Valley	0.2 (0.2, 0.3)	1 (0.8, 1)	0.7 (0.6, 0.8)	0.2 (0.1, 0.2)	41.7 (33.6, 46.2)	91850 (54062.5, 157587.5)	0.2 (0.1, 0.3)	0 (0, 0)	0 (0, 0)	49.5 (21.4, 219.6)
Mississippi	South	0.2 (0.1, 0.4)	0.6 (0.2, 0.9)	0.7 (0.5, 0.9)	0.2 (0.1, 0.2)	39.2 (32.5, 44.7)	92250 (62035, 177720)	0.2 (0.1, 0.3)	0 (0, 1)	1 (0, 12)	42 (12.7, 266.4)
Louisiana	South	0.2 (0.1, 0.4)	0.7 (0.3, 0.9)	0.7 (0.5, 0.8)	0.2 (0.1, 0.2)	38.3 (30.6, 43.7)	124000 (74735, 226892.5)	0.2 (0.1, 0.3)	0 (0, 1)	5 (0, 10)	47.2 (10.4, 919.9)
Texas	South	0.1 (0.1, 0.3)	0.8 (0.6, 0.9)	0.7 (0.6, 0.9)	0.2 (0.1, 0.3)	39.5 (29.5, 53)	117400 (53105, 296925)	0.2 (0.1, 0.3)	0 (0, 1)	0 (0, 6.9)	28.5 (0.9, 1123.6)
Arkansas	South	0.2 (0.1, 0.3)	0.7 (0.4, 0.9)	0.7 (0.6, 0.8)	0.2 (0.1, 0.2)	40.8 (35.2, 47.8)	87700 (65350, 157150)	0.1 (0.1, 0.2)	0 (0, 0)	0 (0, 1)	24.1 (10.7, 225.9)
Oklahoma	South	0.2 (0.1, 0.2)	0.7 (0.6, 0.9)	0.7 (0.6, 0.8)	0.2 (0.1, 0.2)	40.5 (33.1, 44.6)	99700 (63000, 147250)	0.1 (0.1, 0.2)	0 (0, 0)	0 (0, 1)	22.6 (4.7, 86.1)

North Carolina	Southeast	0.2 (0.1, 0.3)	0.7 (0.4, 1)	0.7 (0.6, 0.8)	0.2 (0.1, 0.3)	42.9 (32.7, 53.7)	149400 (84645, 299510)	0.1 (0.1, 0.2)	0 (0, 1)	2 (0, 17.5)	111.3 (23, 1061.2)
South Carolina	Southeast	0.2 (0.1, 0.3)	0.6 (0.3, 0.9)	0.7 (0.6, 0.8)	0.2 (0.1, 0.3)	41.4 (36.4, 49.7)	116600 (71800, 298200)	0.1 (0.1, 0.2)	0 (0, 1)	4 (0, 13.8)	76 (25, 519.6)
Virginia	Southeast	0.1 (0, 0.3)	0.8 (0.4, 1)	0.7 (0.4, 0.9)	0.2 (0.1, 0.3)	42.9 (26.6, 54.4)	2e+05 (79970, 575500)	0.1 (0, 0.2)	0 (0, 1)	1 (0, 9)	101.8 (18.8, 5432.7)
Alabama	Southeast	0.2 (0.1, 0.3)	0.7 (0.2, 0.9)	0.7 (0.6, 0.8)	0.2 (0.1, 0.2)	41.2 (32.9, 46.7)	109600 (67265, 197690)	0.2 (0.1, 0.2)	0 (0, 0.3)	1 (0, 6)	45.9 (12.5, 370.4)
Florida	Southeast	0.1 (0.1, 0.3)	0.8 (0.6, 0.9)	0.7 (0.6, 0.8)	0.2 (0.1, 0.4)	43.2 (33.8, 57.7)	173900 (83070, 340405)	0.1 (0.1, 0.3)	1 (0, 1)	8 (4.7, 11)	180 (15, 1442.4)
Georgia	Southeast	0.2 (0.1, 0.3)	0.7 (0.3, 0.9)	0.7 (0.5, 0.8)	0.2 (0.1, 0.3)	39.6 (29.2, 52.3)	121700 (61965, 274875)	0.2 (0.1, 0.3)	0 (0, 1)	1 (0, 6)	66.5 (12.1, 1968.6)

Table S1.1: State-level summary of county model parameters (barring storm-specific features, and year of exposure, which uniformly was set to 2015). Values are summarized giving the 2.5th, 50th, and 97.5th percentiles.

Table S1.2: Region-level summary of monthly expected excess mortality

Region	Month	Gale-Force Excess Mortality (95% eCI)	Hurricane-Force Excess Mortality (95% eCI)	Net Excess Mortality (95% eCI)
Northeast	June	-23 (-64, 18)	0 (0, 0)	-23 (-64, 18)
Northeast	July	-78 (-229, 59)	18 (3, 31)	-59 (-227, 80)
Northeast	August	-332 (-996, 264)	210 (65, 329)	-133 (-898, 507)
Northeast	September	-432 (-1260, 301)	227 (60, 357)	-217 (-1173, 558)
Northeast	October	-132 (-391, 102)	33 (3, 56)	-99 (-389, 141)
Northeast	November	-29 (-77, 15)	3 (0, 6)	-25 (-77, 20)
Ohio Valley	June	0 (0, 0)	0 (0, 0)	0 (0, 0)
Ohio Valley	July	-1 (-2, 0)	0 (0, 0)	-1 (-2, 0)
Ohio Valley	August	-4 (-9, 2)	0 (0, 0)	-4 (-9, 2)
Ohio Valley	September	-6 (-13, 2)	0 (0, 0)	-6 (-13, 3)
Ohio Valley	October	-1 (-1, 0)	0 (0, 0)	-1 (-1, 0)
Ohio Valley	November	0 (0, 0)	0 (0, 0)	0 (0, 0)
South	June	-21 (-62, 22)	0 (0, 0)	-21 (-62, 22)
South	July	-25 (-73, 26)	6 (3, 10)	-19 (-70, 36)
South	August	-126 (-356, 140)	102 (57, 157)	-24 (-270, 292)
South	September	-114 (-327, 122)	51 (29, 80)	-60 (-286, 201)
South	October	-31 (-86, 32)	8 (4, 13)	-21 (-80, 45)
South	November	-16 (-47, 17)	4 (2, 6)	-12 (-44, 23)
Southeast	June	-69 (-169, 54)	0 (0, 0)	-69 (-169, 54)
Southeast	July	-164 (-434, 112)	95 (59, 142)	-69 (-403, 244)
Southeast	August	-500 (-1396, 379)	487 (313, 715)	-2 (-1205, 1035)
Southeast	September	-586 (-1600, 428)	487 (311, 720)	-85 (-1413, 1089)
Southeast	October	-150 (-382, 103)	57 (34, 89)	-85 (-365, 186)
Southeast	November	-79 (-194, 54)	25 (14, 38)	-51 (-187, 90)

Table S1.2: Region-level summary of monthly expected excess mortality (presented as 50^{th} (2.5^{th}, 97.5^{th}) percentiles from 95% empirical confidence intervals over the 10,000 STORM synthetic hurricane seasons.

SUPPLEMENTAL SECTION 2

Figure S2.1: Denver and Phoenix maps summarizing current mean exposure estimates for normalized difference vegetation index (NDVI) and urban tree canopy (UTC).

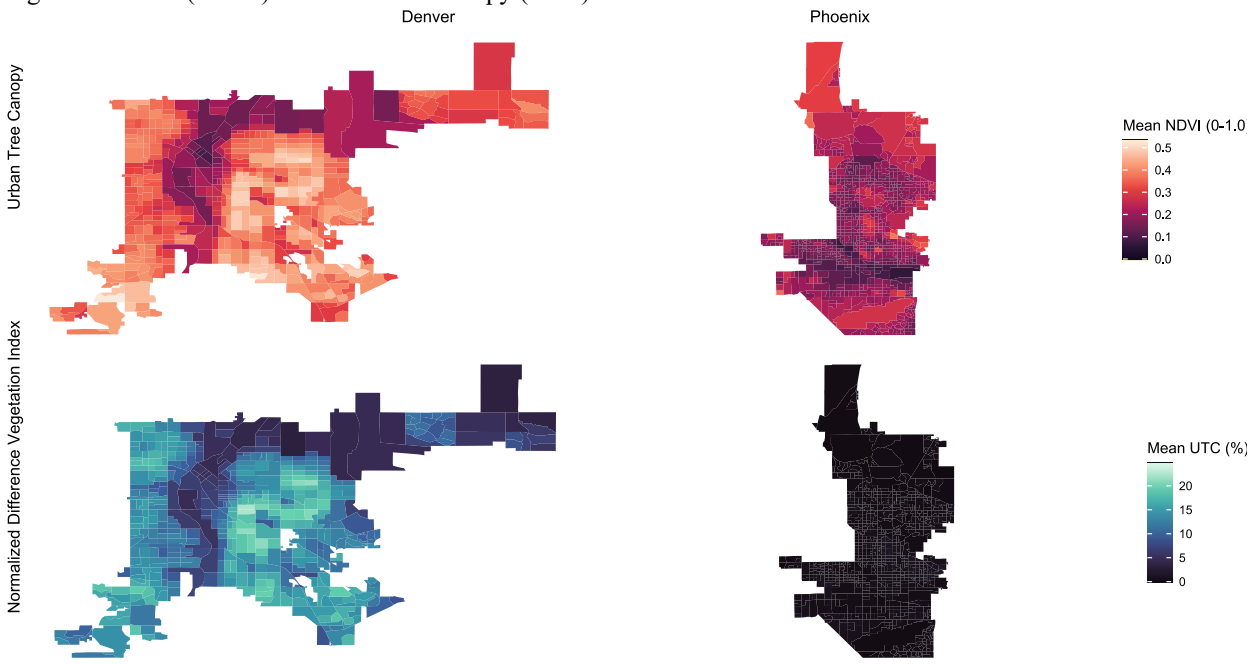

Figure S2.1: Estimated block-group level mean NDVI (A, B)) and UTC (C,D) exposures (population-weighted via National Land Cover Database "developed" status as a proxy) values by city. Exposures are calculated for the peak-NDVI month in each city (June for Denver, January for Phoenix). Color scales are consistent across cities and reflect our population-weighted estimate using (here) 500-meter buffers around all NLCD-designated "developed" land use pixels barring "developed open" categories.

Figure S2.2: Population-Centered Block Group Exposure Sampling Scheme

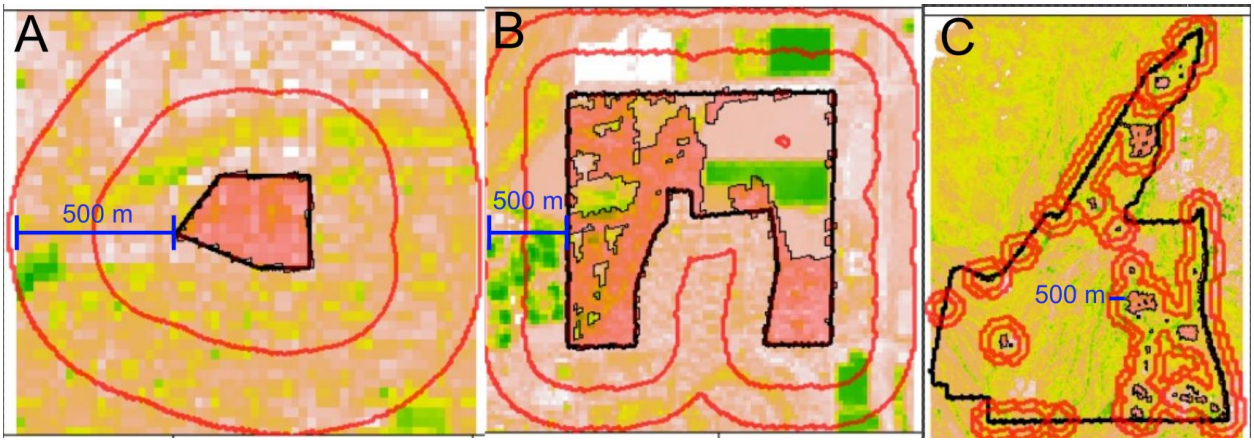

Figure S2.2: *Visual representation of spatial analysis used for estimating development centered NDVI and UTC exposures at the block group level; three Phoenix block groups of increasing size (*A *through* C*) against the NDVI data layer, with 500-meter scale indicated in blue. Black lines indicate block group borders, shaded regions indicate "developed" pixels (essentially impervious cover of >20%), and concentric bands indicate 250m and 500m buffers. Mean values of pixels in each buffer area were collected and assigned to the block group. In extremely small, fractional and/or densely populated block groups like* A*, the developed regions are synonymous with block group boundaries. Still, with increasing size, the distribution of developed areas differs more from boundaries (*B*). For large, sparsely settled (e.g.* C*) block groups, this becomes especially apparent. However, non-residential features (e.g. roadways, industrial sites) tend to become a potential source of false positives event when filtering for <3-pixel clusters.*

Figure S2.3: Predicted vs Observed Block Group Normalized Difference Vegetation Index (NDVI) Exposures by City and Buffer Size:

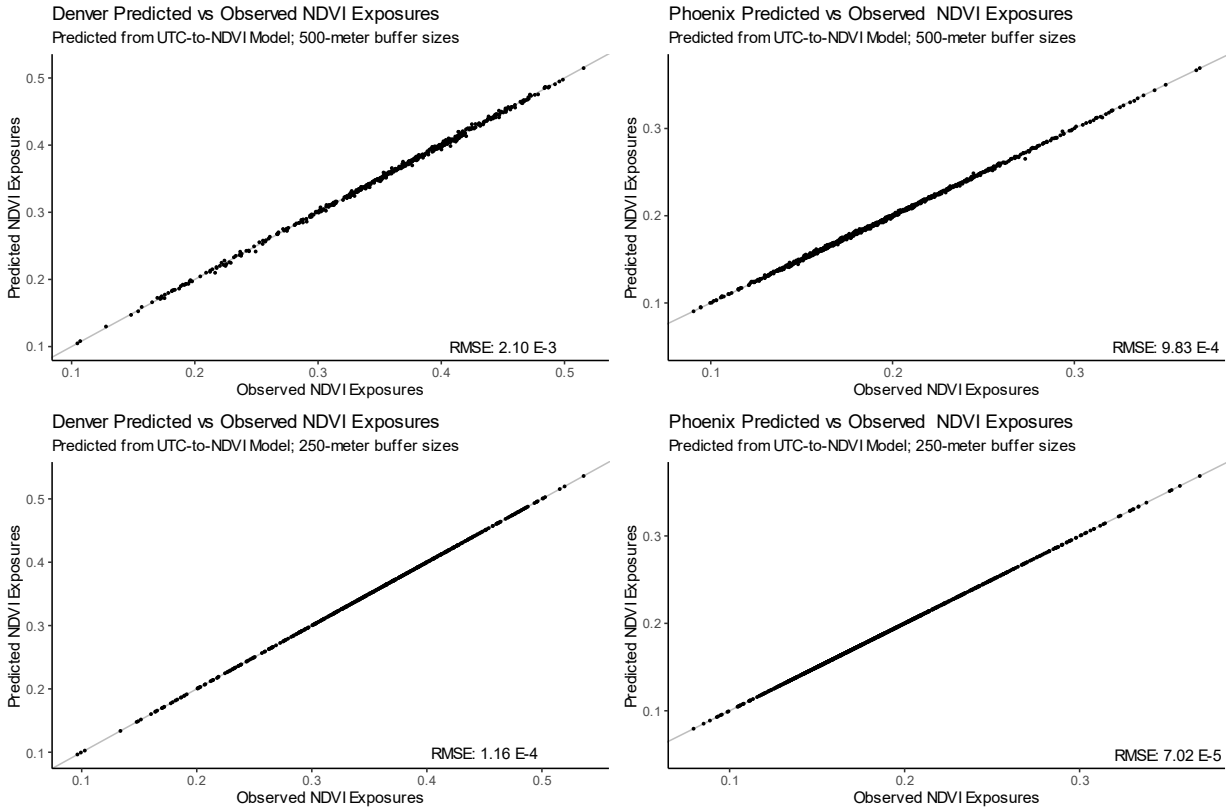

Figure S2.3: Comparisons of predicted vs observed calculation population weighted NDVI exposures (block group level) by city (columns) and exposure buffer size (rows). Each model's root Mean Square Error values are displayed in the lower right. The X-axis represents observed values at the block group level, and the Y-axis the predicted counterparts, with black points representing block group-level values, and a 1:1 line (grey) added for reference.

Figure S2.4 NDVI and Canopy Exposures by Social Vulnerability Index Quartiles

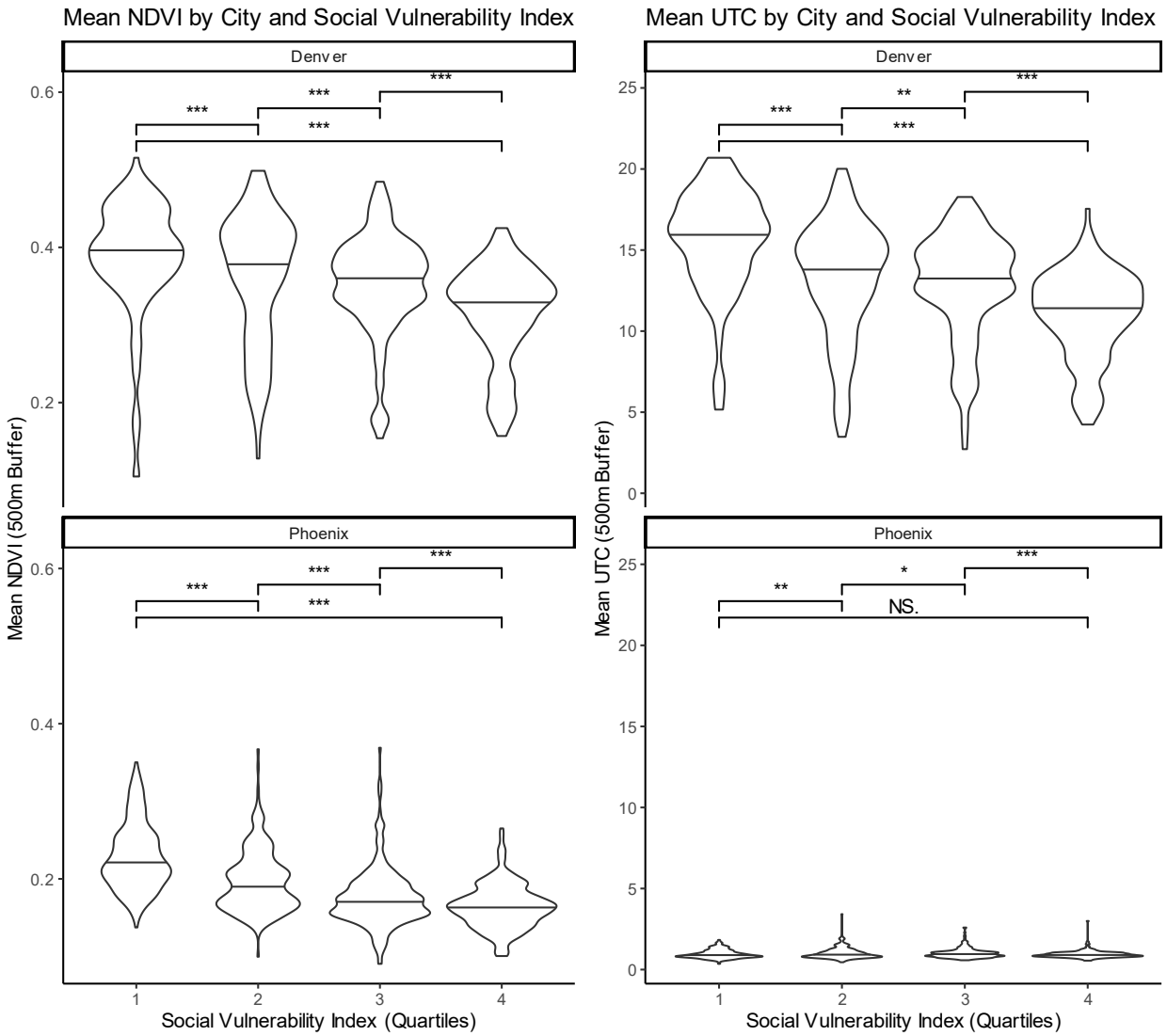

Figure S2.4: Violin plot of mean NDVI (left) and UTC (right) by city and Social Vulnerability Index (SVI) quartile. Stars indicate significance in pairwise differences, based on the Wilcoxon Rank Test ('NS' = p-value >0.05, '' = p-value <0.05, >0.01], '**' = p-value <0.01, >0.001, '***' = p-value <0.001)). Horizontal lines indicate the 50th quantile in exposure values, and the width of the violin plot corresponds to the relative density.*

Figure S2.5 Estimated Annual Averted Attributable Health Outcome Cases by Social Vulnerability Index Quartile, City and Canopy Goal

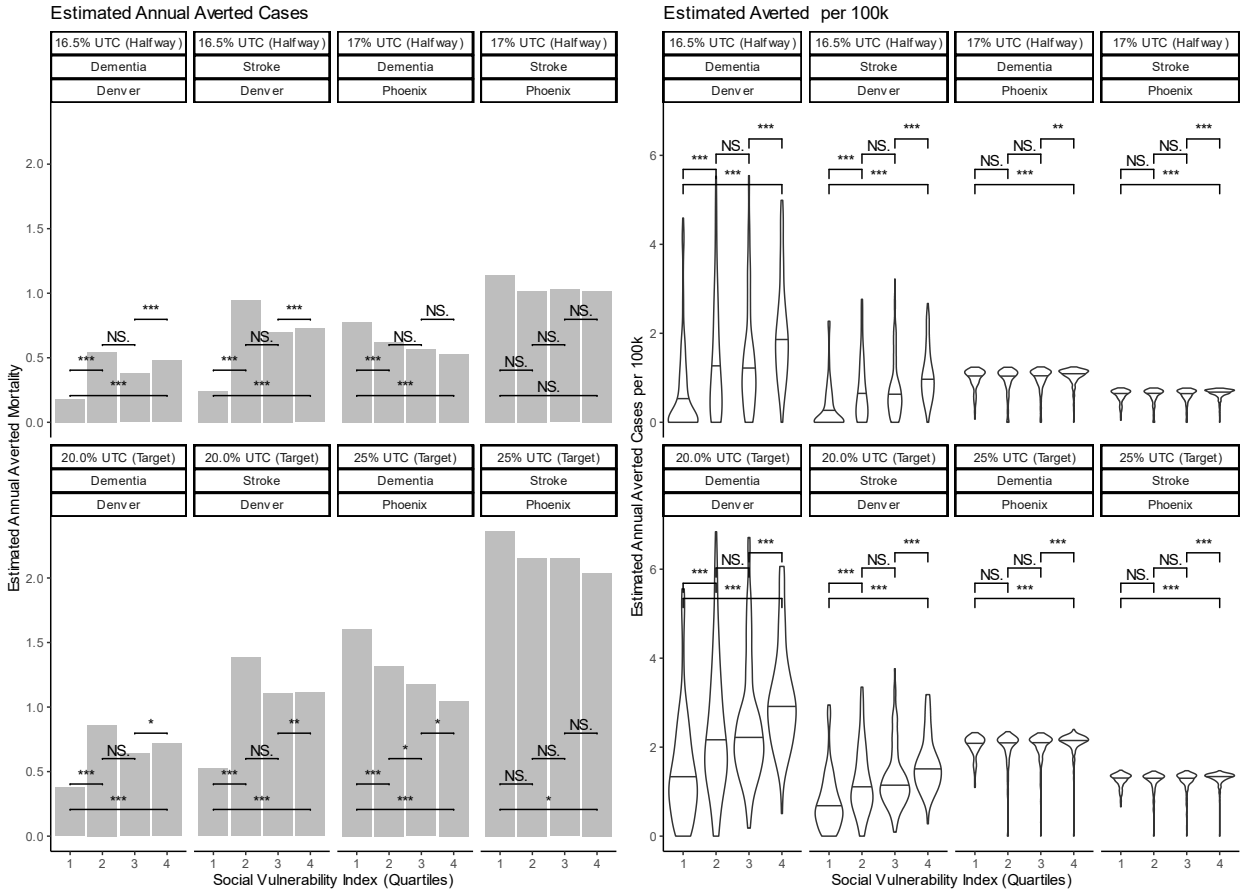

Figure S2.5: Summary of absolute (left) and proportional (right) neurological health outcomes by city for stroke and dementia. S*ignificance levels from two-sided Wilcoxon Rank tests between indicated pairs of SVI quartiles ('NS' = p-value >0.05, '*' = p-value <0.05, >0.01], '**' = p-value <0.01, >0.001, '***' = p-value <0.001). Right: Social Vulnerability Quartile Estimated Attributable Mortality per 100k Residents by City and Canopy Goal; significance levels also from two-sided Wilcoxon Signed Rank Test with the same nomenclature.*

Table S2.1: Denver Characteristics and Health Outcome Summaries by Urban Tree Canopy Scenarios Assuming an 11% UTC Baseline. (all values displayed at nearest three significant digits)

City	Denver	Denver
Urban Tree Canopy (UTC) Goal	**15.5% [Halfway Goal]**	**20% [2050 Goal]**
Urban population 18+ (Mortality risk group)	521,000	
Annual Deaths Averted in at-risk (95% Uncertainty Interval) [UI]	165 (80.1, 255)	276 (136, 423)
Deaths Averted per 100k at-risk (95% UI)	26.3 (12.8, 40.3)	47.3 (23.2, 72.3)
Estimated Economic Impact (USD; 95% UI)	1.61e+09 (7.77e+08, 2.47e+09)	2.68e+09 (1.33e+09, 4.11e+09)
Population 35-84 (Stroke risk group)	298,000	298,000
Annual Strokes Averted in at-risk (95% UI)	3.4 (1.8, 5.5)	5.8 (3.1, 9.4)
Strokes Averted per 100k at-risk (95% UI)	1 (0.5, 1.6)	1.8 (0.9, 2.9)
Estimated Economic Impact (USD; 95% UI)	41,200 (21,600, 66,300)	69,600 (36,900, 113,000)
Population 55-84 (Dementia risk group)	104,000	104,000
Annual dementia cases averted in at-risk (95% UI)	2.1 (1.2, 3.3)	3.7 (2.1, 5.8)
Dementia Averted per 100k at-risk (95% UI)	1.9 (1.1, 3)	3.4 (1.9, 5.3)
Estimated Economic Impact (USD; 95% UI)	1e+05 (56,500, 155,000)	176,000 (99,500, 276,000)

In addition to the 2014 13% UTC baseline cited in Denver's One Healthy City planning document, we used an 11% baseline from Denver Forestry Department's 2020 canopy survey (rounding to the nearest whole number), using an area-weighted mean of reported neighborhood canopy percentages. The "halfway" UTC scenario dropped to 15.5% from 16.5% at the 2014 baseline.

SUPPLEMENTAL SECTION 3:

1: Project 3 Supplemental Methods

S3.1 Distributed Lag Nonlinear Model and Risk Ratio Estimation

To estimate temperature-associated risk ratios for each city, we adapted an existing workflow (Vicedo-Cabrera et al. 2019) developed for projecting public health implications of climate change factors like warming (Vicedo-Cabrera et al. 2019). This method is based on a distributed lag non-linear model (DLNM), which accounts for both short term trends on the scale of days, and long-term seasonal trends, using a bidimensional DLNM functional space to enable interaction between a risk predictor (here, temperature) and "lag time" since exposure by allowing distinct constituent functions for a given time/exposure combination best to represent the outcome association for a given combination (Gasparrini 2011). The dlnm framework accommodates either a "forward" or "backwards" representation of risk over time; we use the former, which e.g. deals with the cumulative risk experienced over a time period of interest as a result of an exposure at a certain time, as opposed to the latter, which essentially "decomposes" a known risk at a timepoint of interest into the contributions arising from a constant exposure level over previous times (Gasparrini 2011). The DLNM initially contained 20 coefficients, yielding a 20x20 covariance matrix to represent the functional space. Still, for projection purposes, the *crossreduce* function was used to compress each component to a one-dimensional summary, leaving 4 unitless coefficients and a 4x4 version of the matrix. Together with the outcome, this creates a "three-dimensional" representation of risk outputs as a function of time, an exposure of interest, and cumulative lagged effects; Figure S3.2 shows an example of this using a visual representation of Chicago's three-dimensional exposure-response function space (consisting of temperature exposures, lag time, and the resulting risk ratio) alongside the compressed two-dimensional representation. Where Vicedo-Cabrera chose to use downscaling and calibration (temperature distribution modifications intended to help reconcile different time and spatial scales in input data) to harmonize between observed and projected temperature datasets, this is noted to have largely unknown implications for apparent trends, so we opted against downstream modifications of the estimates (Vicedo-Cabrera et al. 2019).

For the DLNM function, we used natural cubic splines, rather than e.g. linear regression, to allow the flexibility needed to address the nonlinear and often non-monotonic mortality responses generally observed for temperature stressors across the NMMAPS cities communities based on local climate and adaptational factors (e.g. infrastructure, local behavior) (Cheng et al. 2019). These output functions are centered on a dynamically-determined

“minimum mortality temperature (MMT)—the temperature associated with lowest mortality risk after adjusting for e.g., seasonal trends. To determine the relative risk (RR), the MMT is treated as an RR of 1.0, with temperatures above or below this “optimum” level generally leading to elevated mortality risks, although in some cases low model certainty (arising from e.g., rare temperatures or few deaths in the population of interest) or extrapolation beyond the 1987-2005 NMMAPS conditions can lead to below-MMT levels. The NMMAPS dataset, available as mentioned as county-level approximations of cities, provides all the information we need to develop DLNMs for each city—namely the exposure (recorded temperatures from 1987-2005), outcome (all-cause mortality estimates, divided by age groups), and lag (operationally, the continuous time lengths between temperatures and outcomes). We used a 21-day lag component for this component of the model. While DLNM models generate exposure-response functions with three dimensions incorporating in this case temperature, lag (i.e. time since exposure), and the resulting relative risk of mortality, these were “collapsed” into a summarized two-dimensional form (for temperature and relative risk, with lag time being aggregated across the exposure period in the previous step) to determine a single MMT, and for visualization.

Figure S3.1: Temperature Distribution Comparisons for Nominal NMMAPS and ARISE Cities.

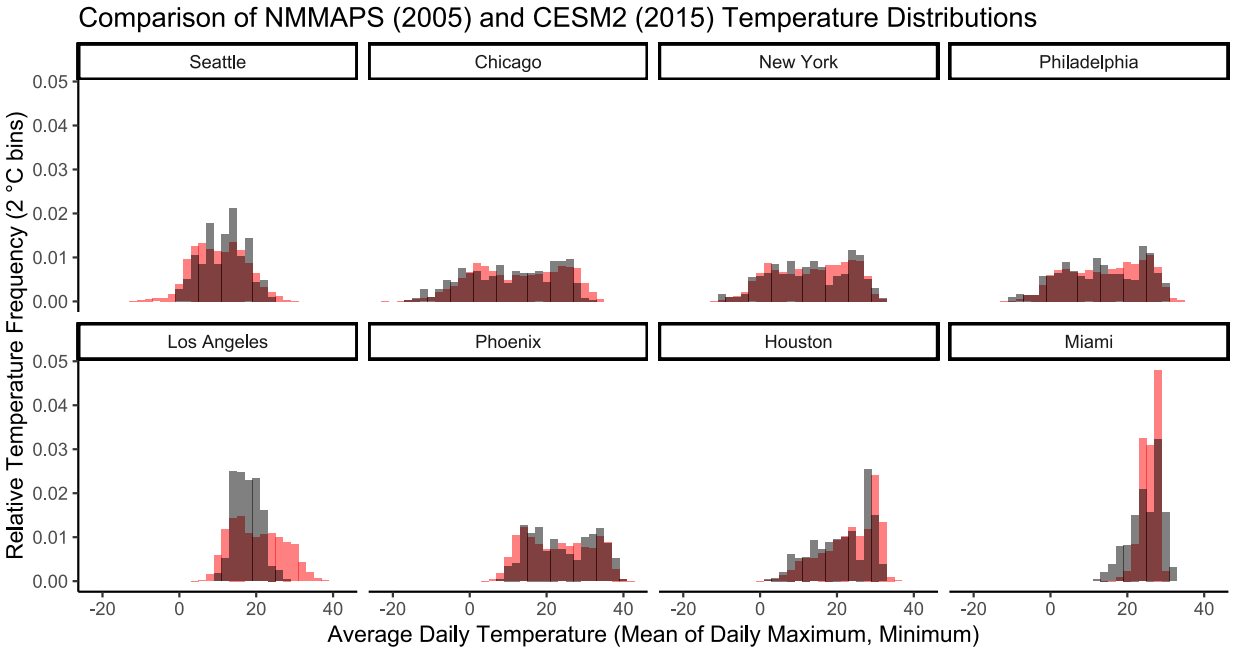

Figure S3.1: Histograms for distributions of temperatures (bin width of 2°C) comparing historical NMMAPS conditions in 2005 (grey) with nominally corresponding 2015 CESM2 temperatures (red). This plot intends to subjectively assess the degree of agreement between the nominal NMMAPS cities and nearest tile in the ARISE dataset.

Figure S3.2: Example Distributed Lag Nonlinear Model Function in 3D and 2D Representations

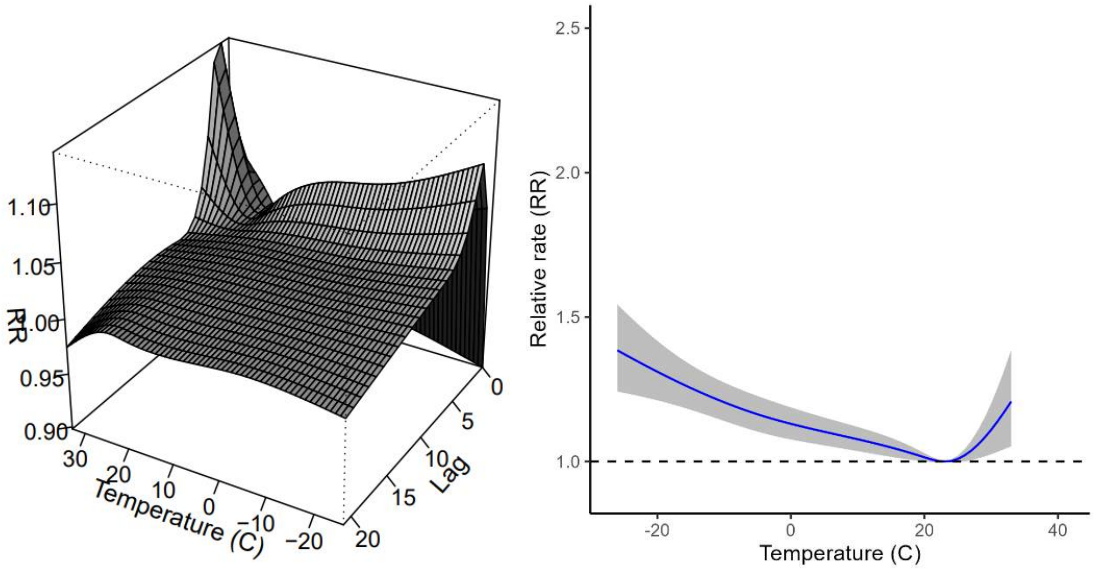

Figure S3.2: Left: Three-dimensional representation of the distributed lag nonlinear model (DLNM) output for the risk ratio of all-cause mortality in Chicago (all age groups) as a function of temperature (Celsius), lag time since exposure (days), and the corresponding relative rate of all-cause mortality (excess deaths per 100,000 residents). Right: Equivalent two-dimensional representation of relative rate as a function of temperature after compressing the lag dimension.

Figure S3.3: Exposure-Response Functions for Last 10 Years (1995-2005) of Historical Data and Central 99.8% of Temperatures:

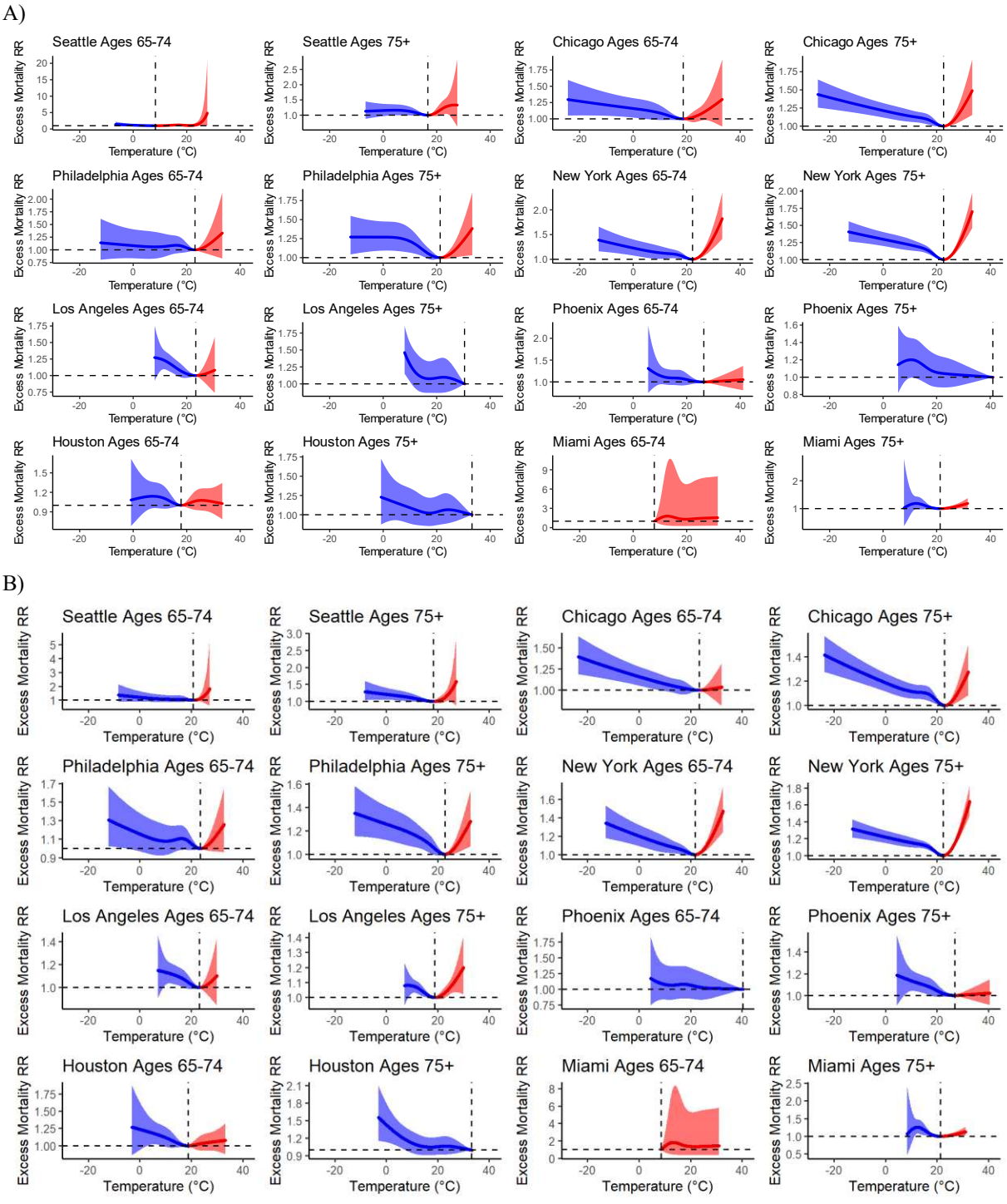

Figure S3.3: Alternate exposure-response functions generated using alternate exposure data: (A) 1995-2005 subset of NMMAPS observations and (B): Excluding the most-extreme 0.1% of temperatures in each city. These were informed by findings by Bobb et al (2014) of significant downward trends in heat-attributable mortality over time, and the presence of extreme outlier temperatures in cities like Miami, respectively.

Figure S3.4: Mortality (2050-2060 Totals) by Scenario and Temperature Category:

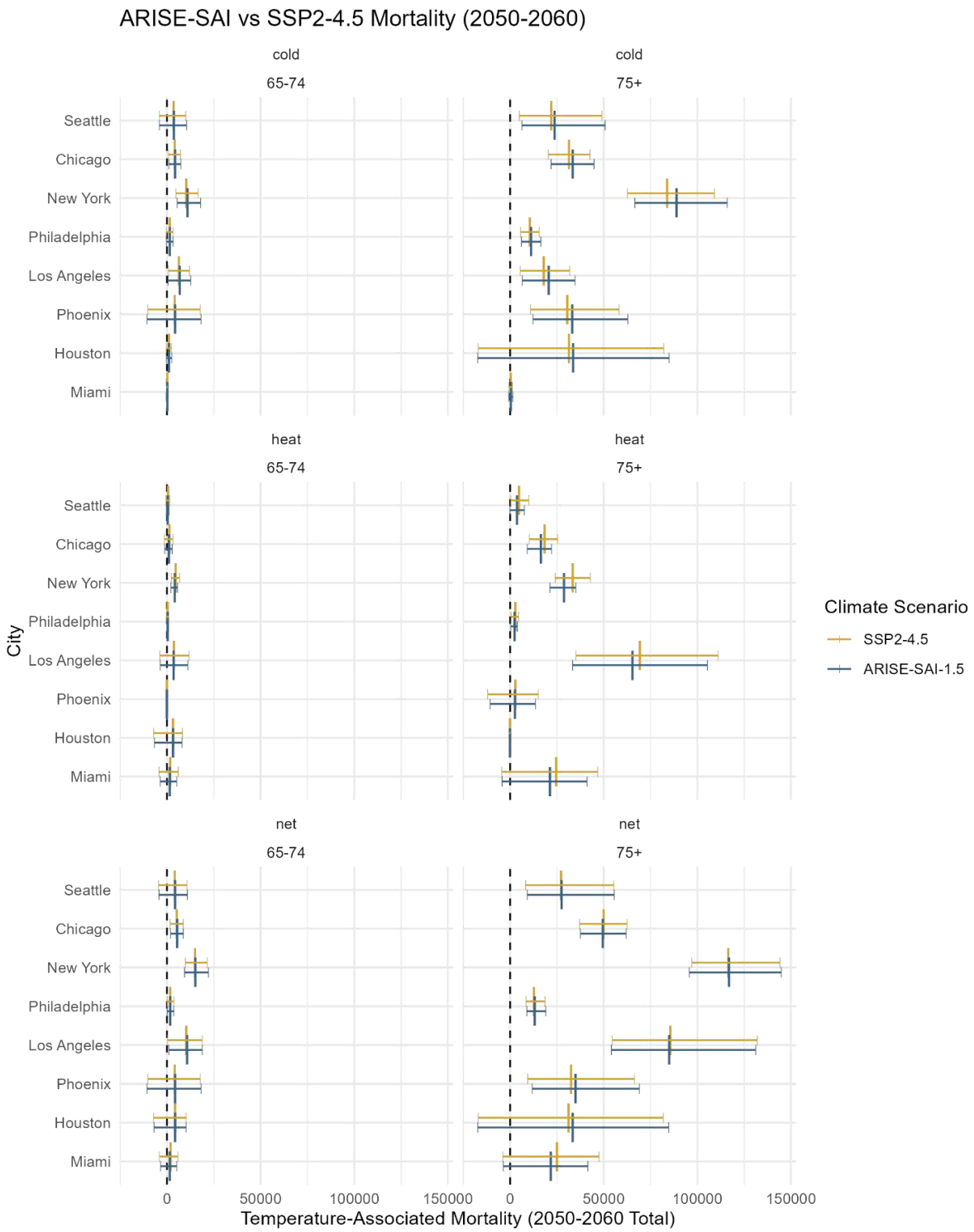

Figure S3.4: Comparison of Modeled Temperature-Associated Mortality (2050-2060 totals). We present mortality counts for SSP2.4-5 (gold) and corresponding ARISE-SAI climate scenarios (blue), and predictions vertically by temperature category (cold-associated, heat-associated, net temperature-associated) in panes, and by city within panes, with age groups (65-74, 75+) arranged in side-by-side panes.

Figure S3.5: 2050-2060 Mortality by Scenario and Temperature Category (1995-2005 Exposures)

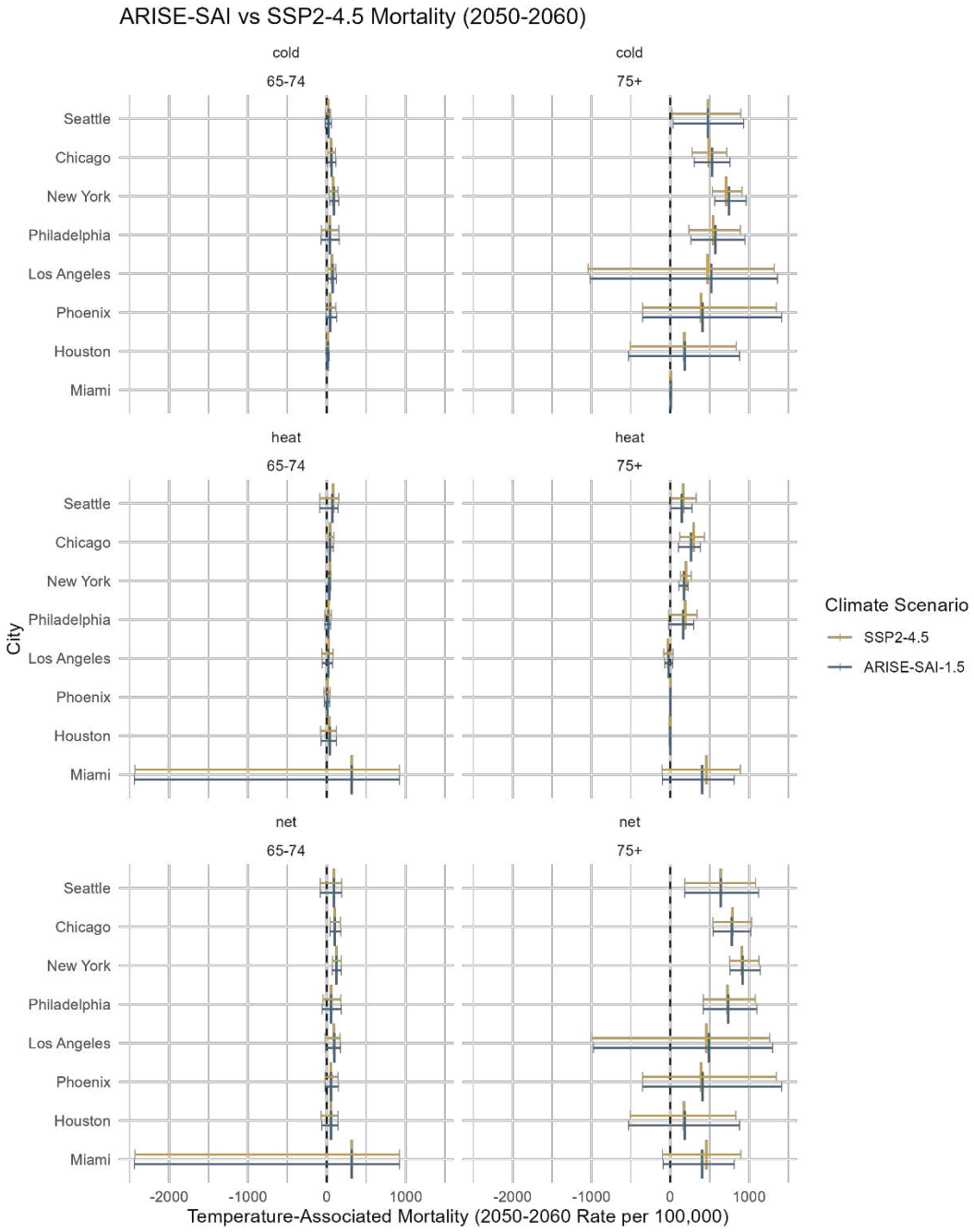

Figure S3.5: Comparison of Modeled Temperature-Associated Mortality (2050-2060 totals) based on 1995-2005 historical exposures. We present mortality counts for SSP2.4-5 (gold) and corresponding ARISE-SAI climate scenarios (blue), and predictions arranged vertically by temperature category (cold-associated, heat-associated, net temperature-associated) in panes, and by city within panes, with age groups (65-74, 75+) arranged in side-by-side panes.

Figure S3.6: 2050-2060 Mortality by Scenario and Temperature Category (Central 99.8% of Temperatures)

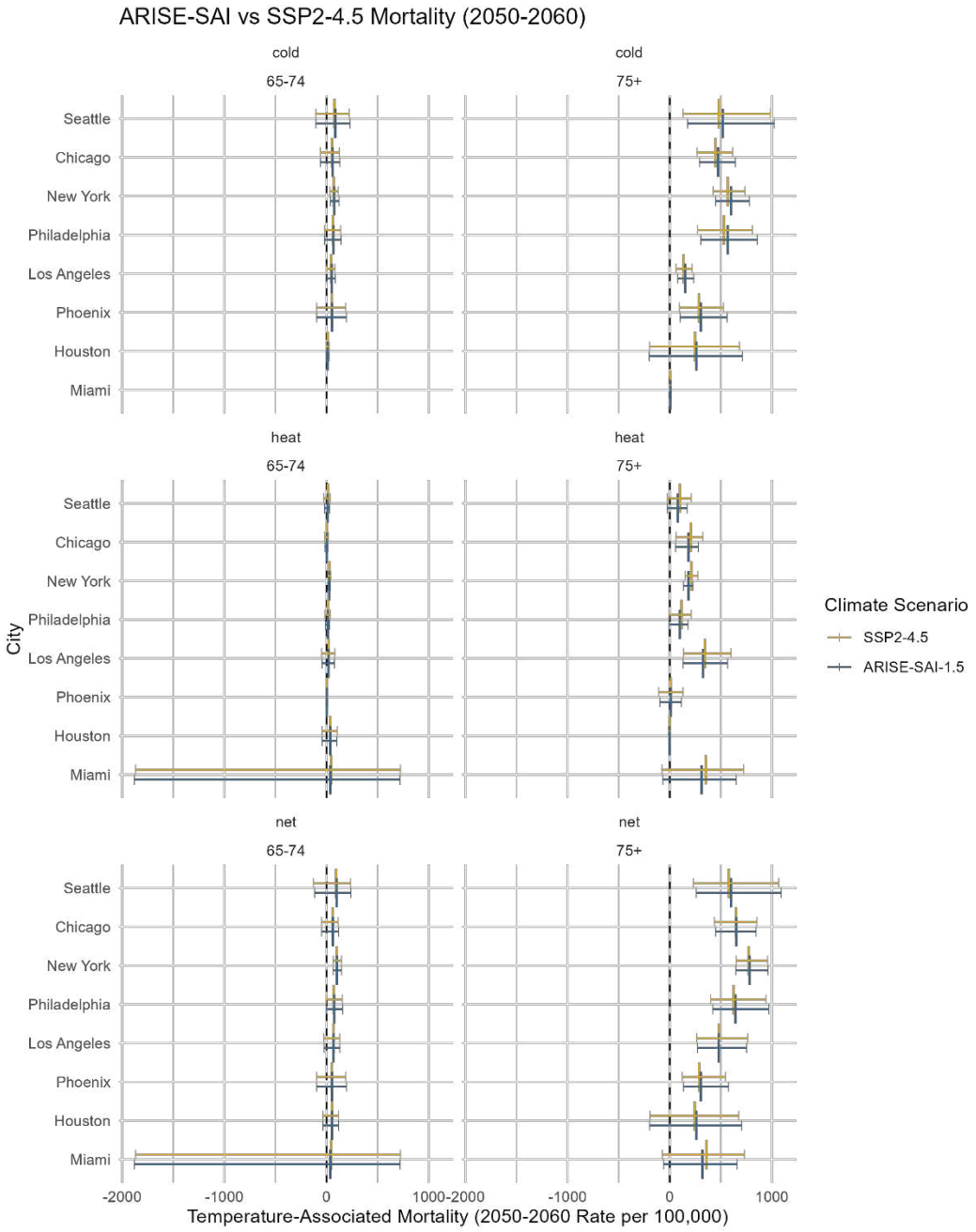

Figure S3.6: Comparison of Modeled Temperature-Associated Mortality (2050-2060 totals) based on exposures excluding the hottest and coldest 0.1% for each city. We present mortality counts for SSP2.4-5 (gold) and corresponding ARISE-SAI climate scenarios (blue), and predictions arranged vertically by temperature category (cold-associated, heat-associated, net temperature-associated) in panes, and by city within panes, with age groups (65-74, 75+) arranged in side-by-side panes.

Table S3.1: Summary of City populations by age group, Minimum mortality temperatures (MMT), and baseline attributable death rates.

City	Age Group	Scenario	Mean Temperature (°C)	MMT (°C)	Population (Mean)	Temperature-Attributable Deaths/Day (mean)
Chicago	Ages 65-74	SSP2-4.5	14.51444	25.4	591949.7	1.839907
Chicago	Ages 65-74	ARISE-SAI-1.5	13.5707	25.4	591949.7	1.836932
Chicago	Ages 75+	SSP2-4.5	14.51444	25.7	600605	14.14324
Chicago	Ages 75+	ARISE-SAI-1.5	13.5707	25.7	600605	13.67635
Houston	Ages 65-74	SSP2-4.5	24.82751	21.1	843527.1	0.833391
Houston	Ages 65-74	ARISE-SAI-1.5	24.40421	21.1	843527.1	0.847275
Houston	Ages 75+	SSP2-4.5	24.82751	29.7	945734	4.028705
Houston	Ages 75+	ARISE-SAI-1.5	24.40421	29.7	945734	4.152891
Los Angeles	Ages 65-74	SSP2-4.5	21.0394	25.2	1375039	3.978053
Los Angeles	Ages 65-74	ARISE-SAI-1.5	20.57075	25.2	1375039	3.911398
Los Angeles	Ages 75+	SSP2-4.5	21.0394	18.7	1768293	28.08763
Los Angeles	Ages 75+	ARISE-SAI-1.5	20.57075	18.7	1768293	27.33627
Miami	Ages 65-74	SSP2-4.5	27.23774	25.6	533003.1	0.26641
Miami	Ages 65-74	ARISE-SAI-1.5	26.74707	25.6	533004.3	0.186072
Miami	Ages 75+	SSP2-4.5	27.23774	21.5	683221	5.202553
Miami	Ages 75+	ARISE-SAI-1.5	26.74707	21.5	683222.4	4.169612
New York	Ages 65-74	SSP2-4.5	15.29886	20.8	1375698	3.17605
New York	Ages 65-74	ARISE-SAI-1.5	14.43684	20.8	1375698	3.21746
New York	Ages 75+	SSP2-4.5	15.29886	24.4	1415918	20.17923
New York	Ages 75+	ARISE-SAI-1.5	14.43684	24.4	1415918	20.21582
Philadelphia	Ages 65-74	SSP2-4.5	16.3785	26.4	210730.4	0.62962
Philadelphia	Ages 65-74	ARISE-SAI-1.5	15.51945	26.4	210730.4	0.631206
Philadelphia	Ages 75+	SSP2-4.5	16.3785	24.5	174174.4	2.825279

Philadelphia	Ages 75+	ARISE-SAI-1.5	15.51945	24.5	174174.4	2.902763
Phoenix	Ages 65-74	SSP2-4.5	24.01088	34.6	823964.6	0.922433
Phoenix	Ages 65-74	ARISE-SAI-1.5	23.29752	34.6	823964.6	0.963618
Phoenix	Ages 75+	SSP2-4.5	24.01088	26.4	1091898	7.678275
Phoenix	Ages 75+	ARISE-SAI-1.5	23.29752	26.4	1091898	8.213378
Seattle	Ages 65-74	SSP2-4.5	12.18888	-10	409755.8	2.012601
Seattle	Ages 65-74	ARISE-SAI-1.5	11.50578	-10	409755.4	1.992493
Seattle	Ages 75+	SSP2-4.5	12.18888	17	437364.4	5.218499
Seattle	Ages 75+	ARISE-SAI-1.5	11.50578	17	437364.3	5.229288

Table S3.1: Summary of City populations by age group, Minimum mortality temperatures (MMT), and baseline attributable death rates.

Table S3.2: Summary of Temperature-Attributable All-Cause Mortality

City	Age Group	Average Population (2050-2060)	Temperature Type	SSP2-4.5 (95% eCI)	ARISE-SAI 1.5 (95% eCI)
Seattle	Ages 65-74	409,756	net	4,156 (-4,489, 10,631)	**4,370 (-4,223, 10,830)**
			heat	**667 (-569, 1,489)**	499 (-383, 1,119)
			cold	3,476 (-4,130, 10,048)	**3,785 (-4,027, 10,477)**
	Ages 75+	437,364	net	27,313 (8,253, 55,329)	**27,575 (9,226, 55,731)**
			heat	**4,758 (353, 9,968)**	3,708 (-89, 7,667)
			cold	22,002 (4,942, 49,053)	**23,686 (6,204, 50,890)**
Chicago	Ages 65-74	591,950	net	5,218 (1,615, 8,750)	**5,376 (1,900, 8,738)**
			heat	**1,293 (-1,356, 3,170)**	1,149 (-1,168, 2,770)
			cold	4,078 (874, 7,088)	**4,429 (1,073, 7,513)**
	Ages 75+	600,605	net	**49,932 (37,211, 62,488)**	49,518 (37,594, 61,952)
			heat	**18,431 (10,182, 25,517)**	16,370 (9,160, 22,162)
			cold	31,462 (20,293, 42,532)	**33,525 (21,839, 44,840)**
New York	Ages 65-74	1,375,698	net	14,892 (9,742, 21,542)	**15,048 (9,435, 22,074)**
			heat	**4,683 (2,399, 6,650)**	4,067 (2,042, 5,681)
			cold	10,251 (4,819, 16,739)	**11,064 (5,471, 18,085)**
	Ages 75+	1,415,918	net	116,450 (96,952, 144,095)	**116,942 (95,671, 144,747)**
			heat	**33,464 (24,201, 42,842)**	28,814 (21,319, 35,243)
			cold	83,846 (62,624, 109,090)	**88,958 (66,549, 115,995)**
Philadelphia	Ages 65-74	210,730	net	1,732 (78, 3,691)	**1,792 (33, 3,770)**
			heat	**422 (-192, 900)**	360 (-170, 749)
			cold	1,482 (-335, 3,175)	**1,575 (-329, 3,322)**
	Ages 75+	174,174	net	12,740 (8,584, 18,696)	**13,048 (8,853, 19,120)**
			heat	**2,766 (473, 4,586)**	2,342 (362, 3,840)
			cold	10,401 (5,547, 15,634)	**11,086 (6,151, 16,524)**
Los Angeles	Ages 65-74	1,375,039	net	10,404 (459, 18,924)	**10,680 (1,128, 18,897)**
			heat	**3,651 (-3,754, 11,824)**	3,457 (-3,514, 11,208)
			cold	6,346 (754, 12,005)	**6,737 (521, 12,758)**
	Ages 75+	1,768,293	net	**85,724 (54,599, 132,006)**	84,948 (54,110, 131,089)
			heat	**69,403 (35,102, 111,164)**	65,425 (33,493, 105,463)
			cold	17,908 (5,415, 31,916)	**20,680 (6,459, 34,675)**
Phoenix	Ages 65-74	823,965	net	4,147 (-10,265, 17,751)	**4,384 (-10,586, 18,261)**
			heat	0 (-2, 2)	0 (0, 0)
			cold	4,147 (-10,266, 17,753)	**4,385 (-10,586, 18,261)**
	Ages 75+	1,091,898	net	32,587 (9,296, 66,346)	**34,963 (11,764, 69,135)**
			heat	**2,813 (-11,940, 15,062)**	2,537 (-10,779, 13,484)
			cold	30,592 (10,980, 58,254)	**33,290 (12,333, 62,896)**
Houston	Ages 65-74	843,527	net	4,353 (-7,158, 10,190)	**4,419 (-6,925, 10,270)**
			heat	**3,288 (-7,112, 8,220)**	3,184 (-6,648, 8,051)
			cold	990 (-344, 2,286)	**1,105 (-359, 2,506)**

	Ages 75+	945,734	net	31,235 (-16,996, 81,838)	**33,422 (-17,247, 84,661)**
			heat	-138 (-526, 134)	**-94 (-366, 88)**
			cold	31,465 (-17,138, 82,179)	**33,534 (-17,318, 85,018)**
Miami	Ages 65-74	533,004	net	**1,858 (-4,038, 5,862)**	1,604 (-3,415, 5,102)
			heat	**1,781 (-4,261, 6,081)**	1,457 (-3,705, 5,231)
			cold	80 (-220, 335)	**100 (-252, 383)**
	Ages 75+	683,222	net	**24,994 (-3,846, 47,524)**	21,816 (-3,557, 41,540)
			heat	**24,500 (-4,437, 46,900)**	21,317 (-4,179, 41,180)
			cold	392 (-649, 1,242)	**478 (-736, 1,347)**

Table 3.2: Summary Table of modeled total 2050-2060 heat-associated, cold-associated, and overall temperature-associated mortality counts under SSP2-4.5 (left) and ARISE-SAI (right), by city, and age group. We provide populations by age group for context, and have bolded the scenario leading to the higher estimated mortality to facilitate comparison.

LIST OF ABBREVIATIONS AND UNITS

HIA: Health Impact Assessment
quantitative health impact assessment: Quantitative Health Impact Assessment
exposure-response function: Exposure-Response Function
TC: Tropical Cyclone
NCEI: National Centers for Environmental Information
RR: Risk Ratio
CI: Confidence Interval
eCI: Empirical Confidence Interval
UI: Uncertainty Interval (equivalent to "Credible Interval" in some uses)
UTC: Urban Tree Canopy
NDVI: Normalized Difference Vegetation Index
NLCD: National Land Cover Database
USFS: United States Forest Service
HR: Hazard Ratio
SAI: Stratospheric Aerosol Injection
SSP: Shared Socioeconomic Pathways
CESM2: 2nd Community Earth System Model
ARISE-SAI: Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (CESM2-based climate model scenario for SAI intervention)