Global Metrics, Local Estimation: Magnifying the Health Impact of Environmental Justice

Joey Frostad

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee
Robert C Reiner, Chair
Joseph Dieleman
Esther Min

Program Authorized to Offer Degree
Global Health

University of Washington

Abstract

Global Metrics, Local Estimation: Magnifying the Health Impact of Environmental Justice

Chair of the Supervisory Committee:
Professor Robert C Reiner, PhD
Health Metrics Sciences

In the first chapter, *Mapping development and health effects of cooking with solid fuels in low-income and middle-income countries, 2000–18: a geospatial modeling study,* the prevalence of solid-fuel use for cooking is mapped at a 5 km × 5 km resolution in 98 LMICs based on 2.1 million household observations of the primary cooking fuel used from 663 population-based household surveys over the years 2000 to 2018. We use observed temporal patterns to forecast household air pollution in 2030 and to assess the probability of attaining the Sustainable Development Goal (SDG) target indicator for clean cooking. We aligned our estimates of household air pollution to geospatial estimates of ambient air pollution to establish the risk transition occurring in LMICs. Finally, we quantified the effect of residual primary solid-fuel use for cooking on child health by doing a counterfactual risk assessment to estimate the proportion of deaths from lower respiratory tract infections in children younger than 5 years that could be associated with household air pollution. We found that while reliance on solid-fuel use for cooking has declined globally, it remains widespread. 593 million people live in districts where the prevalence of solid-fuel use for cooking exceeds 95%. 66% of people in LMICs live in districts that are not on track to meet the SDG target for universal access to clean energy by 2030. Household air pollution continues to be a major contributor to particulate exposure in LMICs, and rising ambient air pollution is undermining potential gains from reductions in the prevalence of solid-fuel use for cooking in many countries. We estimated that, in 2018, 205 000 (95% uncertainty interval 147 000–257 000) children younger than 5 years died from lower respiratory tract infections that could be attributed to household air pollution.

The second chapter, *Scales of environmental justice: global sensitivity analysis of the Washington Environmental Health Disparities Map,* is focused on understanding the factors that drive environmental health inequalities more locally, within the context of the Washington State Environmental Health Disparities (EHD) Map, a composite indicator of environmental justice that synthesizes 19 different environmental and population health indicators to generate cumulative

impact rankings by census tract. We conducted a global sensitivity analysis of the EHD mapping methodology by permuting across alternative methods derived from the composite indicator construction literature, generating estimates of the uncertainty that results from analyst decisions in the development process. We estimated first, second and total order sensitivity statistics to quantify the relative influence of parameter choices on the tract rankings and on the accuracy of classifying communities in the top 20% of impact. On average, census tracts changed by more than one hundred ranks across these permutations. The observed deviations from the baseline EHD index were largest in the middle of the impact spectrum and smallest for tracts in the top 10% of impact. The formula used in aggregation and the method of data normalization were the most sensitive parameter decisions for both tract ranking and impact classification. We demonstrate that the EHD rankings were more robust in the highest impact tracts and relatively uncertain throughout the rest of the index, suggesting that this data is better suited for classifying hotspots than for estimating an ordinal impact. There are strong assumptions underlying the baseline EHD ranking methodology and these assumptions substantially drive the results

In the third chapter, *Validating the structure of an environmental justice index in Washington State, a multivariate case study,* we explore the mechanisms through which the most sensitive parameter choices in the EHD methodology impact the final index. The statistical characteristics of the raw indicator data are analyzed in order to understand the influence of various ranking transformations on the data distributions, and case studies that exhibit large changes in ranking and impact classification between linear and nonlinear transformations are analyzed to assess the bias introduced by normalization methodology. Nonlinear transformations are observed to favor the high-impact classification of tracts with above average values for the majority of indicators while reducing the effect of outliers. A variance-weighted Principal Component Analysis (PCA) is employed to compare the results of inductive aggregation to the more theoretically derived baseline index. A PCA based index is observed to agree generally with the baseline index and classify the impacted tracts with high accuracy using only the first two principal components. The loading structure of individual indicators within these first two components suggests that twin gradients of urban environmental degradation and socioeconomic deprivation are driving the index rankings in the current formulation of the EHD map.

Table of Contents

Chapter 1: *Mapping development and health effects of cooking with solid fuels in low-income and middle-income countries, 2000–18: a geospatial modelling study*

(Reprinted from https://doi.org/10.1016/S2214-109X(22)00332-1)

Summary

Background

More than 3 billion people do not have access to clean energy and primarily use solid fuels to cook. Use of solid fuels generates household air pollution, which was associated with more than 2 million deaths in 2019. Although local patterns in cooking vary systematically, subnational trends in use of solid fuels have yet to be comprehensively analysed. We estimated the prevalence of solid-fuel use with high spatial resolution to explore subnational inequalities, assess local progress, and assess the effects on health in low-income and middle-income countries (LMICs) without universal access to clean fuels.

Methods

We did a geospatial modelling study to map the prevalence of solid-fuel use for cooking at a 5 km × 5 km resolution in 98 LMICs based on 2·1 million household observations of the primary cooking fuel used from 663 population-based household surveys over the years 2000 to 2018. We use observed temporal patterns to forecast household air pollution in 2030 and to assess the probability of attaining the Sustainable Development Goal (SDG) target indicator for clean cooking. We aligned our estimates of household air pollution to geospatial estimates of ambient air pollution to establish the risk transition occurring in LMICs. Finally, we quantified the effect of residual primary solid-fuel use for cooking on child health by doing a counterfactual risk assessment to estimate the proportion of deaths from lower respiratory tract infections in children younger than 5 years that could be associated with household air pollution.

Findings

Although primary reliance on solid-fuel use for cooking has declined globally, it remains widespread. 593 million people live in districts where the prevalence of solid-fuel use for cooking exceeds 95%. 66% of people in LMICs live in districts that are not on track to meet the SDG target for universal access to clean energy by 2030. Household air pollution continues to be a major contributor to particulate exposure in LMICs, and rising ambient air pollution is

undermining potential gains from reductions in the prevalence of solid-fuel use for cooking in many countries. We estimated that, in 2018, 205 000 (95% uncertainty interval 147 000–257 000) children younger than 5 years died from lower respiratory tract infections that could be attributed to household air pollution.

Interpretation

Efforts to accelerate the adoption of clean cooking fuels need to be substantially increased and recalibrated to account for subnational inequalities, because there are substantial opportunities to improve air quality and avert child mortality associated with household air pollution.

Funding

Bill & Melinda Gates Foundation.

Introduction

The deleterious health effects of household air pollution are long established: solid-fuel use, defined by WHO as primary reliance on wood, crop residue, coal, or dung for cooking, heating, and lighting,1 was first associated with increased risk of respiratory infections in children in Papua New Guinea almost 50 years ago.2 The fine particulate matter smaller than 2·5 μm (PM2·5) generated by solid-fuel use is a complex mixture that causes harm to health through multiple pathways, including mucociliary dysfunction (which increases susceptibility to infection) and hyperinflammation or immunodeficiency (which can worsen disease prognosis).3 Solid-fuel use results in PM2·5 exposure both within the home and more broadly through emissions that contribute substantially to ambient air pollution.4, 5

High-income countries have almost fully transitioned to clean fuels (ie, the prevalence of solid-fuel use is less than 5%).6, 7 Across low-income and middle-income countries (LMICs), the net effects of household air pollution—including health effects (US$1·4 trillion), lost productivity ($0·8 trillion), and environmental degradation ($0·4 trillion)—represent an immense annual cost, and thus access to clean and sustainable energy needs to be an essential part of the development agenda.8 Clean cooking is core to proposed indicators for monitoring Sustainable Development Goal (SDG) 7 (target 7.1: universal access to clean fuels and technology), and has important synergies with goals related to health (SDG 3), education (SDG 4), gender (SDG 5), urban development (SDG 11), climate change (SDG 13), and terrestrial ecology (SDG 15).9 Prevention strategies targeting household air pollution are shifting towards

supplying households with technology or fuels for clean cooking, such as liquefied petroleum gas or electricity.10 Clean-fuel campaigns are often targeted subnationally, and even country-level programmes have shown heterogeneous patterns of adoption.11, 12 Previously, descriptive analyses of household air pollution and solid-fuel use have focused on a subset of relevant countries13, 14 or have been done globally but constrained by their spatial scale,6, 15 and were of little use for highlighting local patterns or identifying subnational inequality.

Research in context

Evidence before this study

We did not do a formal systematic search of the literature. Previous analyses have quantified the cause-specific disease burden associated with household air pollution globally, including studies using integrated exposure–response curves and pooled meta-analysis from a systematic review. These efforts showed the substantial health effects associated with cooking with solid fuels but did not examine trends in the underlying prevalence of the use of solid fuel in depth. A study estimated the prevalence of primary reliance on specific fuel types at the global and national level over the past 30 years, including forecasts to 2030. Evidence that community-level drivers are the strongest predictors of clean fuel adoption implies that the operational scale of solid-fuel use is more granular and that failure to account for local patterns could obscure inequalities.

Added value of this study

We used geostatistical methods to estimate the prevalence of primary use of solid fuels for cooking and household air pollution at substantially higher resolutions than previous studies, which allowed us to do subnational trend analysis in 98 low-income and middle-income countries from 2000 to 2018. By aggregating these geospatial estimates to second-level administrative boundaries (districts), we were able to provide actionable insights aligned to the scale of precision public health. We also made projections of progress to 2030, which suggested that few countries are on track to reach the Sustainable Development Goal of universal access to clean fuels within the coming decade. Finally, we combined our high-resolution estimates of the prevalence of solid-fuel use for cooking with exposure–response curves and equivalently resolved estimates of under-5 mortality from lower respiratory tract infections to quantify effects on child health.

Implications of all the available evidence

Although some regions exhibited substantial progress from 2000, in many regions in low-income and middle-income countries, primary reliance on solid fuels for cooking was still ubiquitous in

2018. We noted substantial subnational disparities, leading to health inequalities. Local estimates highlight the outstanding challenge of attaining universal access to clean cooking fuels, and risk assessments showed that hundreds of thousands of children still die annually from lower respiratory tract infections associated with household air pollution. The economic downturn and increased public health strain associated with the ongoing COVID-19 pandemic suggest that our forecasts are likely optimistic, and that the transition to clean and modern fuels must be broadly accelerated to fulfil the bold ambitions of the Sustainable Development Goals.

In this study, we generate the first high-resolution geospatial estimates of the prevalence of solid-fuel use (as indicated by primary fuel type) and the resulting household concentrations of PM2·5 in 98 LMICS. Our aim was to assess growth in access to clean cooking fuels over the past two decades. We also use temporal trends from 2000 to 2018 to forecast the likelihood of achieving SDG target 7.1 by 2030. We further quantify the household-level relationship between household and ambient air pollution by juxtaposing our estimates with data for ambient exposure to PM2·5 with equivalent spatial resolution to establish a robust indicator of total personal exposure to PM2·5 air pollution. Finally, we combine our results with population data, PM2·5 exposure–response functions, and equivalently resolved geospatial estimates of mortality from lower respiratory tract infections (LRTIs) in children younger than 5 years—a case study designed to assess the health effects of residual solid-fuel use in this vulnerable population.

Methods

Data sources

Solid-fuel use was estimated on the basis of population-based household survey data, in which respondents indicated the primary cooking fuel being used in the household. These responses were mapped to one of eight categories: no cooking in household, electricity, gas, kerosene, coal, wood, crop waste, and dung. Coal, wood, crop waste, and dung were considered solid fuels, whereas the others were deemed clean fuels. We then constructed a binary indicator of solid-fuel use (primary reliance on solid fuels vs primary reliance on clean fuels).

98 LMICs were included in the analysis based on their Socio-demographic Index scores (a development index derived from education, fertility, and poverty estimates), which were calculated using values from the low, low-middle, and middle quintiles from the Global Burden of Disease 2019.16 Sources of input data were only included for modelling if they were representative of the entire population during the time period and across the geographical area

of measurement. Furthermore, certain sources were excluded if the associated estimates seemed implausible based on expert review of estimates and comparison with other sources in the same country and time period. We excluded LMICs with populations of less than 1 million and those that did not have household survey data available (appendix p 2). In total, 663 household surveys were compiled and extracted (appendix p 84). The full database represented 2·1 million people from 2000 to 2018 and included geocoded information from 181 556 coordinates (points) and 417 650 subnational administrative boundaries (polygons). Further details about the data-extraction and data-processing sequence are in the appendix (p 4).

Definitions

In this study, we defined solid-fuel use as the household-level prevalence of primary reliance on solid fuels for cooking, which was described by the adminstrators of the included surveys as the fuel used most often for cooking in a household. In accordance with the Global Burden of Disease (GBD) 2019 study,16 household air pollution was defined as the incremental concentration of PM2·5 generated from cooking with solid fuels. We used the estimated ambient concentration of PM2·5 in a location as the baseline exposure (appendix p 6). We subtracted this value from the total personal PM2·5 exposure estimated for a solid fuel user: the difference represented the additional contribution of household air pollution to PM2·5 exposure.

Statistical analysis

Available geospatial covariates with plausible a priori relationships with solid-fuel use were compiled for use in the prediction model (appendix p 4). We included seven indicators of urbanicity or development, which could be associated with increased access to clean-fuel technologies,17 such as travel and night-time lights. We also included 16 environmental indicators that might be associated with access to fuelwood or other solid fuels,18 including diurnal temperature range, elevation, and the normalised difference vegetation index (an indicator of whether a given observation contains live green material, which is calculated by comparing satellite images generated from visible and near-infrared light to estimate plant mass in the pixel). To account for potential multicollinearity, we used the variance inflation factor to analyse these covariates and filtered for each modelling region using a threshold of 5 (which was chosen to prioritise predictive over explanatory power).

We used a Bayesian hierarchical modelling framework to model household exposure to solid-fuel use through a generalised linear mixed-effects model that was spatially explicit. Prevalence of exposure to solid-fuel sources was modelled using the observed number of household members exposed as binomial count data (Cd) among a sample size (Nd). Annual

prevalence in each primary sampling unit (cluster; d) for each survey was the modelled quantity, which was mapped to a geospatial raster location (i) for every year (t):

$$C_d \left| p_{i(d)} N_d \sim Binomial(p_{i(d)} N_d) \forall obs.clustersd \right.$$

$$\text{logit}(p_{i,t}) = \beta_0 + X_{i,\beta} + \varepsilon_{c(i)} + \varepsilon_{n(i)} + \varepsilon_i - Z_i$$

$$\sum_{n=1}^{3} \beta_h = 1$$

$$\varepsilon_c \overset{iid}{\sim} N(0, \gamma_c^2)$$

$$\varepsilon_n \overset{iid}{\sim} N(0, \gamma_n^2)$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Z \sim GP(0, \sum^{space})$$

$$\sum^{space} = \frac{\omega^2}{\Gamma(v) 2^{v-1}} * (kD)^v * K_v(kD)$$

The appendix (p 5) contains a detailed explanation of these calculations, including definitions of all other included variables.

Prevalence of solid-fuel use was modelled as a linear combination of three submodels (generalised additive models, gradient boosted decision trees, and lasso regression; appendix p 4), rasterised spatiotemporal covariate values, a correlated spatial random effect term (Zi), country random effects (εc), survey-specific random effects (εn), and an independent nugget random effect (εi). The coefficient of each submodel (β), represented the predictive weighting within the logit link. A key strength of this approach is the ability to leverage residual correlation structures within the predictions to make more accurate estimates for data-sparse locations, while simultaneously propagating this dependence through to estimates of uncertainty in all indicators (appendix p 6). The posterior distributions were fit based on approximations in integrated nested Laplace approximation (R-INLA), with approximation of the stochastic partial differential equations to the Gaussian process residuals done in R (version 3.6.1).

Models were assessed on the basis of a five-fold out-of-sample cross-validation strategy that was stratified over space (appendix p 6). Estimates of bias (mean error), variance (root-mean-square error), coverage of data by 95% prediction intervals (appendix p 6), and correlations between predictions and observed data were used to assess the models (appendix p 6). In-sample and out-of-sample model validation plots were also produced comparing every country and first and second administrative unit estimated with the observed data for those units (appendix p 6).

Pixel-level estimates of solid-fuel use were calibrated to estimates from the GBD 2019 study using a previously described method19 to preserve relative spatial patterns while ensuring comparability and incorporating information from national-level reports that could not be used within geospatial models. We defined calibration factors on the basis of comparison of draws from GBD outputs to population-weighted aggregations of our estimates at the highest level of spatial granularity available (either national or the first administrative level for select countries for which GBD-produced subnational estimates were available).

We combined the model output (P(SFU)) with geospatial estimates of population (pop) and ambient PM2·5 exposure (APM2·5) from the GBD 2019 study16 to calculate personal total exposure to PM2·5 pollution (TAP) (the sum of household and ambient air pollution in each 5 km × 5 km grid cell i [pixel]):

$$HAP_{i,c,t} = (P(SFU)_{i,c,t} * pop_{i,c,t}) * HPM_{2.5_{c,t}}$$

$$TAP_{i,c,t} = pop_{i,c,t} * APM_{2.5_{c,t}} + HPM_{i,c,t}$$

$$TAPRR_{i,c,t,o} = IER(\frac{TAP_{i,c,t}}{pop_{i,c,t}})_o$$

$$TAPPAF_{i,c,t,o} = \frac{TAPRR_{i,c,t,o}}{TAPRR_{i,c,t,o}-1}$$

$$HAPN_{i,c,t,o} = TAPPAF_{i,c,t,o} * N_{i,c,t,o} * HAP_{i,c,t}/TAP_{i,c,t}$$

Estimates of the expected incremental PM2·5 concentration generated in a household using solid fuels (HPM2·5) for a given country (c) and year (t) from GBD 2019 were used to calculate the concentration of household air pollution in the exposed population.20 The per-person annual average ambient PM2·5 estimate from GBD 2019 (APM2·5) was summed with the household air pollution concentration to calculate the total air pollution concentration. The fraction of total personal exposure to PM2·5 air pollution contributed by household air pollution in each pixel

was estimated to provide the household air pollution share (HAP%). Finally, the per-person air pollution concentration in each pixel was used as an input to the GBD 2019 risk (IER).16 curve for LRTIs to estimate a relative risk (RR) and population attributable fraction (PAF) for every PM2·5-associated outcome (o) in each pixel. The population attributable fraction is a counterfactual estimate of the disease burden explained by a given risk factor, based on the level of exposure and corresponding excess relative risk of disease.16 The population attributable fraction for LRTIs was combined with pixel-level estimates21 of under-5 LRTI mortality counts(Ni,c,t,o) to estimate the count (TAP N) and rate of deaths from LRTIs that were attributable to total air pollution and specifically to household air pollution and ambient air pollution in each district.

All pixel-level indicators were population-weighted and aggregated to first (regions) and second (districts) administrative levels using shapefile boundaries from the Database of Global Administrative Areas shapefiles. We quantified within-country inequalities by using the range between the best-performing and worst-performing district for a given year and the average interpersonal difference, which estimates the average difference between any two districts in a country-year.22 Absolute and annualised rates of change from 2000 to 2018 were computed to quantify temporal trends over the study period. This annualised rate of change was applied to 2018 values to project the summary exposure value (a risk-weighted summary measure of exposure prevalence)16 to 2030 for assessment of the attainment of SDG indicator 7.1.2 (the proportion of population with primary reliance on clean fuels and technology, which is used as a tracking benchmark for SDG 7). We chose this target indicator because it corresponds most closely to our modelled proportion on the basis of previously published23 methods that were consistent with the GBD 2019 study.16 The annualised rate of change for solid-fuel use (P) was calculated at the draw level (i) for each pixel (m) by estimating the rate between each pair of adjacent years (t):

$$Annualisedrateofchange_{i,m,t} = \text{logit}(\frac{P_{i,m,t}}{P_{i,m,t-1}})$$

The annualised rate of change was then weighted across all years, such that more recent rates were additionally weighted. Weights (*W*) were defined as:

$$W_t = (t - 200 + 1)^{\omega}$$

For this analysis, ω was defined using draws from a distribution generated empirically for this indicator in the GBD 2019 study, with a mean value of 2·3 and an SD of 0·41. The weighted annualised rate of change for each pixel was generated for each unit as:

$$Annualisedrateofchange = \text{logit}(\sum_{2000}^{2018} \text{W}_\text{t} * Annualised rateofchange_{i,m})$$

Annualised rate of change estimates for 2018 were used to predict for 2030:

$$\text{Proj}_{\text{i,m,2030}} = \text{logit}^{-1}(\text{logit}(P_{i,m,2018}) + Annualisedrateof change_{i,m} * (2030 - 2018))$$

Attainment probabilities for SDG indicator 7.1.2 were derived from the percentage of simulations in 2030 with summary exposure below 5%, a threshold chosen on the basis of estimates of solid-fuel use in high-income countries.24 We used the R-INLA package in R (version 4.1.3) for our analyses. All code used in the analysis is available online.

Role of the funding source

The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or the writing of the report.

Results

Subnational solid-fuel use prevalence was notably spatiotemporally heterogeneous (figure 1A), underscoring the importance of subnationally tracking cooking behaviours and corresponding measures of pollution. In 2018, 10·0% (95% uncertainty interval [UI] 7·4–13·9; appendix p 6) of the population in LMICs (593 million people) lived in districts where the prevalence of solid-fuel use exceeded 95%. National prevalence estimates masked substantial within-country heterogeneity (figure 2). For example, the average solid-fuel use prevalence in Guatemala was 52·0% (95% UI 40·0–63·2), but across just 140 km, prevalence ranged from 1·0% (0–3·3) in Zona 22, Guatemala City, to 91·6% (79·4–97·8) in Santa Eulalia, Huehuetenango. During the study period, the average interpersonal difference increased by 43·0% across all LMICs.

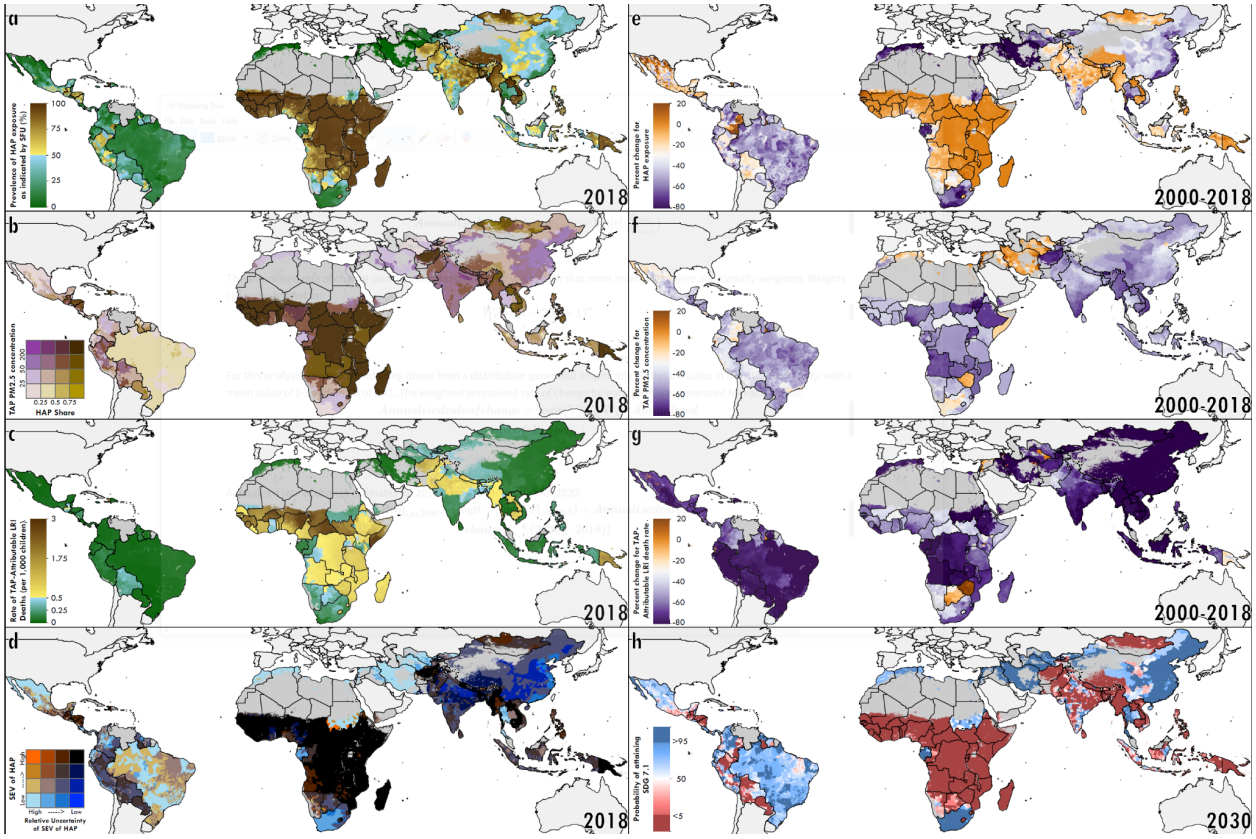


Figure 1 Prevalence of, and change from 2000 to 2018, in solid-fuel use, household air pollution, total air pollution, and deaths attributable to LRTI in children younger than 5 years in low-income and middle-income countries

(A) Mean prevalence of solid-fuel use for cooking, 2018 (as indicated by primary fuel type). (B) Total concentration of PM2·5 in air by source, 2018. (C) Mean number of deaths from LRTIs (per 1000 children younger than 5 years) attributable to total concentration of PM2·5 in air, 2018. (D) Overlapping terciles of mean risk-weighted prevalence (SEV) of HAP in 2018 and relative uncertainty. These data were used for our projections of SDG attainment in 2030. (E) Percentage change in mean proportion of solid-fuel use for cooking, 2000 to 2018. (F) Percentage change in total concentration of PM2·5 in air, 2000 to 2018. (G) Percentage change in the proportion of LRTI deaths (per 1000 children younger than 5 years) attributable to total concentration of PM2·5 in air, 2000 to 2018. (H) Probability of attaining SDG 7.1 by 2030. All panels are aggregated to the second administrative-level unit. Maps reflect administrative boundaries, land cover, lakes, and population. Grey-coloured grid cells had fewer than ten people per 1 km × 1 km grid cell and were classified as barren or sparsely vegetated, or not included in this analysis. HAP=household air pollution. PM2·5=particulate matter of less than

2·5 μm in diameter. SEV=summary exposure value. LRTI=lower respiratory tract infection. SDG=Sustainable Development Goal.

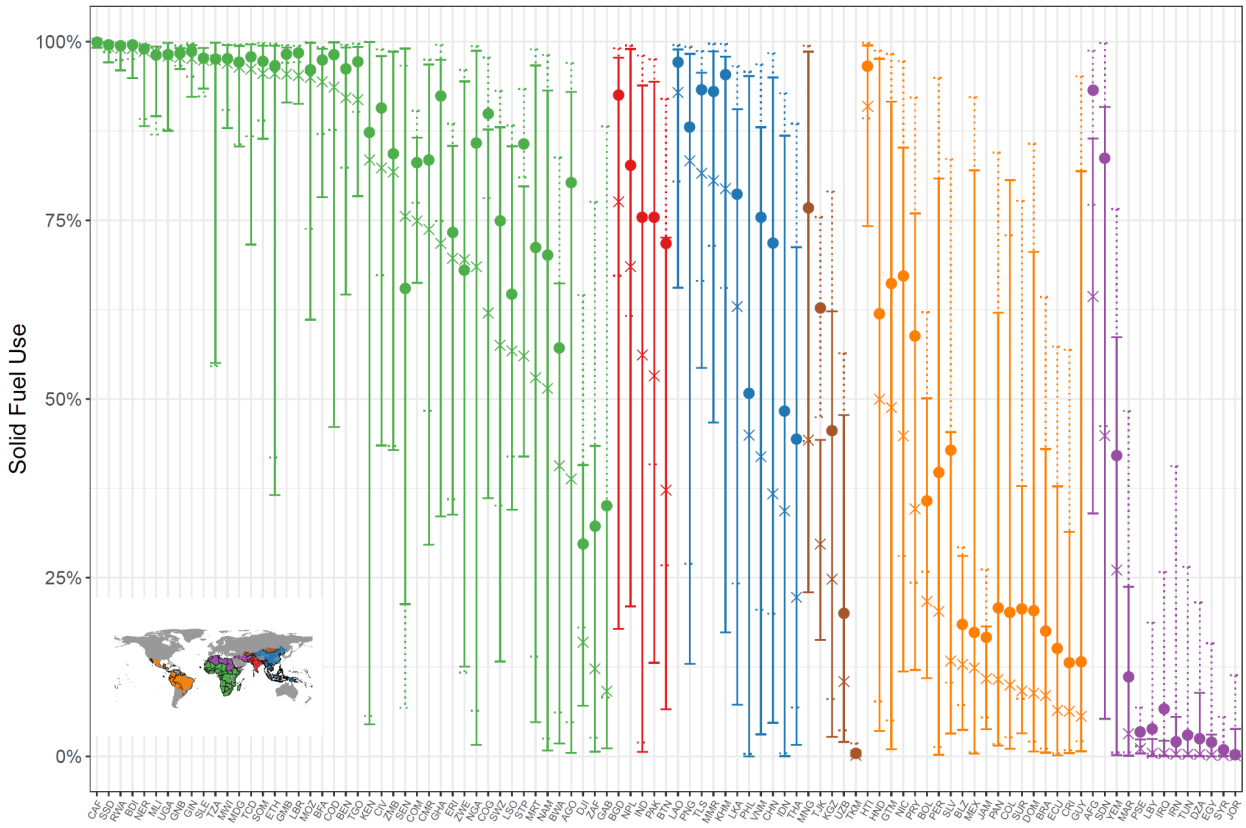


Figure 2 Geographical variations in the prevalence of solid-fuel use for cooking in 98 low-income and middle-income countries

Each bar represents the range of the prevalence of solid-fuel use for cooking (as represented by primary fuel type) across all districts within each country. The Xs and dashed bars represent the mean and range in 2000, whereas the dots and solid bars represent the mean and range in 2018. Countries are grouped according to geographical region and coloured according to the Global Burden of Disease 2019 super-regions (see inset thumbnail).16 Each country is labelled by its ISO-3 code.

Prevalence of solid-fuel use fell across LMICs (figure 1E) from 67·8% (95% UI 67·1–69·0) in 2000 to 56·5% (53·8–57·9) in 2018, a relative decrease of 16·7% (14·3–21·6). Due to population growth, however, an additional 359 million (234 million–388 million) people were exposed to solid-fuel use in 2018. We estimated that, in 2018, 56·5% (53·9–57·9) of people in

LMICs, roughly 3·38 billion (3·21 billion–3·45 billion) of the 5·96 billion in the study area (which rose from 4·4 billion in 2000) relied primarily on solid fuels for cooking. Projections for the year 2030 (figure 1D) suggest that only five countries (Iraq, Iran, Jordan, Syria, and Turkmenistan) across three regions, representing 3% of the 2018 study population, have a greater than 95% probability of achieving universal access to clean cooking fuels (figure 1H) in every district. District-level forecasting highlights the importance of using subnational estimation to monitor progress towards universal access. National trends suggest that Mexico has a 98% probability of meeting the threshold for universal access by 2030. However, this figure obscures local inequalities, because 14% of the population (nearly 19 million people in 2018) live in districts where the probability is less than 50%. In India, solid-fuel use prevalence has substantially decreased, and the national attainment probability is up to 33%. Locally, however, more than 35% of people (482 million people in 2018) lived in districts where the forecasted probability of attainment is less than 5%.

Moderate rates of change and the shortening timeline for achievement of the SDGs suggest that few districts that had not already achieved the target of universal access to clean cooking fuels in 2018 will do so by 2030. In 2018, 3·8 billion people in LMICs (ie, 65·7% of the population) lived in districts that did not meet the threshold for universal access to clean cooking fuel. Our projections suggest that only 22·0% (835 million) of these people live in a district that is forecasted to meet the threshold by 2030. Countries with the largest share of districts that were above the threshold in 2018 but on track to meet it by 2030 included Gabon (where 93·9% of the population that has yet to meet the goal lives in districts that are projected to meet it by 2030), China (82·9%), South Africa (66·5%), Mongolia (52·7%), Suriname (40·5%), Ecuador (27·6%), and Uzbekistan (24·6%). In 27 countries, every district is projected to remain above the 5% threshold by 2030 with high certainty (figure 1H). In the 48 countries containing both districts that are projected to meet the threshold and districts projected to fail to meet it, the proportion of the population living in a district that is not on track (as of 2018) ranged from less than 1% (14 459 people) in Uzbekistan to 95% (49 million people) in Kenya (figure 3).

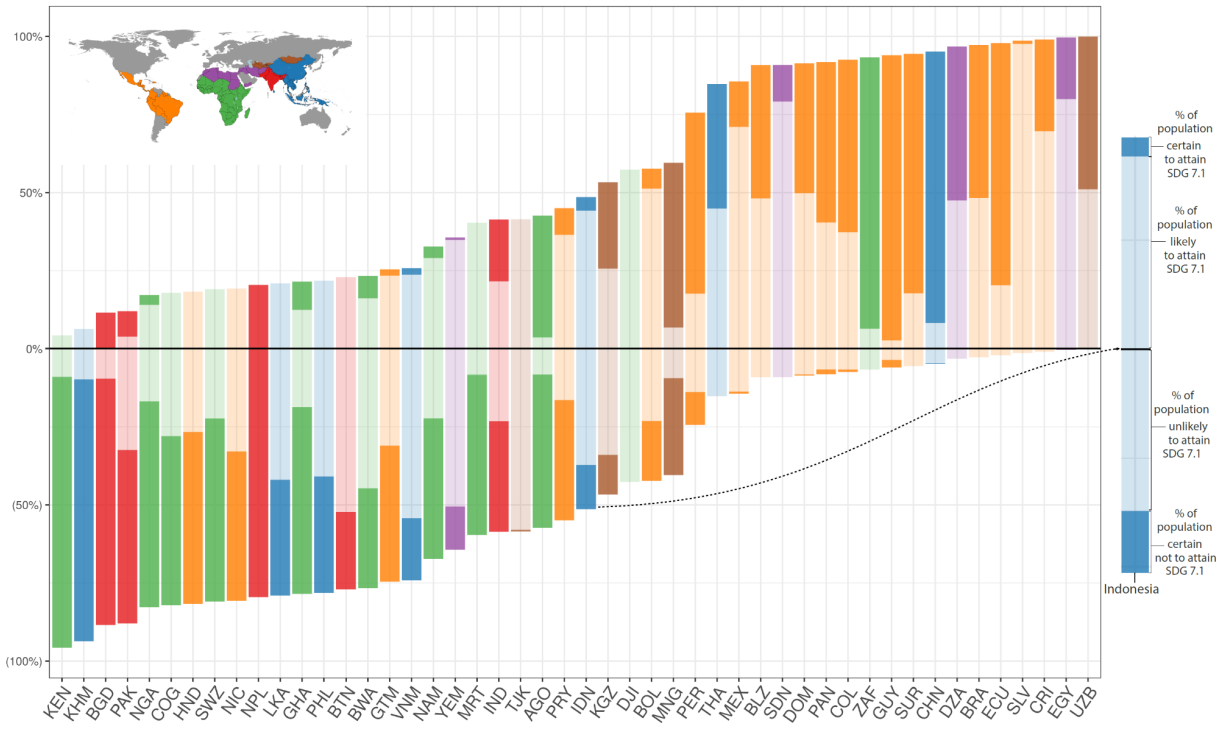


Figure 3 Projections of the probability of attainment of SDG 7.1 by 2030

Progress towards SDG 7.1 was monitored by estimating access to clean cooking fuel across the study period, calculating the annualised rate of change and forecasting to 2030. The bars represent the share of population living in districts that were projected (with 95% certainty) to be above (negative values) or below (positive values) a threshold of 5% for attainment in 2030. The translucent, lighter-coloured sections of the bars represent projections in districts with less certainty (greater than 50%, but less than 95%). The bars are coloured according to Global Burden of Disease 2019 super-regions (see inset thumbnail).16 Only the 48 countries where at least one district was projected to be above the threshold and at least one was projected to be below the threshold were included. Each country is labeled by its ISO-3 code. SDG=Sustainable Development Goal.

The total concentration of PM2·5 air pollution to which people were exposed was nearly halved (mean decrease 47·2% [95% UI 46·4–47·7]) from 2000 to 2018 (figure 1F). However, in 2018, only 14·9% (10·9–18·4) of people in LMICs lived in districts that met WHO's air quality interim target25 of 35 µg/m3 (figure 4A), and 80% of people were exposed to annual concentrations greater than 44 µg/m3 (for comparison, in 2000, 80% of the population lived in pixels with

concentrations that were almost double that value, at 82 µg/m3 or more). Globally, the proportion of the total particulate matter concentration contributed by household air pollution decreased from 66·3% (64·8–70·0) in 2000 to 46·5% (43·2–47·8) in 2018. Household air pollution contributed most of the total PM2·5 air pollution in 69 of 98 countries in 2000, and 50 of 98 in 2018, suggesting a risk transition for the main source of particulate matter exposure in LMICs.

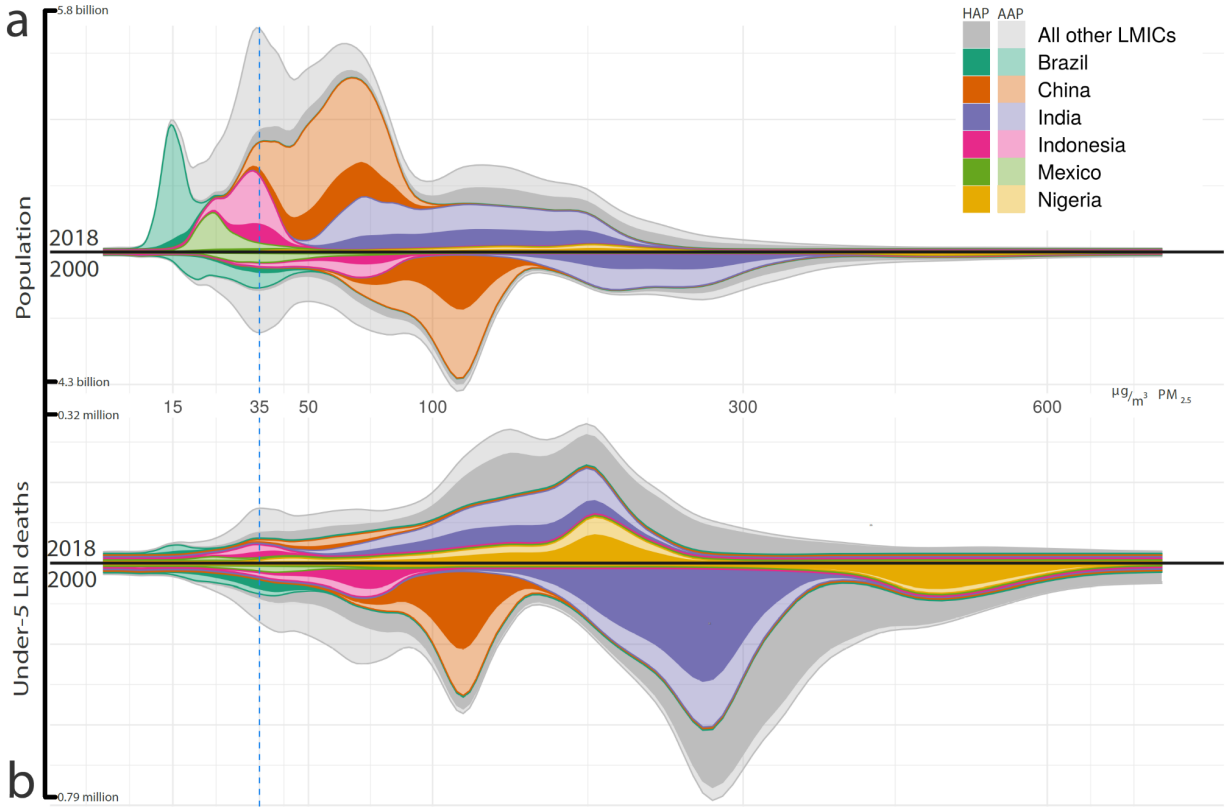


Figure 4 All-age population (A) and LRTI deaths attributable to total PM2·5 concentrations in air in children younger than 5 years (B) in 2000 and 2018, distributed as a function of PM2·5

Within each distribution, the darker shading represents the portion of air pollution contributed by household sources, whereas the lighter shading represents the portion contributed by ambient sources. The plotted data represent local smoothing of normalised distributions that were computed over 400 logarithmically spaced bins. The dashed vertical line indicates WHO's interim threshold for PM2·5 air pollution (35 µg/m3). The y-axis labels provide the total area

under the curve. Data for all other LMICs included in the study are broken down by country in the appendix (p 42). LRTI=lower respiratory tract infection. HAP=household air pollution. AAP=ambient air pollution. PM2·5=particulate matter of less than 2·5 μm in diameter. LMICs=low-income and middle-income countries.

We focused our case study on the effects of particulate matter on child health, given the availability of analogous high-resolution geospatial estimates for mortality attributable to LRTIs in children younger than 5 years.21 Over the study period, the population attributable fraction of fatal LRTIs associated with total exposure to PM2·5 air pollution in all LMICs fell from 45·7% (95% UI 32·2–57·6) in 2000 to 38·4% (24·9–51·9) in 2018, a relative reduction of 16·5% (95% UI 9·0–26·5). Nearly half of fatal LRTIs in 2018 were attributable to total exposure to PM2·5 air pollution in sub-Saharan Africa (49·8% [36·4–61·4]). Across LMICs, most fatal LRTIs were attributable to total exposure to PM2·5 air pollution in 10·0% (0·0–22·3) of districts. We estimated that, across LMICs in 2018, 320 000 (218 000–408 000) children died from LRTIs attributable to air pollution (figure 1D) compared with 747 000 (533 000–932 000) in 2000. This reduction was driven largely by reductions in solid-fuel use: the number of LRTI deaths attributable to household air pollution in children younger than 5 years fell by 65·5% (62·3–68·6). However, an estimated 205 000 (147 000–257 000) children still died from LRTIs attributable to household air pollution in 2018. These deaths are now associated with lower pollution concentrations (figure 4B): the median was 202 μg/m3 in 2018, compared with 326 μg/m3 in 2000. Across LMICs, the share of attributable under-5 LRTI mortality that was driven by household versus ambient particulates fell from 79·6% (77·4–81·6) in 2000 to 64·3% (61·6–67·9) in 2018, signalling that household air pollution is still a crucial factor in under-5 LRTI deaths.25

District-level burden estimates further underscored the risk transition occurring in LMICs: the concentrations of particulate matter to which populations were exposed decreased, and were increasingly driven by ambient sources (figure 5). This transition generally decreased the share of LRTIs attributable to particulate air pollution (figure 1E). However, as experienced in China, India, and Nigeria, sharply rising outdoor pollution superseded clean cooking adoption and offset potential burden reductions (figure 5). In some countries, the prevalence of fatal LRTIs (figure 1G) fell without a corresponding decline in household air pollution: in Laos, for example, the rate of fatal LRTIs per 1000 children younger than 5 years fell 73·8% (95% UI 66·3–80·9) between 2000 and 2018, while solid-fuel use prevalence decreased by only 3·5% (1·0–6·9). In the 19 countries where under-5 LRTI death rates still exceed two per 1000 children, however,

the proportion of LRTI deaths attributable to household air pollution ranged from 21·4% (12·8–30·0) in Lesotho to 60·9% (44·8–75·7) in Somalia, suggesting that reduction of household air pollution remains fundamental to the elimination of preventable child mortality.

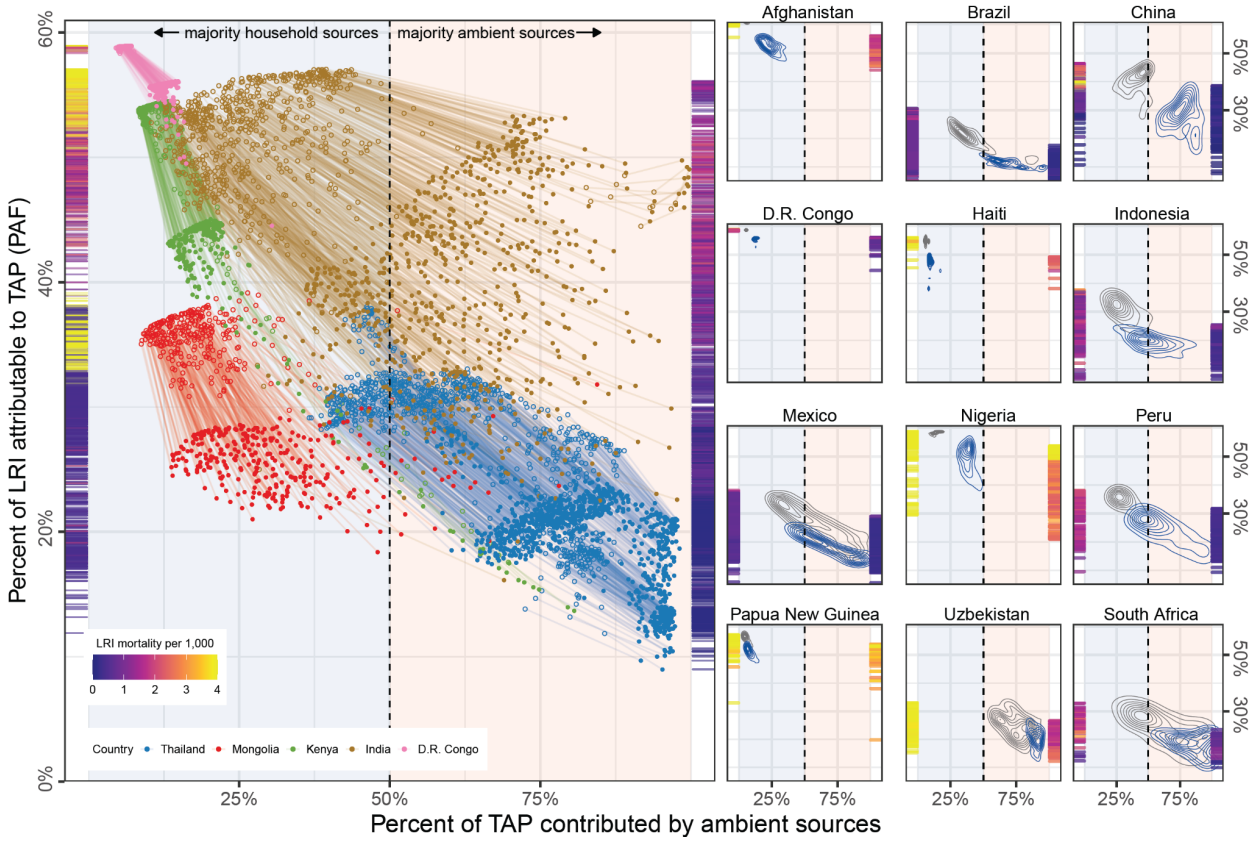


Figure 5 Air pollution risk transition, 2000–18

(A) Trends in the proportion of LRTI deaths attributed to total air pollution at the second administrative unit (district) level in five low-income and middle-income countries. These countries were chosen to exemplify different stages of air-pollution risk transition (all other countries included in the study are illustrated in the appendix [p 63]). The y-axis rugs indicate the gradient of background LRTI mortality rates for 2000 (left) and 2018 (right), illustrating the correlation between LRTI rates and the fraction attributable to total ambient air pollution. The lines connect a district to its preceding timepoint across the series. (B) Countries with the highest LRTI mortality in 2000 for each Global Burden of Disease subregion. The grey contours represent district-level distributions in 2000, whereas the navy represents distributions in 2018. In Thailand and South Africa, for example, reductions in household air pollution have resulted in less than a quarter of LRTIs being attributable to air pollution, whereas in China and India,

similar reductions have been counteracted by rising ambient air pollution concentrations, which means that a larger share of LRTI deaths in children younger than 5 years continue to be attributable to total air pollution. LRTI=lower respiratory tract infection.

Discussion

The results of our geospatial modelling study suggest that, despite progress since 2000, solid-fuel use continues to be widespread in LMICs, with uneven improvements driving regional and within-country inequalities. Our projections indicate that attaining the SDG target of universal access to clean cooking fuels by 2030 is improbable for many countries. Total particulate matter exposure fell between 2000 and 2018, but most people in LMICs continue to be exposed to PM2·5 concentrations far above the interim air quality target of 35 μg/m3. The fraction of total PM2·5 contributed by ambient air pollution rose, as growing outdoor concentrations in many geographies offset the health gains of cleaner cooking. The combination of a slow transition from solid-fuel use and displacement by ambient pollution sources supports a mandate for strengthening efforts to drive decreases in exposure to realise crucial health gains. The health effects of inaction are particularly relevant to vulnerable populations, such as children younger than 5 years, among whom more than a third of deaths from LRTIs are still attributable to air pollution.

The inclusion of clean energy in multinational initiatives such as the Global Forum on Child Pneumonia,26 climate change mitigation,27 and a web of interlinked SDG targets9 indicates that promotion of clean fuels will be crucial to the global development agenda in the coming decade. The energy-ladder hypothesis initially asserted that adoption of clean fuel sources was primarily driven by rising income, but subsequent research suggests that residential energy choices are influenced by a variety of complex socioeconomic forces, including community factors, agricultural practices, and dietary preferences.14, 28 Cooking is social, and campaigns to develop new energy markets by reducing prices and strengthening supply chains should be integrated with marketing and educational outreach activities to build knowledge and change household norms.29 Policies that include rural electrification and fuel subsidies have been efficacious, but our estimates support critiques that these campaigns need to be accelerated and redirected to target rural households that are continuing to fall behind in the adoption of clean cooking fuels.30

Ambitious programmes, such as India's Pradhan Mantri Ujjwala Yojana, have substantially increased adoption of clean fuels, but beneficiaries struggle to sustain regular use.12 Our estimates were derived from data for primary fuel type, and the widespread nature of stacking

(ie, parallel use of multiple fuel types31) supports the value of a more specific proxy for clean cooking than that used by us and in most household surveys. The Multi-Tier Framework surveys for some countries show that the extent of stacking depends on local context: for example, most urban households in Cambodia stack,32 whereas more than 90% of households in Rwanda use one stove to meet their energy needs.33 Air-monitoring data help to clarify the direction and magnitude of the bias introduced by exposure misclassification as a result of stove stacking: compared with households that exclusively use their primary fuel for cooking, households that primarily use clean fuels but stack with a solid fuel have a much higher particulate matter differential than households who stack with a solid primary and clean secondary fuel, suggesting that primary fuel data underestimate the true burden of household air pollution.34

Ultimately, switching of primary fuel sources is an inadequate target for the reduction of household air pollution to levels that are acceptable for human health. While our analysis of primary fuel data shows the tremendous scale of residual solid-fuel use, it represents a narrow interpretation of SDG 7.1, and attainment of universal access to modern energy should incorporate solutions that acknowledge the contextual challenges of LMIC households (eg, power outages, fuel shortages, device malfunctions). Expanding the spatiotemporal coverage of nuanced surveys with more specific assessments of household energy behaviours would allow for future analyses to maintain our global scope while transitioning to more sophisticated indicators, like the Multi-Tier Framework data, to enable calculation of adequate proxies for total personal exposure to PM2·5. Likewise, other indicators relevant to the immediate housing environment35 should be integrated with these estimates to account for the synergistic effects they have on health outcomes.

Our study had several limitations. Although data6 suggest that use of kerosene as a fuel for cooking has substantially diminished globally during the study period, our exclusion of kerosene in calculations of household air pollution probably underestimates the burden in countries where kerosene is still commonly used,36 such as Equatorial Guinea, Djibouti, and urban areas in some Oceanian nations.6 Furthermore, the lack of comprehensive data for fuels used for heating and lighting restricted our analyses to cooking. Solid-fuel use for heating especially contributes substantially to seasonal concentrations of PM2·5 in LMICs with cooler climates.37 Thus our burden estimates are probably too low in important cases, such as China38 and South Africa.39 Restricting our epidemiological case study to LRTI deaths in children younger than 5 years is an additional limitation, because LRTIs represent only a fraction of the deleterious health effects of air pollution. Aside from the robust evidence for effects on other child health endpoints, including adverse birth outcomes40 and neurodevelopmental effects,41 PM2·5

pollution substantially affects adult health and is associated with chronic diseases like cardiovascular disease, chronic obstructive pulmonary disease, lung cancer, and diabetes.7 Finally, the accuracy of our estimates is a function of the quality and volume of survey data available, and uncertainty is substantial in areas where data are missing or less reliable. Our projections are based on a simplified method for applying annual trends to estimates for 2018, and are dependent on countries maintaining this rate of progress. Accordingly, they do not reflect the potential effect of new technologies or investments in clean energy, nor do they capture how widespread economic and societal disruptions—such as those associated with the COVID-19 pandemic—could negatively affect immediate and long-term trajectories. Pandemic-related effects on pollution are likely to be multifaceted and nuanced across settings, and could vary in relation to a combination of public health measures taken against COVID-19.42

There have been notable triumphs in the push for global transition to clean cooking fuels, yet our estimates clearly show the breadth of residual exposure to household air pollution and the substantial challenges faced in addressing the relevant SDG targets before 2030. Unless global and national investments in clean energy access and adoption increase substantially this decade, household particulate concentrations will remain far above acceptable levels and the bold SDG ambitions for clean energy will remain unmet. Heterogeneity in reductions in the prevalence of solid-fuel use emphasises growing inequality in much of the world, and the changing relationship between solid-fuel use and ambient air pollution underscores the importance of continuing to push for universal access to clean fuels. Collectively, our results emphasise the need to support access to clean cooking fuels to achieve development goals and can help to inform the design of geospatially targeted campaigns that combat areas of enduring deprivation by proactively targeting inequality.

References

1. WHO guidelines for indoor air quality: household fuel combustion. (World Health Organization, 2014).
2. Anderson, H. R. Respiratory abnormalities in Papua New Guinea children: the effects of locality and domestic wood smoke pollution. Int. J. Epidemiol. 7, 63–72 (1978).
3. Smith, K. R., Samet, J. M., Romieu, I. & Bruce, N. Indoor air pollution in developing countries and acute lower respiratory infections in children. Thorax 55, 518–532 (2000).
4. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367 (2015).

5. Landrigan, P. J. et al. The Lancet Commission on pollution and health. The Lancet 391, 462–512 (2018).
6. Zhao, B. et al. Change in household fuels dominates the decrease in PM2.5 exposure and premature mortality in China in 2005–2015. Proc. Natl. Acad. Sci. 115, 12401–12406 (2018).
7. The State of Access to Modern Energy Cooking Services. World Bank https://www.worldbank.org/en/topic/energy/publication/the-state-of-access-to-modern-energy-cooking-services.
8. Fuso Nerini, F. et al. Mapping synergies and trade-offs between energy and the Sustainable Development Goals. Nat. Energy 3, 10–15 (2018).
9. Quinn, A. K. et al. An analysis of efforts to scale up clean household energy for cooking around the world. Energy Sustain. Dev. 46, 1–10 (2018).
10. Mortimer, K. et al. A cleaner burning biomass-fuelled cookstove intervention to prevent pneumonia in children under 5 years old in rural Malawi (the Cooking and Pneumonia Study): a cluster randomised controlled trial. Lancet 389, 167–175 (2017).
11. Urmee, T. & Gyamfi, S. A review of improved Cookstove technologies and programs. Renew. Sustain. Energy Rev. 33, 625–635 (2014).
12. Wang, W., Assaf, S., Mayala, B. & Davis, L. M. Household air pollution: national and subnational estimates in Bangladesh, India, Indonesia, Nepal, and the Philippines. (2020).
13. Shupler, M. et al. Household, community, sub-national and country-level predictors of primary cooking fuel switching in nine countries from the PURE study. Environ. Res. Lett. 14, 085006 (2019).
14. Stoner, O. et al. Household cooking fuel estimates at global and country level for 1990 to 2030. Nat. Commun. 12, 5793 (2021).
15. Wang, Q. et al. Impacts of residential energy consumption on the health burden of household air pollution: Evidence from 135 countries. Energy Policy 128, 284–295 (2019).
16. Stevens, G. A. et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. The Lancet 388, e19–e23 (2016).
17. Murray, C. J. L., GBD 2019 Risk Factors Collaborators & S.S., Lim. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 396, (2020).
18. Rehfuess, E., Bruce, N. & Smith, K. Solid Fuel Use: Health Effect. 13.
19. Masera, O., Ghilardi, A., Drigo, R. & Angel Trossero, M. WISDOM: A GIS-based supply demand mapping tool for woodfuel management. Biomass Bioenergy 30, 618–637 (2006).

20. Faraway, J. J. Linear Models with R. 255.
21. Bhatt, S. et al. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. J. R. Soc. Interface 14, 20170520 (2017).
22. Osgood-Zimmerman, A. et al. Mapping child growth failure in Africa between 2000 and 2015. Nature 555, 41–47 (2018).
23. Shupler, M. et al. Global estimation of exposure to fine particulate matter (PM2.5) from household air pollution. Environ. Int. 120, 354–363 (2018).
24. Reiner, R. C. et al. Identifying residual hotspots and mapping lower respiratory infection morbidity and mortality in African children from 2000 to 2017. Nat. Microbiol. 4, 2310–2318 (2019).
25. Friedman, J. et al. Measuring and forecasting progress towards the education-related SDG targets. Nature 580, 636–639 (2020).
26. Forouzanfar, M. H. et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 388, 1659–1724 (2016).
27. Bonjour, S. et al. Solid Fuel Use for Household Cooking: Country and Regional Estimates for 1980–2010. Environ. Health Perspect. 121, 784–790 (2013).
28. Team, W. H. O. O. and E. H. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide : global update 2005 : summary of risk assessment. Lignes directrices OMS relatives à la qualité de l'air : particules, ozone, dioxyde d'azote et dioxyde de soufre : mise à jour mondiale 2005 : synthèse de l'évaluation des risques (2006).
29. Bassat, Q. et al. The first Global Pneumonia Forum: recommendations in the time of coronavirus. Lancet Glob. Health 8, e762–e763 (2020).
30. Bailis, R., Drigo, R., Ghilardi, A. & Masera, O. The carbon footprint of traditional woodfuels. Nat. Clim. Change 5, 266–272 (2015).
31. Muller, C. & Yan, H. Household fuel use in developing countries: Review of theory and evidence. Energy Econ. 70, 429–439 (2018).
32. Lewis Jessica J. & Pattanayak Subhrendu K. Who Adopts Improved Fuels and Cookstoves? A Systematic Review. Environ. Health Perspect. 120, 637–645 (2012).
33. Puzzolo, E., Pope, D., Stanistreet, D., Rehfuess, E. A. & Bruce, N. G. Clean fuels for resource-poor settings: A systematic review of barriers and enablers to adoption and sustained use. Environ. Res. 146, 218–234 (2016).

34. Pachauri, S. & Jiang, L. The household energy transition in India and China. Energy Policy 36, 4022–4035 (2008).
35. Mani, S., Jain, A., Tripathi, S. & Gould, C. F. The drivers of sustained use of liquified petroleum gas in India. Nat. Energy 5, 450–457 (2020).
36. Kar, A., Pachauri, S., Bailis, R. & Zerriffi, H. Capital cost subsidies through India's Ujjwala cooking gas programme promote rapid adoption of liquefied petroleum gas but not regular use. Nat. Energy 5, 125–126 (2020).
37. Ruiz-Mercado, I. & Masera, O. Patterns of Stove Use in the Context of Fuel–Device Stacking: Rationale and Implications. EcoHealth 12, 42–56 (2015).
38. Shankar, A. V. et al. Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. Energy Policy 141, 111468 (2020).
39. Shupler, M. et al. Household and personal air pollution exposure measurements from 120 communities in eight countries: results from the PURE-AIR study. Lancet Planet. Health 4, e451–e462 (2020).
40. Tusting, L. S. et al. Mapping changes in housing in sub-Saharan Africa from 2000 to 2015. Nature 568, 391–394 (2019).
41. Chipeta, M. Mapping local variation in household overcrowding across Africa from 2000 to 2018: a modelling study [Under submission]. Lancet Planet. Health [48, 84]., (2021).
42. Arku, R. E. et al. Adverse health impacts of cooking with kerosene: A multi-country analysis within the Prospective Urban and Rural Epidemiology Study. Environ. Res. 109851 (2020) doi:10.1016/j.envres.2020.109851.
43. Tao, S. et al. Quantifying the rural residential energy transition in China from 1992 to 2012 through a representative national survey. Nat. Energy 3, 567–573 (2018).
44. Kerimray, A., Rojas-Solórzano, L., Amouei Torkmahalleh, M., Hopke, P. K. & Ó Gallachóir, B. P. Coal use for residential heating: Patterns, health implications and lessons learned. Energy Sustain. Dev. 40, 19–30 (2017).
45. Chen, Y. et al. Estimating household air pollution exposures and health impacts from space heating in rural China. Environ. Int. 119, 117–124 (2018).
46. Kimemia, D. & Annegarn, H. Domestic LPG interventions in South Africa: Challenges and lessons. Energy Policy 93, 150–156 (2016).
47. Amegah, A. K. & Jaakkola, J. J. Household air pollution and the sustainable development goals. Bull. World Health Organ. 94, 215–221 (2016).
48. Xu, X., Ha, S. U. & Basnet, R. A Review of Epidemiological Research on Adverse Neurological Effects of Exposure to Ambient Air Pollution. Front. Public Health 4, (2016).

49. Naidoo, R. & Fisher, B. Reset Sustainable Development Goals for a pandemic world. Nature 583, 198–201 (2020).
50. Leal Filho, W., Brandli, L. L., Lange Salvia, A., Rayman-Bacchus, L. & Platje, J. COVID-19 and the UN Sustainable Development Goals: Threat to Solidarity or an Opportunity? Sustainability 12, 5343 (2020).
51. Fullman, N. et al. Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. Lancet 391, 2236–2271 (2018).

Chapter 2: *Scales of Environmental Justice: Global Sensitivity Analysis of the Washington Environmental Health Disparities Map*

Abstract

Background: Theoretically-derived models of cumulative impact are increasingly applied to synthesize environmental and population health datasets, generating composite indicators used in benchmarking and policy-making.

Methods: We conducted a global sensitivity analysis to validate the methodology of the Washington Environmental Health Disparities (EHD) map by permuting across alternative methods derived from the composite indicator construction literature, generating estimates of the uncertainty that results from analyst decisions in the development process. We estimated variance-based first, second and total order Sobol sensitivity statistics to quantify the relative influence of parameter choices on the tract rankings and on the accuracy of classifying communities in the top 20% of impact.

Results: On average, the rankings of census tracts changed by more than one hundred ranks across these permutations. In comparison to the baseline EHD index, variation was largest in the middle of the impact spectrum and smallest for tracts in the top 10% of impact. The formula used in aggregation and the method of data normalization were the most sensitive parameter decisions for both tract ranking and impact classification.

Discussion: The EHD ranks are observed to be more robust in the highest impact tracts and relatively uncertain throughout the rest of the index, suggesting that this data is better suited for classifying hotspots than for quantifying ordinal impact. There are strong assumptions underlying the baseline EHD ranking methodology and these assumptions substantially drive the results.

Conclusion: Architects of environmental justice indicators must carefully consider the assumptions made while normalizing indicator data and selecting an aggregation formula, so that the uncertainties imparted by analyst decisions may be better understood and transparently communicated to policy-makers and the community.

Intro

The association between the health of an environment and its inhabitants is a fundamental concept in epidemiology.[1] Environmental factors are understood to impact human health through a variety of direct and indirect pathways, and disparities in those factors will manifest in health inequalities at the population level.[2] Given the complex and synergistic nature of these relationships, multiple frameworks have been developed with the aim of systematically understanding the pathways through which environmental health impacts occur, including the exposure-disease-stress model and the World Health Organization's (WHO) drivers-pressure-state-exposure-effect-action (DPSEEA) framework.[3,4] These frameworks have been used to derive numerous environmental health indicators, through which key drivers of this process are quantified for the purposes of benchmarking and informing policy-making efforts.[5] A recurrent conclusion of these frameworks is the distinction between direct effects (exposures) and more distal factors (vulnerabilities), with the acknowledgement that these two categories are cumulative and have multiplicative interactions that should be incorporated in order to properly assess risk[6]. The inherent multidimensionality of environmental factors and growing availability of broad sources for high-quality data lends itself to the development of composite metrics in order to decrease complexity, track changes over time, aid comparability, and facilitate interpretation and policy translation[7].

Developing composite indicators is a well-established method to reduce the dimensionality of models where proxy variables are being used to quantify latent factors that are infeasible to measure outright. This approach is not without controversy, as in any mathematical model, combining disparate metrics requires analysts to make subjective decisions at every step of the process, introducing assumptions that have been demonstrated to have substantial impacts on the distribution of resulting indices[8]. As a result, methods have been developed to analyze the uncertainty (capturing how sub-indicator uncertainty propagates during the aggregation process) and sensitivity (quantifying the differential level of impact for each uncertainty source on the overall variance of the index) of composite indicators in order to carry out quality assessment and to evaluate the consequences of analytical assumptions therein[9].

The core methodological assumptions required for developing a composite indicator include variable selection/imputation, normalization, weighting scheme, and aggregation formula[10]. Based on the landscape of data available and at the appropriate spatial resolution, variable selection can be performed using a conceptual model or with data mining techniques[11]. When selected variables are drawn from a diverse array of instruments and data-generating processes, they will present in different units of measurement and require normalization prior to any synthesis. The selections could be normalized by ranking, z-score standardization, min-max normalization, or distance to target approaches[12]. The variables are then weighted, sometimes equally, or through a weighting schema that is derived from literature values, expert groups or stakeholder assumptions, or data-driven using methods such as the variance-based Pearson correlation ratio[13]. The formula for aggregating up to a final index is commonly inductive, using dimensionality reduction techniques such as principal components analysis (PCA)[14] or based on a deductive approach, which can also be formulated hierarchically[15]. Given that environmental health is understood to be multidimensional, latent, and therefore not directly observable, permuting the decision matrix of those assumptions to analyze sensitivity and uncertainty is critical to robustly validating composite indicators[16].

Conducting a local sensitivity analysis is the simplest way to quantitatively compare different options at a given stage, by holding other steps as fixed and analyzing the resulting index permutations through a metric such as the correlation coefficient - where higher correlation demonstrates insensitivity of the index to methodological assumptions made for that stage[16]. A key limitation of this approach is that local sensitivity analyses are only able to discretely examine individual stages, meaning they will underexplore the sample space and are thus particularly inefficient for composite indicators where the ratio of influential to uninfluential permutations is high and unable to account for interactive effects[17,18]. Alternatively, a global sensitivity analysis may be employed, permuting all possibilities jointly in a Monte Carlo simulation, and decomposing variance using bootstrapping analysis[19,20]. Global sensitivity analyses have been conducted for a variety of environmental and socioeconomic composite

indices, but never before to our knowledge has this approach been used in validating an index for environmental justice[9,21,22].

Beginning in the 1980s, the environmental justice movement developed to center equity as a core tenet for environmental health policy discussions, with community-driven initiatives to recognize and redistribute the uneven health burden of environmental degradation[23]. To benchmark and quantitatively inform these efforts, several different methods have been used to develop environmental justice indices, including national level indicators such as the Environmental Justice Screening and Mapping Tool (EJSCREEN)[24]. This was followed by more localized methods, such as the California Communities Environmental Health Screening Tool (CalEnviroScreen)[25], based - as of its fourth iteration - on a hierarchical combination of 21 different state-wide indicators. In Washington state, a novel community-academic-government partnership led to the utilization of similar methods in developing the Environmental Health Disparities (EHD) Map[26]. In 2021, the historic passage of Washington's Healthy Environment for All (HEAL) Act institutionalized environmental justice values as a guiding principle of state-level policy, calling for integration of these concepts for decision-making and underscoring the critical importance of robust tools for measurement and evaluation in the process. Of particular importance is the identification and classification of highly-impacted communities, with their eligibility for funding targeted to remediation and harm-reduction supporting the emphasis on accurate monitoring.

The impact of methodological assumptions must be understood in order to accurately translate health information into policies that most effectively target environmental inequality. There are two central questions to this line of inquiry - first, how confident should we be in the results of our composite indicator (uncertainty), and which elements of the construction are most influential in driving this uncertainty (sensitivity). In this study, we analyze the framework and model used to produce the Washington State EHD Map, quantifying the uncertainty of the EHD ranking scores and the sensitivity of the underlying assumptions necessary to calculation of that index. We employed a global sensitivity analysis, simultaneously permuting the available and relevant methods for each stage of the index construction, in order to capture the effect

of interactions on index uncertainty[27]. A multidimensional parameter space was derived from the matrix of relevant decisions for each step of the process and then sampled quasi-randomly using Sobol sequencing, permuting through indicator selection, normalization, weighting, aggregation, and classification of highly-impacted communities. For each permutation, the selected approach was compared against the baseline ranking in order to calculate the mean average deviation in ranking. Furthermore, the classification accuracy of the model was tested by calculating the probability that the most impacted tracts in a given permutation would match the baseline classification. Variance from this analysis was decomposed to generate Sobol sensitivity scores, guiding our recommendations for the construction of future environmental justice indices and helping to inform decision-makers when interpreting the index in its current formulation.

Methods

The Washington State Environmental Health Disparities Map

The baseline for comparison in this sensitivity analysis is the second version of the Washington State Environmental Health Disparities Map, a composite indicator developed within a community-driven government-academic partnership to estimate cumulative environmental risk statewide at the census tract level. The EHD map was originally released in 2019, with a second iteration in 2022. EHD was formulated based on the previously established CalEnviroScreen method, which similarly employed a cumulative impact scoring framework to generate a single composite ranking. The design of the EHD model required the selection of relevant indicators, normalization to a standard unit scale, weighting, hierarchical deductive aggregation, and finally classification into decile rankings of which the top 20% represented highly impacted communities. The specific parameters for each of these steps for the sensitivity analysis and for the baseline EHD ranking in the following Methods text.

Indicator selection

Nineteen different indicators were selected for inclusion in the EHD ranking. These include five indicators capturing environmental exposures (diesel emissions, ozone, $PM_{2.5}$, toxic releases from facilities, and traffic density), five indicators capturing environmental effects (lead risk and exposure, proximity to hazardous waste generators and facilities, proximity to Superfund sites, proximity to facilities with highly toxic substances, and wastewater discharge), two indicators capturing sensitive populations (rate of cardiovascular disease and rate of low birth weight infants), and finally seven indicators capturing socioeconomic factors (low educational attainment, housing burden, linguistic isolation, poverty, race (percentage people of color), transportation expense, and unemployment). These indicators were hierarchically classified by those four groups, and then further classified with an additional level of hierarchy (Figure 1), where the first two groups representing threats and the latter two representing vulnerabilities, with overall risk calculated by multiplying the aggregated scores for those two groups[26]. In order to investigate the influence of indicator selection, we permuted the index by dropping the indicators, one at a time. Likewise, we examined the influence of indicator development by analyzing the response of this index to differential measurement error in these metrics, by additively simulating uniformly-distributed random noise to each of the indicators as part of the parameter space.

Normalization

These nineteen indicators were measured on a variety of different scales, as such they have incompatible units (from proportions, to population rates, to continuous measurements) and must be normalized prior to aggregation. The baseline EHD model accomplished this task by rank-ordering the data for each indicator and then binning it into ten groups, converting to an ordinal classification. This is the simplest normalization method, with the advantage being that it is more intuitive and very robust to outliers, but in the same vein, it compresses the distribution, removes information about the measured variance and makes tracking progress more difficult as the level values cannot be preserved[10].

Z-score standardization is an alternative, in which indicators are scaled such that the values are distributed with a mean of zero and a standard deviation of one. Outliers are more influential when using this method, as indicators that have extreme values will tend to contribute more to the composite - a useful quality if the goal of the index is to identify exemplars or severe inequalities. Re-scaling is a similar technique, except that the indicator values are adjusted to create distributions that share a common range from zero to one. Generally, composite indicators that use one of these two approaches choose between them based on the aggregation method, with deductive aggregators defaulting to re-scaling and inductive aggregators preferring to apply a z-score standardization[16].

Weighting

Sub-indicators are generally weighted prior to aggregation, either based on analyst or expert knowledge on the relative importance of each sub-indicator to the latent data-generating process that the composite indicator is being designed to capture (normative), or using a statistical approach (data-driven). Normative weights can be divided into approaches that are equally-weighted or arbitrarily-weighted. Equal weights is the most common and the simplest approach, skirting the need for analysts to quantify the importance of each indicator by assuming that they are all equally important or by viewing the exercise agnostically - a tacit admission that data or theory gaps make it infeasible to accurately determine the relative values[28]. Alternatively, if the relative importance of each indicator is understood to be known by the index developers, the weights can be set arbitrarily to reflect that differential importance in the aggregation step. The baseline EHD index used a hierarchical deductive approach with 2 additional levels of hierarchy, composed at the first level of aggregation with 4 sub-categories, where the individual indicators for each sub-category were equally weighted. Then, the 4 subcategories were aggregated up into two larger sub-categories, all of which were equally weighted except the sub-category of environmental effects, which was arbitrarily down-weighted by half in order to capture the relatively distal proximity of effects when compared to exposures.

If the relative importance of each indicator is not quantifiable or well-understood, data-driven approaches can be used to exploit the distributions of the variables in order to estimate appropriate weights. One method is to use the indicator correlations to generate inverse weights for each indicator, which can prevent double-counting that arises from positively correlated indicators reinforcing each other while uncorrelated ones are downweighted[29]. This is a desirable quality in indices where the sub-variables are understood to represent different elements of the latent factor. A similarly motivated approach for reducing double-counting is to employ a multivariate statistical method such as principal components analysis (PCA), which calculates linear combinations of the variables that maximize the variance explained[10]. This ensures that each principal component explains different statistical elements of the entire dataset, which can be thought of as potentially corresponding to different sub-categories of the latent effect the index is measuring. The resulting principal components themselves can aggregate the final index (inductive approach), or the variable loadings can be input to be the weights for a different aggregation formula. The loadings can either be restricted to the first component on the rationale that it explains the most variation, or an eigenvalue threshold of 1 can be employed to select significantly explanatory components and then aggregate their loadings into weights by weighting each component itself by the percent of variance explained[10].

Aggregation

Numerous sensitivity analyses indicate that the aggregation formula is generally the most impactful assumption for the creation of composite indicators[16,21]. The baseline EHD calculation employed a deductive hierarchical approach, with two levels of hierarchy. The first level was linearly summed with the arithmetic mean to produce two sub-indicators. These two sub-indicators were summed multiplicatively based on an established risk-scoring model[30]:

$$Risk = Threat \times Vulnerability$$

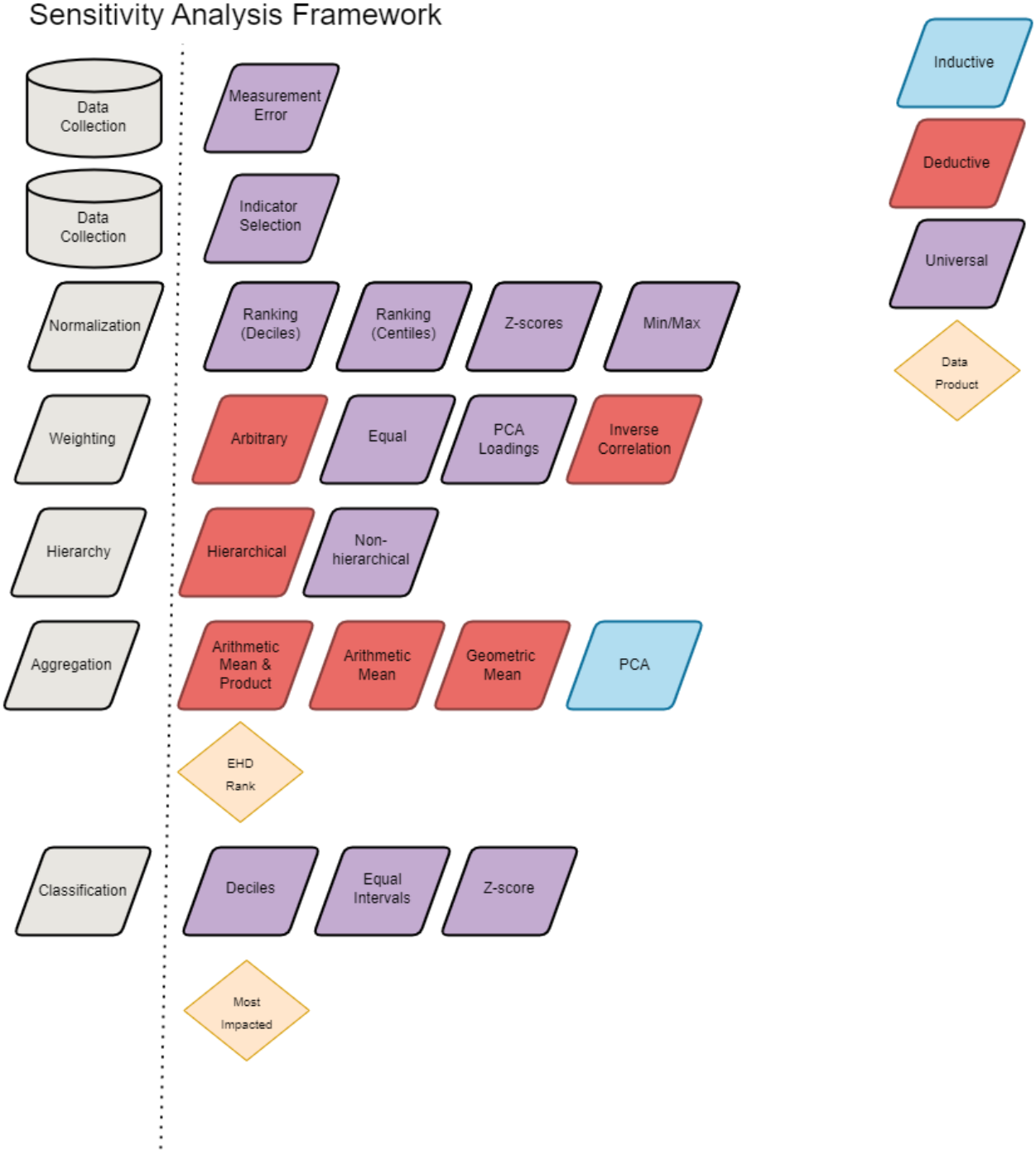
That equation was modified such that the population characteristics sub-indicator (composed of the aggregation of variables for sensitive populations and socioeconomic factors) captured population *vulnerability*, and the pollution burden sub-indicator

(composed by aggregating the environmental effects and environmental exposures indicators) represented environmental *threats*. To quantify the impact of the chosen hierarchy, we also employed a non-hierarchical deductive approach, where all indicators were summed by their arithmetic mean as a single step on the same level. We also tested using geometric means to aggregate in both hierarchical and non-hierarchical approaches. Finally, an alternative method, popularized by the social vulnerability to environmental hazards (SoVI) index, is a more data-driven, inductive aggregation[14]. There are a number of multivariate statistical techniques that can be applied within this paradigm, but the most common is to use PCA.

Classification

Classification is an important final step of composite indicator construction, typically employed for visualization or post-analysis of the ranked outputs. The relevance of a threshold for highly impacted communities to environmental health policy-making demonstrates the importance of carefully choosing the method for classifying the resulting indices. Equal intervals is a common approach, where the ranks are binned into a set number of classes such that each class has the same range. Using quantiles is a similar approach, except bins are constructed where each class has the same number of units - the baseline EHD index categories were created by binning the ranks into deciles, which is aligned with the decile ranking method employed for data normalization. Standard deviation based approaches, like z-scoring, can also be employed to account for the resulting distribution of the composite indicator when classifying the outputs. In this method, a standard deviation of less than or greater than 1.5 is often used to classify the high and low impact areas, with a threshold of less than or greater than .5 classifying moderate impacts[31]. To align our permutation with the number of tracts being identified in the baseline EHD, we used a z-score threshold of 0.842, representing the 80th percentile of a normal distribution.

Figure 1 - Global Sensitivity Analysis Framework



The methodological framework of the global sensitivity analysis is presented as a flowchart, where the parallelograms on the left side of the dotted line represent the different steps of composite indicator construction, and the parallelograms on the right side represent the different possible choices being sampled in the multidimensional parameter space. The diamonds represent the data products, both final and intermediate being produced during the construction process. The colors represent the applicability of that parameter selection to all aggregation methods (purple) or to just the inductive (blue) or deductive (red) aggregations. Permutations that sampled incompatible methods were excluded from analysis.

Global Sensitivity/Uncertainty Analysis

In order to test the sensitivity and uncertainty of the EHD across the universe of potential choices for composite indicator construction, we permuted applicable options as part of a global sensitivity analysis (see Figure 1). We used a low-discrepancy Sobol sequence to iteratively permute quasi-random samples from a multidimensional sample space that was constructed by drawing from the input distribution of each relevant model parameter. From the resulting permutations, we estimated the mean absolute change in ranking (MARC) and the probability of agreeing with the baseline index to classify the top 20% of tracts as highly impacted areas (accuracy):

$$MARC = \frac{1}{n}\sum_{i=1}^{n}\left|rank_{baseline} - rank_{i}\right|$$

We also estimated by area the 95% confidence interval across permutations to quantify the underlying uncertainty of each census tract. Finally, we used the estimated distributions of EHD rankings from the permutation analysis in a variance-based sensitivity analysis by calculating Sobol sensitivity indices. This approach decomposes the modeled variance in order to understand the sensitivity of the index to fluctuating each model parameter[22]. We calculated first order (S_i) and total order (S_{Ti}) effects using the following estimators:

$$S_i = \frac{V[E(y|x_i)]}{V(y)}$$

$$S_{Ti} = 1 - \frac{V[E(y \mid \mathbf{x}_{-i})]}{V(y)} = \frac{E[V(y \mid \mathbf{x}_{-i})]}{V(y)}$$

where *i* represents the parameter of interest, and X_{-i} represents the set of all varied parameters except parameter *i.* These estimators can be expressed as the fractional reduction in variance expected if parameter *i* could be fixed.

Results

Uncertainty

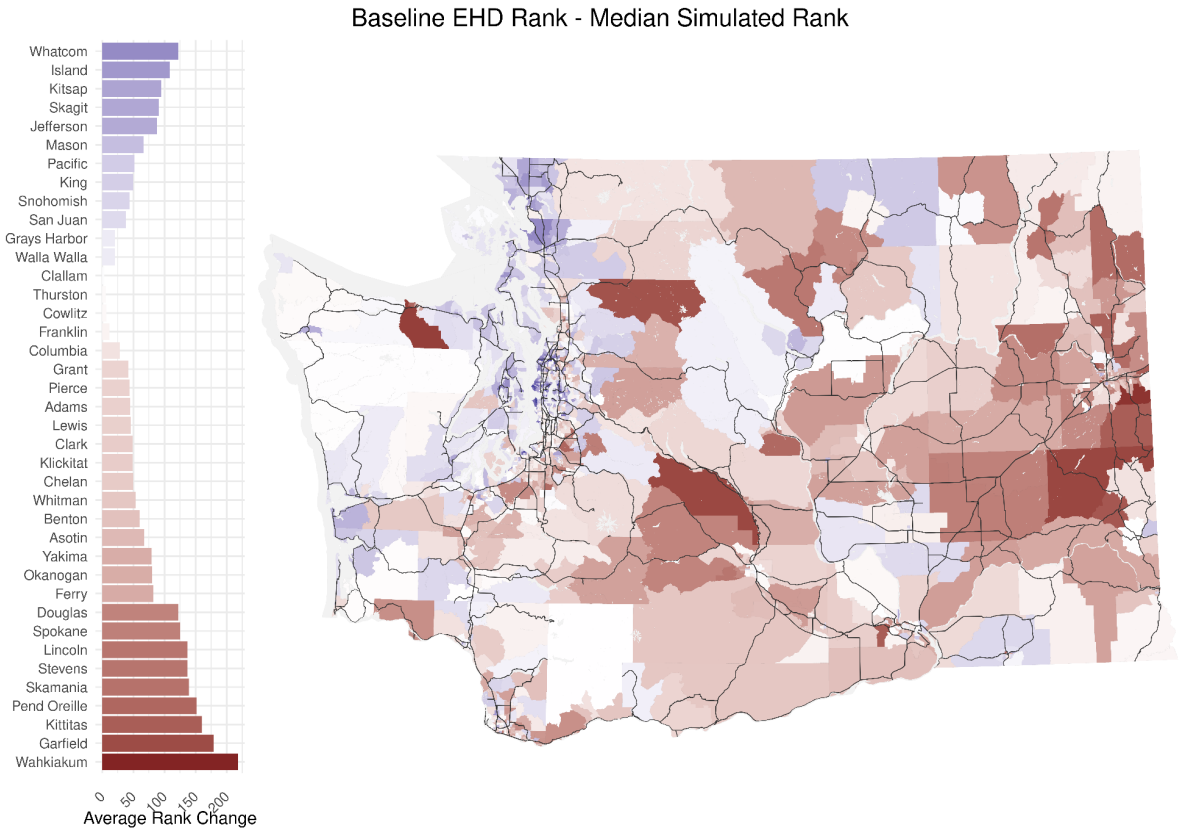
In order to ensure adequate sampling of this parameter space, we drew from each of the 19 methodological options 5000 times, generating 115,000 permutations of the EHD index. Across the 1,458 census tracts in Washington state, the population-weighted average rank change (*Figure 2a*) in comparison to the EHD version 2.0 rankings was 118, and that 25% of tracts moved by more than 165 spots. However the change in rankings were not uniformly distributed across the ordinal ranks from the baseline EHD index (*Figure 3a),* suggesting that varying the methods for constructing this index can lead to larger rank deviations for census tracts that are currently being classified mid to low impact (*Figure 6b)*. The tract with the largest rank decrease was located in Walla Walla county, falling from 582nd in the baseline EHD to 180th in the median ranking across permutations. The tract with the largest increase was located in King County, where a tract rose from 442nd in the baseline EHD to 976th in the median ranking across permutations. Across the entire county *(Figure 2a)*, the county with the largest population-weighted average increase in ranking was Wahkiakum County (-218), while Whatcom County had the largest average decrease (168). The counties that tended to be most stable across permutations were Columbia (population-weighted average deviation of -7) and Thurston (-2).

The 95% confidence interval of the various permutations provides a metric to estimate the uncertainty that is introduced to this index by inherent methodological assumptions (*Figure 2b).* The median range of the 95% confidence interval across the state was 833 ranks, suggesting that the choices made in constructing this composite indicator have a substantial impact on the ranking generated for any given census tract. Geographically, this analysis indicates that the areas with the highest uncertainty in Washington State are in the rural north-central area of the state, comprising Okanogan, Ferry, Chelan, and Douglas counties, while San Juan county stood out as being the

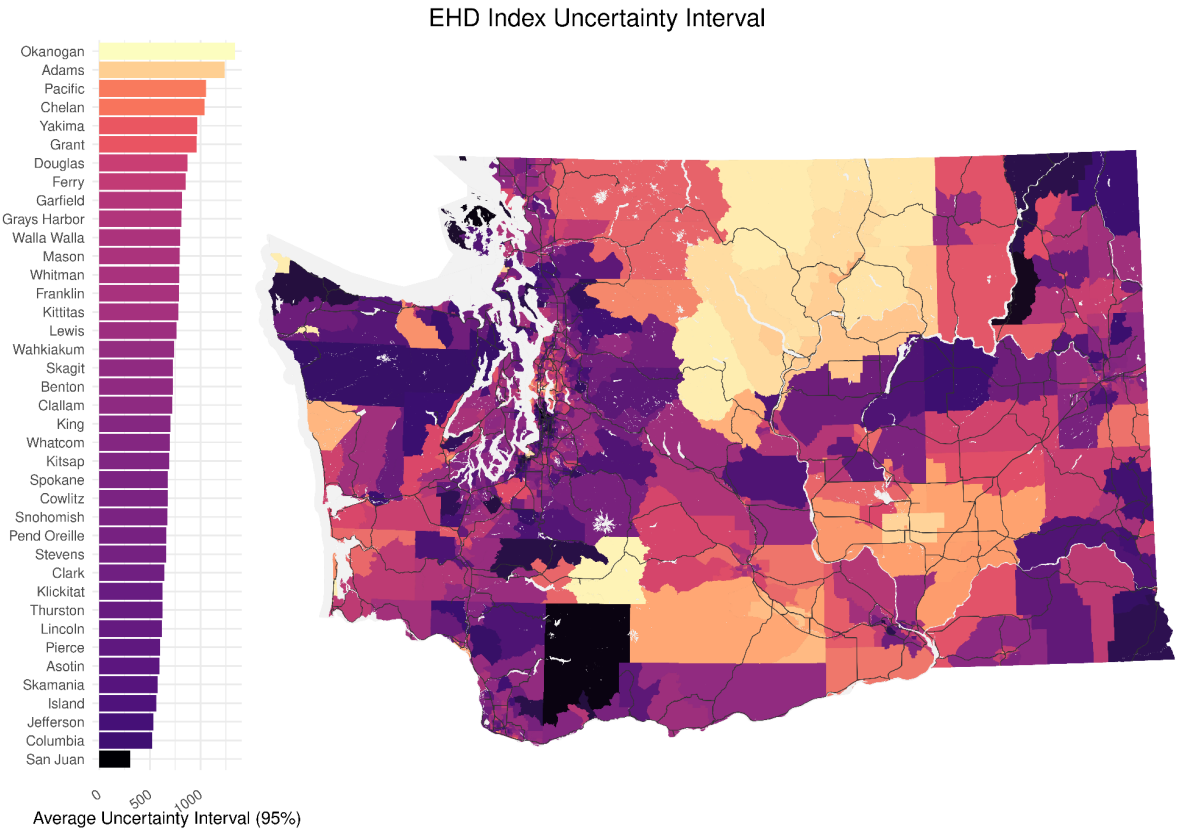
most certain. In contrast to the rank deviations, we observed that the index tended to be most uncertain in the center of the distribution, while the extremes have higher confidence as measured by sensitivity to deviations given methodological changes *(Figure 3)*.

Figure 2: Tract-level maps of mean deviation, uncertainty range, and bivariate uncertainty as a function of the baseline EHD rank.

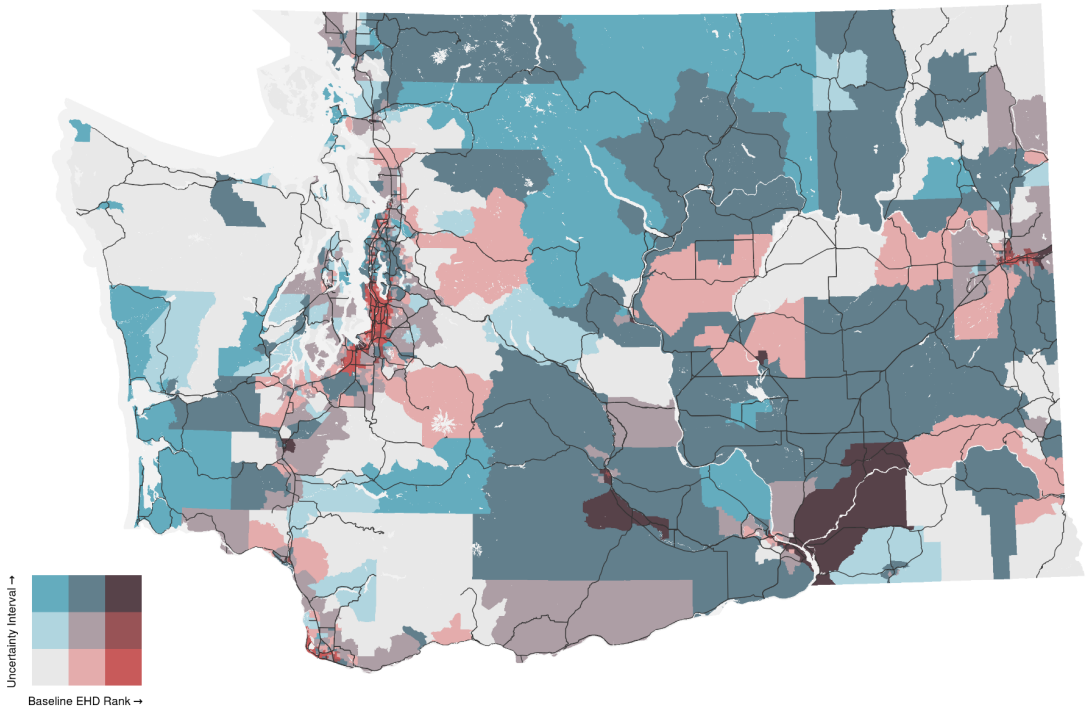
a)



b)


EHD Index Uncertainty Interval
Okanogan
Adams
Pacific
Chelan
Yakima
Grant
Douglas
Ferry
Garfield
Grays Harbor
Walla Walla
Mason
Whitman
Franklin
Kittitas
Lewis
Wahkiakum
Skagit
Benton
Clallam
King
Whatcom
Kitsap
Spokane
Cowlitz
Snohomish
Pend Oreille
Stevens
Clark
Klickitat
Thurston
Lincoln
Pierce
Asotin
Skamania
Island
Jefferson
Columbia
San Juan
0
500
1000
Average Uncertainty Interval (95%)

c)

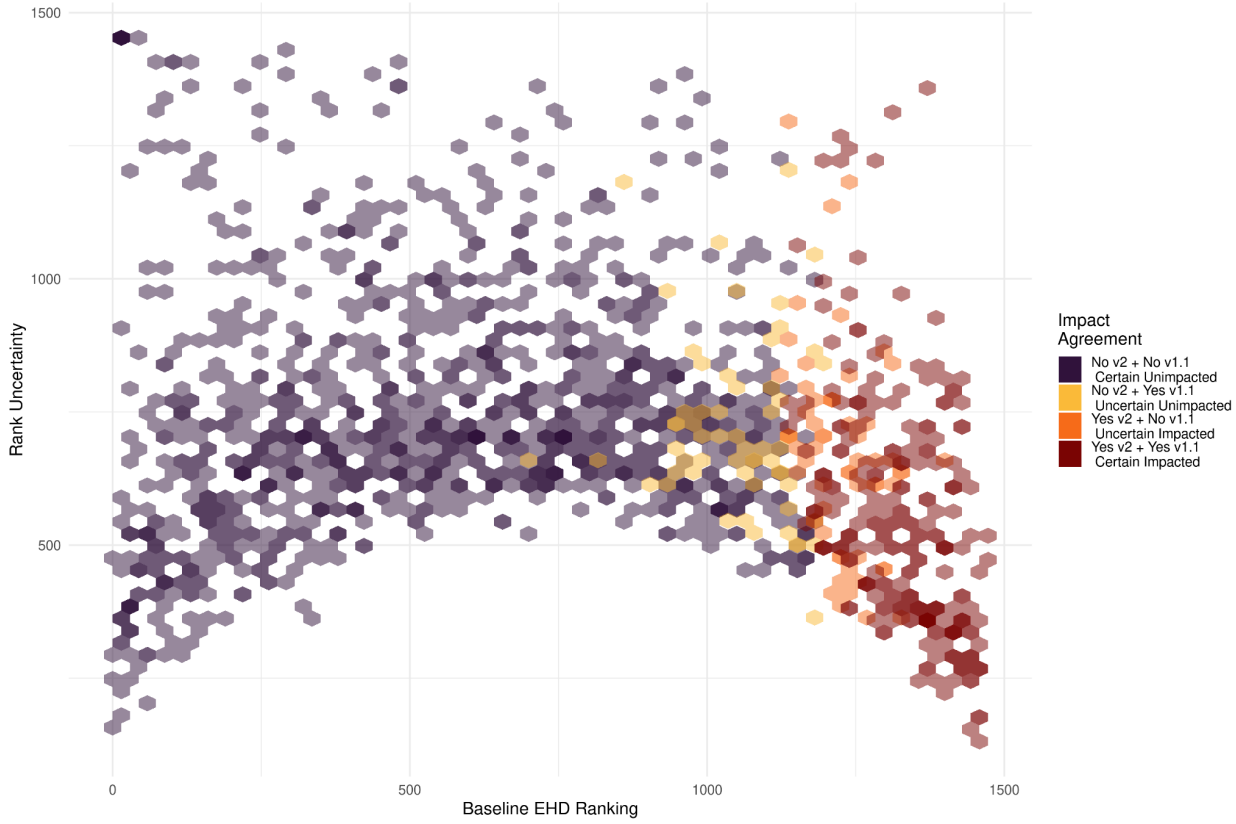


The results of the global sensitivity analysis at the census tract level are displayed as choropleths of the mean change in rank from the baseline EHD ranking **(a)**, rank uncertainty as measured by the range of the 95% confidence interval of all permutations **(b)**. Here, the left inset histograms display the population-weighted averages at the county level on the same color scale as the maps. For the change plots, red denotes increases (where the median rankings from the permutations were larger than those of the baseline EHD) and the blue decreases. The bottom map uses a bivariate color scale to display the relationship between baseline EHD ranking and rank uncertainty **(c)**. The bivariate's blue gradient across the y axis denotes increasing uncertainty and the red gradient of the x axis the decile rankings of the baseline EHD binned from 1-2, 3-8, and 9-10 (high impact tracts classification). Bright red tracts are those that were classified as high impact with high certainty, while dark purple tracts were classified as high impact but observed large uncertainty according to the global sensitivity analysis. The teal tracts are those that were classified in the lowest impacted according to the baseline EHD but are sensitive to methodological change and therefore more uncertain, while the gray tracts are classified with high certainty as very low impact.

However, there were certainly outliers, especially on the lower end of the impact spectrum, where census tracts could be ranked in essentially any part of the index based on which methods were chosen in construction. The census tract with the highest certainty was located in Grays Harbor County (ranging across 1457 spots in the ranking

based on methodological parameters), while the lowest was in San Juan county (88 rankings). Simultaneously mapping the baseline EHD ranking and uncertainty interval for each census tract (*Figure 2c*) gives a sense for how confidently tracts can be classified across the spectrum of cumulative impact. South king county is highly impacted with relatively high confidence, while high-impact tracts in east Spokane, Yakima county and the near the Tri-Cities area can range dramatically based on which model is selected. Likewise, areas of north-central Washington and the southwest coast are currently being classified in the bottom 20% of cumulative impact but observe substantial variation in ranking across the range of permutations.

Figure 3: Global rank uncertainty as a function of the baseline EHD v2.0 ranking



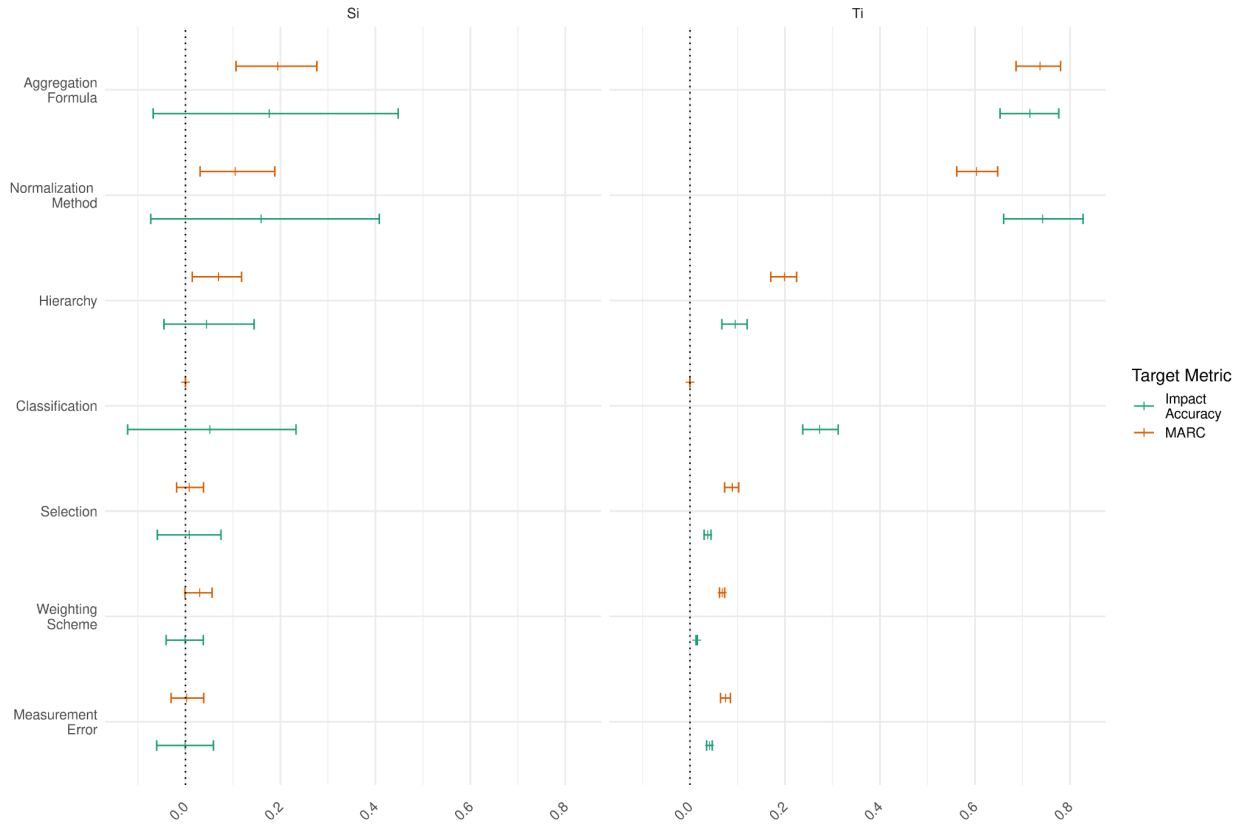
The results of the global sensitivity analysis at the census tract level are displayed as scattered hexagons where increased transparency of the hexagon represents the density of points in that plot location. The Y axis represents the range of the 95% confidence interval across all simulations and the X axis represents the tract-level EHD ranking in the baseline index. The colors represent the agreement in classifying the tracts as highly impacted (top 20%), where purple are those that were classified as unimpacted in both of the releases of the EHD rankings

and red are those that were classified as impacted in both. The yellow and orange hexagons are those with a classification status that changed in between versions of the EHD and can be considered to be less certain based on that disagreement. We observe a parabolic shape where the range of values across the simulations is considerably greater in the center of the index, and more confident at the extremes. The least sensitive area on the impact spectrum is in the top 10%, where both versions of the index tended to agree and the rankings were less subject to change based on varying the methods used to compute the index.

Sensitivity

Our primary variance-based global sensitivity analysis used the mean absolute rank change (MARC) as the target metric to generate sensitivity indices for the first, second (interaction), and total order effects of parameter choices in each of the steps of constructing this composite index. In both the first and total order effect indices *(Figure 4a),* the choice of aggregation formula was significantly more influential than all other parameters, with an effect of 0.077 (0.028-0.127) individually (Si) and an effect of 0.77 (0.650-0.898) when considering all interactions (Ti) with other methodological steps. Normalization was the second most important parameter for both Si (0.105; 0.031-0.188) and Ti (0.603; 0.687-0.781) The choice of whether or not to enforce a hierarchy also had a large total effect (0.199; 0.170-0.224) , followed by indicator selection, weighting scheme, and measurement error. The latter three parameters observed total effects of less than 0.1 but the overlaps of their confidence intervals mean that this model cannot definitively state which is more influential. Normalization and aggregation also had the largest second-order interaction effects, with their interaction being the largest by far, and the interactions between aggregation formula and selection and measurement error the next most influential.

Figure 4: Global Sensitivity Indices

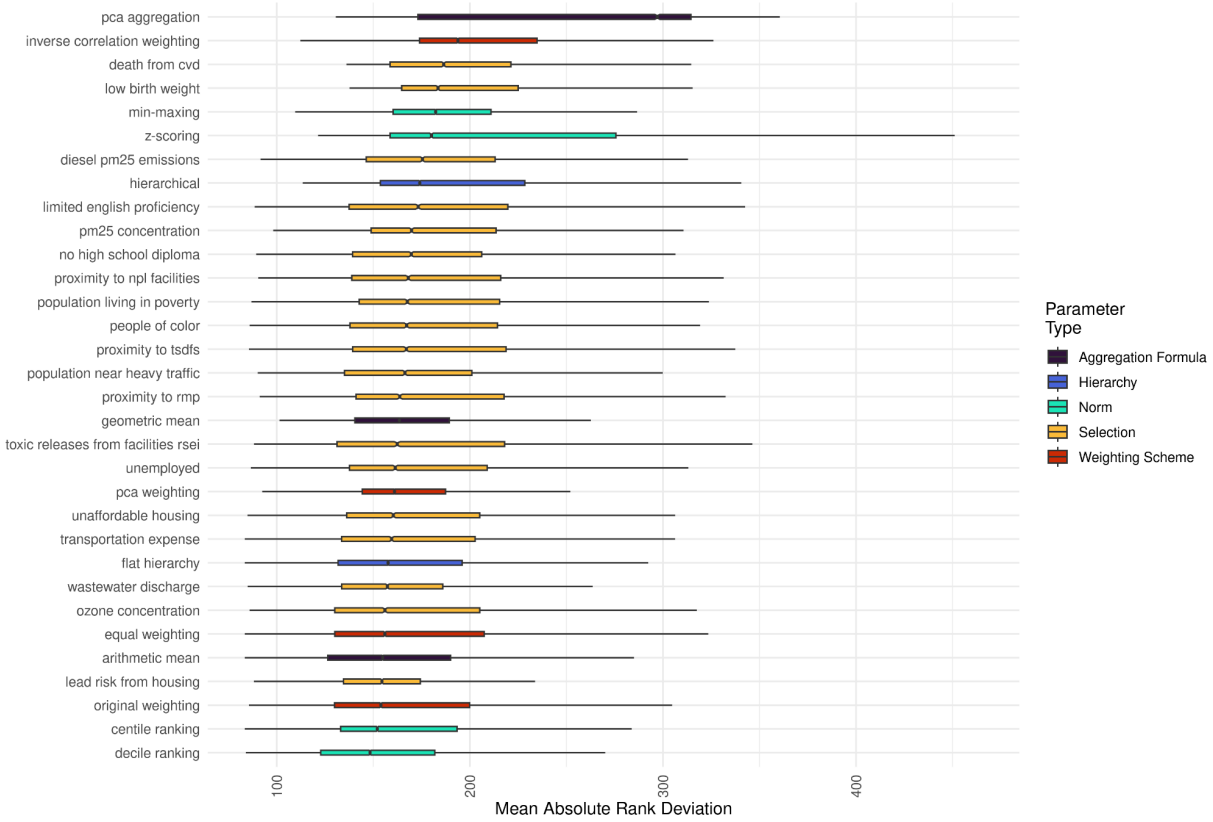


The mean (tick) and 95% confidence interval (bar) of the first-order (Si) and total-order (Ti) effects of the Sobol sensitivity indices are displayed. The Si and Ti effects can be understood as the proportion of the total variance in ranking **(a)** or impact classification **(b)** that could be reduced by fixing that parameter.

Examining the distribution of MARCs by individual parameter choices (*Figure 5*) allows for intuition on which formulations of the composite indicator are most different from the current EHD ranking system. We observed that the largest deviations were found in models that used an inductive PCA aggregation formula, instead of the deductive hierarchical aggregation in the baseline EHD. The next cluster of large deviations were found in models that used inverse correlation weighting or linear normalization methods like min-max scaling or z-scores, and in models that dropped the low birth weight or death from cardiovascular disease indicators. The most similar model permutations were those using rank-based normalizations, either the original weighting schema or an equal weighting scheme for the sub-indices, and models that

dropped the least influential indicators of lead risk from housing, ozone concentration, or wastewater discharge.

Figure 5: MARC by selected parameter



The distributions for the mean absolute rank change (MARC) across all permutations are shown for each individual parameter selection, where the density is shown as a semi-transparent violin and the interquartile range is shown as a solid boxplot, with the average MARC is found notched into the boxplot. The colors denote which exact step of the construction process that each parameter selection occurs in. On average, permutations that used PCA for an inductive aggregation deviate the most from the baseline EHD, while those that used a decile rank normalization method are most similar. We observe that permutations using inverse correlation weighting, min-max or z-score normalization, or that dropped either of the indicators from the Sensitive Populations grouping (low birth weight or death from cardiovascular disease) exhibit the next largest deviations in rank.

An additional sensitivity analysis was conducted using the classification accuracy for most impacted tracts, based on the probability that a given permutation classified the highly impacted tracts equivalently to the baseline EHD index. Here, the method used to classify the most impacted tracts was also considered, and was the third most influential

parameter by Ti effect (0.273; 0.237-0.312). In this analysis, however, the normalization method was most influential, followed by the aggregation formula, with Ti effects of 0.743 (0.661-0.828) and 0.716 (0.653-0.777), respectively. Here choice of hierarchy (0.095; 0.067-0.120), measurement error (0.041; 0.035-0.047), and indicator selection (0.037: 0.030-0.044) all trailed further, identifiably less influential for classifying the outlying tracts. The most important second-order interaction effects were between normalization and aggregation formula (0.324; -0.015-0.635), between normalization and classification (0.034; -0.178-0.252), and between the aggregation formula and the weighting scheme (0.019; -0.015-0.054).

Discussion

The results of this global sensitivity analysis demonstrate how census tract-level rankings derived from the WA State Environmental Health Disparities index are considerably impacted by assumptions throughout the construction of the composite indicator, and these choices are important sources of uncertainty for both the ordinal rankings and for the classification of highly impacted areas for policy targets. We observed that, on average, selecting other methods from alternatives consistent with the literature for constructing composite indicators can shift census tracts more than one hundred spots in the overall ranking, underscoring the critical importance for analysts to carefully vet and transparently communicate these decisions to policy-makers and to the communities engaging with this index. Furthermore the observed deviations in rank across these methodological permutations were not uniform across the distribution of the index, suggesting that tracts that are currently being classified as mid to low impact are most sensitive to which methods are selected for the analysis and could be drawn biased towards the null in alternative but statistically defensible constructions of the index. This index is most robust at estimating the top 10% of tracts, which defends its use for highlighting areas of extreme inequality and suggests that care should be taken when communicating and displaying this index as a continuous measure of environmental health impact.

Not only does this analysis help to quantify the uncertainties that result from inherent decisions made in the construction of a composite index, but a key utility that can be derived from sensitivity analysis is the decomposition of variance into specific parameters in order to estimate the sensitivity of the output to that step of the process. This information helps inform analysts of the relative influence of various decisions, and can be used to focus their finite resources for iterative developments on the most important steps (factor prioritization) and not on uninfluential parameters (model simplification)[16].

We observed the large influence that the choice of aggregation formula has on the EHD rankings. The individual choice decomposition suggests that whether to use a deductive or inductive aggregation is a critical decision, and this analysis makes evident that using a data-driven multivariate regression approach like PCA would heavily modify the results. The inverse correlation weighting permutations followed PCA as the largest deviation from the baseline estimates, demonstrating the important distinction between methods that are informed by the underlying variance of the raw data and those that owe their structure to a top-down, theoretical design. Either of these approaches would be a sizable departure from the EHD system and the preceding statewide indices that it inherits from. The relevance of the ranking sensitivity to those decisions and the resulting estimation of uncertainty depends on how confident the modelers are in the body of scientific knowledge that informed their design of that hierarchical system, specifically how well they believe it to approximate the underlying system through which environmental threats and population vulnerabilities impact health.

The baseline deductive aggregation method also has an implicit weighting structure that derives from the number of indicators in each category, the influence of which is demonstrated by the sensitivity of the rankings to dropping either of the sensitive populations indicators - a category that has only two constituents and therefore the largest individual weight to each. That hybrid normative weighting/aggregation approach currently provokes two notable limitations. First, equal weights are generally understood to introduce bias because it is statistically unlikely that the latent effect that the index is proxying has an equal association with every sub-indicator. Furthermore, arbitrary weighting approaches are criticized for being

paternalistic - even with the strong assumption that the analysts have accurately weighted the sub-indicators relative their average association with the response factor at the population level, if those associations are not normally distributed it could lead to effects that are more important for some groups being underrepresented in the composite indicator[28]. In an environmental justice index, this is a particularly important potential limitation.

Our analysis also illustrates the particular impact of decisions made for which method to use for normalizing the raw indicator data, which is a necessary step in order to standardize indicators that are drawn from disparate unit spaces. The decile ranking system currently being employed in the baseline EHD index is a useful method because it is unaffected by outliers, however this inherently nonlinear approach will thus alter the distribution of the raw data and cause the loss of information on the magnitude of inter-tract differences. This is a particularly notable issue for an environmental justice index, where the scale of differences for some outlying tracts may actually be important for determining cumulative impact. This effect is made more pronounced by the tendency of environmental indicators to exhibit considerable skewness, which manifests in risk accumulation for sub-populations who are generally marginalized due to socioeconomic disparities[32,33]. Likewise, we observe the outsized influence of the classification scheme used to identify the most impacted tracts, where alternative classification schemes like z-scoring (which preserve the distribution of EHD scores) could potentially drive the misclassification of tracts. The policy relevance of that threshold for funding from the HEAL act highlights the importance of that decision as questions continue to be raised about the validity of tracts which drop in or out of that classification based on iterative methodological updates to the index.

Given that meaningful engagement with community members is central to the field of environmental justice, disseminating to a wide and non-technical audience emphasizes the balance between statistical merit and model interpretability. The methods in this index are aligned with the latter, but here we observe that the effects of these decisions manifest in downstream effects on index uncertainty. Efforts must be made to more transparently communicate this uncertainty to end-users in reports and visualization products, and policy-makers should be informed so as to understand the

implications of any biases that were introduced by the methodology as currently selected. We recommend that additional iterations of the EHD index construction focus on aligning normalization procedures with the statistical characteristics of the raw indicators so as to preserve distributions amidst standardizing - clearly, decisions on whether to apply linear or nonlinear scaling procedures merit careful reflection based on their influence on ranks. Finally, the classification framework should be reexamined to ensure that no tracts are being left behind by statistical artifacts that are introduced by the current definition of high-impact. Given the stability and robustness of this index in classifying the uppermost decile, understanding the behavior of tracts that are on either side of the 20% threshold is crucial to validating that classification system.

There are several limitations that must be considered when interpreting the results of this study. The approximation of measurement error in this analysis was derived from simulating classical measurement errors which are randomly applied and thus uncorrelated. Error in survey data has been observed to defy the assumptions required of classical measurement errors, especially with regards to intercorrelation and association with the dependent variable - as such our estimates for the sensitivity of EHD to measurement error are likely conservative and the true effect on the index is attenuated. Likewise, we only tested a single type of PCA aggregation, based on applying the Kaiser criterion to filter down the component factors based on their eigenvalue. This method is understood to be an overly conservative threshold, retaining too many inconsequential factors and leading to the development of Horn's parallel analysis method[31]. Parallel analysis employs a Monte Carlo simulation to run PCA on simulated versions of the dataset and then compare the resulting distributions of eigenvalues to the observed values from running PCA on the original dataset. Factors with observed eigenvalues that are greater than the simulated ones are retained, typically with lower retention compared to the Kaiser criterion[16]. It's likely that permuting across other types of PCA would increase the variance in deviation from the baseline EHD, so this analysis does not fully capture the effect on rankings that changing from deductive to inductive aggregation may entail.

Conclusion

Analytical assumptions are intrinsic to the process of using mathematical models to quantify and approximate real-world phenomena. Understanding the consequences of those decisions is critical for transparently communicating results and for prioritizing finite resources during iterative model development. Global sensitivity analysis is an established technique for quantifying the uncertainty that is introduced within the construction of composite indicators and our analysis demonstrates the wide range of values that could be produced in alternative formulations of the EHD index. By decomposing that variance into the model constituents, we observe that future iterations of the EHD should carefully consider the appropriate methods for normalization of raw model inputs to preserve data as much as possible when standardizing each indicator to a common unit scale. For the purpose of upholding environmental justice, the findings of this analysis should be utilized to develop communication and policymaking efforts that are transparent, accurate, and aligned with reducing systematic inequalities in the cumulative impact of environmental health.

References

1. Snow, J. On the Mode of Communication of Cholera. *Edinb. Med. J.* **1**, 668–670 (1856).
2. Frumkin, H. *Environmental Health: From Global to Local*. (John Wiley & Sons, 2016).
3. Gee, G. C. & Payne, -Sturges Devon C. Environmental Health Disparities: A Framework Integrating Psychosocial and Environmental Concepts. *Environ. Health Perspect.* **112**,

1645–1653 (2004).

4. Corvalán, C. F., Kjellström, T. & Smith, K. R. Health, Environment and Sustainable Development: Identifying Links and Indicators to Promote Action. *Epidemiology* **10**, 656–660 (1999).
5. Hambling, T., Weinstein, P. & Slaney, D. A Review of Frameworks for Developing Environmental Health Indicators for Climate Change and Health. *Int. J. Environ. Res. Public. Health* **8**, 2854–2875 (2011).
6. Morello-Frosch, R., Zuk, M., Jerrett, M., Shamasunder, B. & Kyle, A. D. Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy. *Health Aff. (Millwood)* **30**, 879–887 (2011).
7. Saisana, Michaela & Tarantola, Stefano. *Saisana, Michaela, and Stefano Tarantola. State-of-the-art report on current methodologies and practices for composite indicator development.* vol. 214 (Ispra: European Comission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit, 2002).
8. Barclay, M., Dixon-Woods, M. & Lyratzopoulos, G. The problem with composite indicators. *BMJ Qual. Saf.* **28**, 338–344 (2019).
9. Saisana, M., Saltelli, A. & Tarantola, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J. R. Stat. Soc. Ser. A Stat. Soc.* **168**, 307–323 (2005).
10. Handbook on constructing composite indicators: methodology and user guide - OECD. https://www.oecd.org/els/soc/handbookonconstructingcompositeindicatorsmethodologyanduserguide.htm.
11. Otoiu, A., Titan, E. & Dumitrescu, R. Are the variables used in building composite indicators of well-being relevant? Validating composite indexes of well-being. *Ecol. Indic.* **46**, 575–585 (2014).

12. Burgass, M. J., Halpern, B. S., Nicholson, E. & Milner-Gulland, E. J. Navigating uncertainty in environmental composite indicators. *Ecol. Indic.* **75**, 268–278 (2017).

13. Becker, W., Saisana, M., Paruolo, P. & Vandecasteele, I. Weights and importance in composite indicators: Closing the gap. *Ecol. Indic.* **80**, 12–22 (2017).

14. Cutter, S. L., Boruff, B. J. & Shirley, W. L. Social Vulnerability to Environmental Hazards*. *Soc. Sci. Q.* **84**, 242–261 (2003).

15. Beccari, B. A Comparative Analysis of Disaster Risk, Vulnerability and Resilience Composite Indicators. *PLoS Curr.* **8**, ecurrents.dis.453df025e34b682e9737f95070f9b970 (2016).

16. Tate, E. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Nat. Hazards* **63**, 325–347 (2012).

17. Saltelli, A. *et al. Global Sensitivity Analysis: The Primer*. (John Wiley & Sons, 2008).

18. Saltelli, A. *et al.* Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environ. Model. Softw.* **114**, 29–39 (2019).

19. Sarrazin, F., Pianosi, F. & Wagener, T. Global Sensitivity Analysis of environmental models: Convergence and validation. *Environ. Model. Softw.* **79**, 135–152 (2016).

20. VILLA, F. & McLEOD, H. Environmental Vulnerability Indicators for Environmental Planning and Decision-Making: Guidelines and Applications. *Environ. Manage.* **29**, 335–348 (2002).

21. Saisana, M. & Saltelli, A. Uncertainty and sensitivity analysis of the 2008 environmental performance index. *Jt. Res. Cent. Eur. Comm. Ispra* (2008).

22. Nations, U. *Uncertainty and Sensitivity Analysis of the Human Development Index*. *Human Development Reports* https://hdr.undp.org/content/uncertainty-and-sensitivity-analysis-human-development-index (2010).

23. Cutter, S. L. Race, class and environmental justice. *Prog. Hum. Geogr.* **19**, 111–122

(1995).

24. Kuruppuarachchi, L., Kumar, A. & Franchetti, M. A Comparison of Major Environmental Justice Screening and Mapping Tools. *Environ. Manag. Sustain. Dev.* **6**, 59 (2017).

25. Greenfield, B. K., Rajan, J. & McKone, T. E. A multivariate analysis of CalEnviroScreen: comparing environmental and socioeconomic stressors versus chronic disease. *Environ. Health* **16**, 131 (2017).

26. Min, E. *et al.* The Washington State Environmental Health Disparities Map: Development of a Community-Responsive Cumulative Impacts Assessment Tool. *Int. J. Environ. Res. Public. Health* **16**, 4470 (2019).

27. Zhang, X.-Y., Trame, M., Lesko, L. & Schmidt, S. Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT Pharmacomet. Syst. Pharmacol.* **4**, 69–79 (2015).

28. Decancq, K. & Lugo, M. A. Weights in Multidimensional Indices of Wellbeing: An Overview. *Econom. Rev.* **32**, 7–34 (2013).

29. Spiegelhalter, D. *et al.* Statistical methods for healthcare regulation: rating, screening and surveillance. *J. R. Stat. Soc. Ser. A Stat. Soc.* **175**, 1–47 (2012).

30. Brody, T. M., Bianca, P. D. & Krysa, J. Analysis of Inland Crude Oil Spill Threats, Vulnerabilities, and Emergency Response in the Midwest United States. *Risk Anal.* **32**, 1741–1749 (2012).

31. Schmidtlein, M. C., Deutsch, R. C., Piegorsch, W. W. & Cutter, S. L. A Sensitivity Analysis of the Social Vulnerability Index. *Risk Anal.* **28**, 1099–1114 (2008).

32. Schwartz, J., Bellinger, D. & Glass, T. Expanding the Scope of Environmental Risk Assessment to Better Include Differential Vulnerability and Susceptibility. *Am. J. Public Health* **101**, S88–S93 (2011).

33. Abel, T. D. & White, J. Skewed Riskscapes and Gentrified Inequities: Environmental Exposure Disparities in Seattle, Washington. *Am. J. Public Health* **101**, S246 (2011).

Chapter 3: *Validating the structure of an environmental justice index in Washington State, a multivariate case study analysis*

Intro

Environmental health is a critical determinant of health outcomes, with environmental factors accounting for a significant proportion of morbidity and mortality worldwide[1]. In the United States, there are substantial disparities in exposure to environmental hazards and resulting health outcomes across populations[2]. The state of Washington is home to diverse populations, and understanding the extent and nature of environmental health inequalities in the state is essential for benchmarking purposes and to inform the development of effective public health policies and interventions[3]. The Washington State Environmental Health Disparities (EHD) Map is an effort to measure and analyze environmental health inequalities across populations in Washington state using a range of indicators, including air quality, exposure to pollutants and hazardous waste sites, socioeconomic variables, and the rates of diseases that increase population vulnerabilities to environmental exposures.

As part of the broader environmental justice movement, the EHD map and adjacent state-wide and federal indicators are being employed in policy-making efforts that aim to identify hotspots and reduce inequality[4]. However, these indicators have been demonstrated to exhibit considerable sensitivity to analyst decisions, which can manifest in index uncertainty and in allocative harm from misspecified policies[5]. By identifying and quantifying the most important factors that are driving variation in this index of environmental health disparity, our study contributes to a deeper understanding of the social and environmental determinants of health and informs policy to promote health equity in Washington and beyond.

Environmental health data often has unique statistical characteristics that must be taken into account during analysis[6]. For instance, the data generating processes of environmental health indicators are often complex and interrelated, with multicollinearity resulting from the myriad of pathways through which variables influence each other[7]. Additionally, the data can tend towards skewness or contain severe outliers due to the clustered nature of environmental exposures[8]. Furthermore, environmental health data are often spatially correlated, as nearby areas will tend towards similar environmental

conditions and health outcomes[9]. Understanding these statistical characteristics is important for appropriate analysis and interpretation of environmental health data to identify and mitigate environmental risks to human health.

Furthermore, the collection of environmental health data from different unit spaces mandates that data treatment must include standardization to ensure that apples are compared with apples before aggregating to a common index, and the resulting indices have been observed to be sensitive to different standardization methods[10]. Common normalization techniques for composite indicator construction can include linear transformations like min-max scaling or z-scoring or unitless approaches such as force-ranking the observations and generating quantiles of the results[11]. The latter approach is simple, intuitive, and robust to outliers, but inherently results in loss of information as the underlying distribution of the raw indicator data must be transformed, while linear transformations can be sensitive to the kind of outlying values expected in environmental health data[12]. The variable impact of normalization schema on indicator distributions must be understood and carefully considered in order to account for the impact of potential biases associated with each type of method on the resulting composite values.

The sensitivity of the EHD map index to different normalization methods has been observed to be second only to the deductive aggregation formula in terms of the overall impact on the resulting index when compared to alternatives from the broader composite indicator construction literature. The imposition of a theoretical framework through which to cluster, weight, and aggregate indicators clearly shapes the output index, and the validity of this approach is dependent on the strength of the contextual knowledge that supports its formulation[13]. Alternatively, many composite indicators apply inductive aggregation approaches in order to reduce the effect of the analyst assumptions on the resulting output, allowing instead for the variation of the input data to drive the aggregation formula[14]. The most common approach is Principal Components Analysis (PCA), a dimensionality reduction technique that transforms correlated variables into a smaller number of uncorrelated principal components. These components can then be weighted and combined to create a single composite indicator, capturing the underlying information of the original variables more concisely[15].

Intermediate elements of the PCA process can also be used to identify patterns and relationships in the indicator data, and to help visualize high-dimensional data in lower-dimensional space.

In this study, we analyzed the statistical characteristics of the 19 environmental health indicators that were included in the recent EHD map version 2.0 in order to understand their behavior under different normalization methods and quantify the implications of those transformations on resulting index rankings. We provide case studies of the census tracts that are most impacted by the current decile normalization methodology in order to illustrate the effect of the established method in comparison to alternative approaches. We also estimated a contribution-weighted PCA version of the EHD ranks in order to understand the underlying structure of the 19 indicators and to quantify the efficacy of an inductively generated measure of cumulative environmental health impact for classifying census tracts and allocating funding within the context of environmental justice policy-making according to the Healthy Environment for All (HEAL) Act (*RCW 43.70.815*).

Methods

The WA EHD Map

The state of Washington adopted the CalEnviroScreen model, which creates cumulative impact scores across multiple environmental hazards and population characteristics, to generate the first version of the Washington Environmental Health Disparities Map[4]. The baseline model integrated measures of environmental exposures, adverse environmental effects, sensitivities, and socio-demographic vulnerabilities to create a composite score, underpinned by the following equation:

$$EHD\ Rank\ = P\ x\ S$$

Where P represents 'Pollution Burden Score', summarizing environmental risks and hazards, and S a 'Population Characteristics Score', capturing various community-level characteristics influencing vulnerability to environmental risk. Within those groupings, 19 indicators were additionally grouped into four sub-categories: Environmental Exposures, Environmental Effects, Sensitive Populations, and Socioeconomic Factors. Each indicator's raw values were normalized by assigning a decile score. Finally, a decile

ranking of 1–10 was used for the final disparities rank in the resultant map. The top two deciles were used to identify highly impacted communities that may be eligible for remedial funding under the HEAL act. Version 2.0 of the EHD map was released in July 2022, with data updates to reflect the most recently available indicator data and methodological updates for the heavy traffic and diesel PM2.5 exposure indicators[16]. Agreement in the highly impacted community classification between the two versions was used to generate a metric of impact uncertainty for visualization and analysis.

Alternative normalization methods

Normalization is a process of transforming data to a standard form for effective comparison and analysis. The three most commonly used normalization methods for composite indicator construction are z-score normalization, min-max normalization, and rank-based normalization.

Z-score standardization, also known as standard score or standardization, is a technique used to transform individual data points to a standardized scale by subtracting the mean from each data point and then dividing by the standard deviation[17]. To perform z-score standardization for a given data point x, the following formula is used, where x is the original value of the data point, μ represents the mean of the dataset, and δ represents the standard deviation of the dataset:

$$Z = \frac{x-\mu}{\delta}$$

Alternatively, min-max normalization is a linear transformation that scales data to a specified range using the following formula:

$$M = \frac{x-min(x)}{max(x)-min(x)}$$

To aid in comparability to the baseline EHD index, we rescaled data to a range between 0 and 10 for this analysis.

Finally, rank-based normalization requires sorting the data and assigning a rank to each data point based on its value relative to the other observations in the dataset. In accordance with the baseline EHD methodology, these values were then divided into ten equal deciles, where the value for each observation is replaced with its decile value, ranging from 1 to 10. The lowest ranked observations (the bottom 10%) are in the first

decile and given a value of 1, the next 10% are in the second decile and given a value of 2, and so forth, until the highest ranked observations (the top 10%) are in the tenth decile and given a value of 10. In order to observe the nonlinear scaling effect of this baseline approach with higher resolution, we also tested rank scaling the indicators and then dividing them into centiles by binning the output ranks into 100 equally sized quantiles.

PCA

Principal component analysis (PCA) is a statistical method used for reducing high-dimensional data. PCA involves transforming the original variables into a new set of uncorrelated variables called principal components. The covariance matrix is decomposed into its eigenvalues and eigenvectors, where the eigenvalues represent the amount of variance explained by each principal component, and the corresponding eigenvectors represent the direction or weights of the components.These principal components are ordered in terms of the amount of variation they capture in the data. The first principal component captures the most variation, with monotonically decreasing variation for subsequent components. In this study, PCA was performed using the covariance matrix of the data to extract the principal components, and the percentage of variation explained by each principal component was examined.

Since components capture subsequently less variation in the inputs, the results are typically filtered to retain only the most explanatory components. A standard threshold for filtration is based on the Kaiser criterion, which suggests that only components with eigenvalues greater than 1 should be retained, while components with eigenvalues less than 1 should be discarded. The Kaiser criterion is derived from the mathematical properties of the eigenvalues, where values greater than 1 indicate that the corresponding component explains more variance than an individual variable, making it worthwhile to retain[18].

To generate a composite indicator from PCA, it’s essential to select a method for aggregating the resulting factors into a single measure. Most simple is to just select the first component, which statistically will explain the greatest amount of variation in the input dataset. To preserve more of the factors, which will tend to capture different

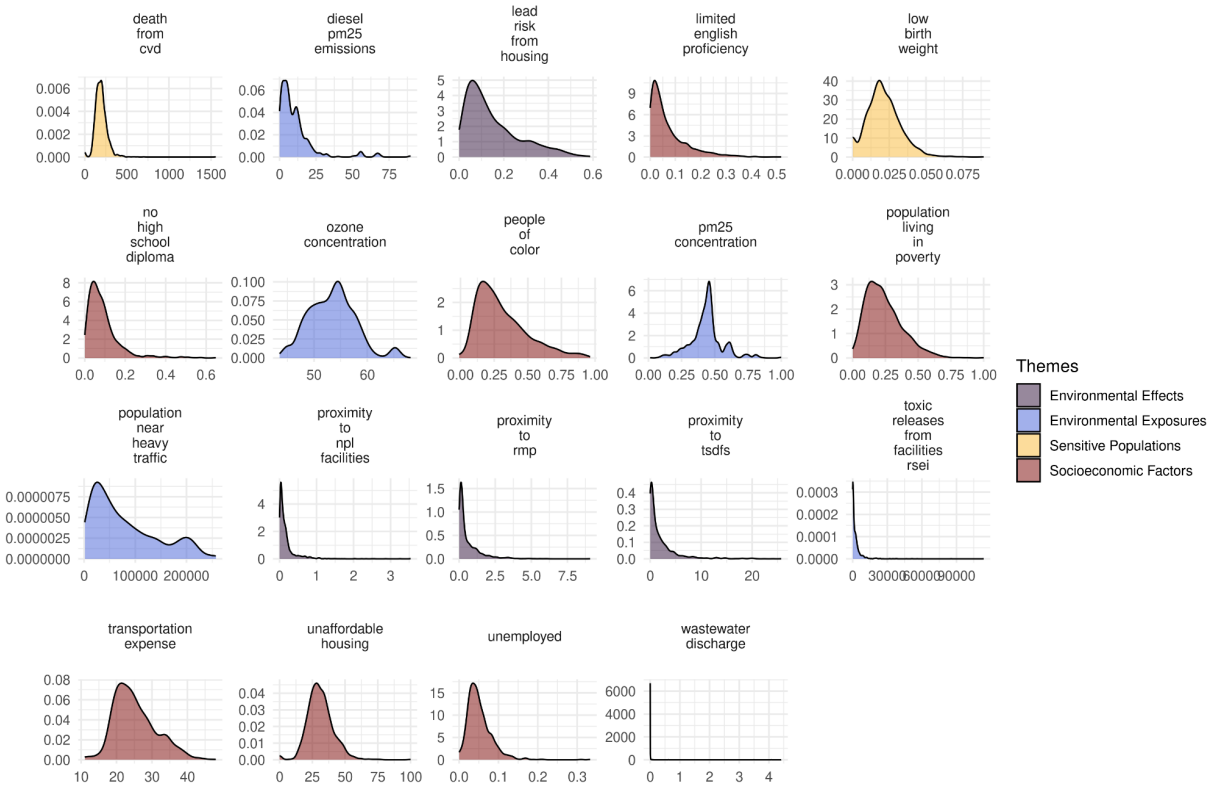
elements of the latent variable that the index is designed to proxy, they could be equally weighted and summed linearly, or weighted using the explained variance from each factor - downweighting the influence to the later components which will inherently explain less and less variation as they are rotated[19]. We used the latter approach to aggregate the remaining PCs after filtration based on the Kaiser criterion. The R statistical software package *prcomp* was used to perform the PCA and generate the visualizations.

Results

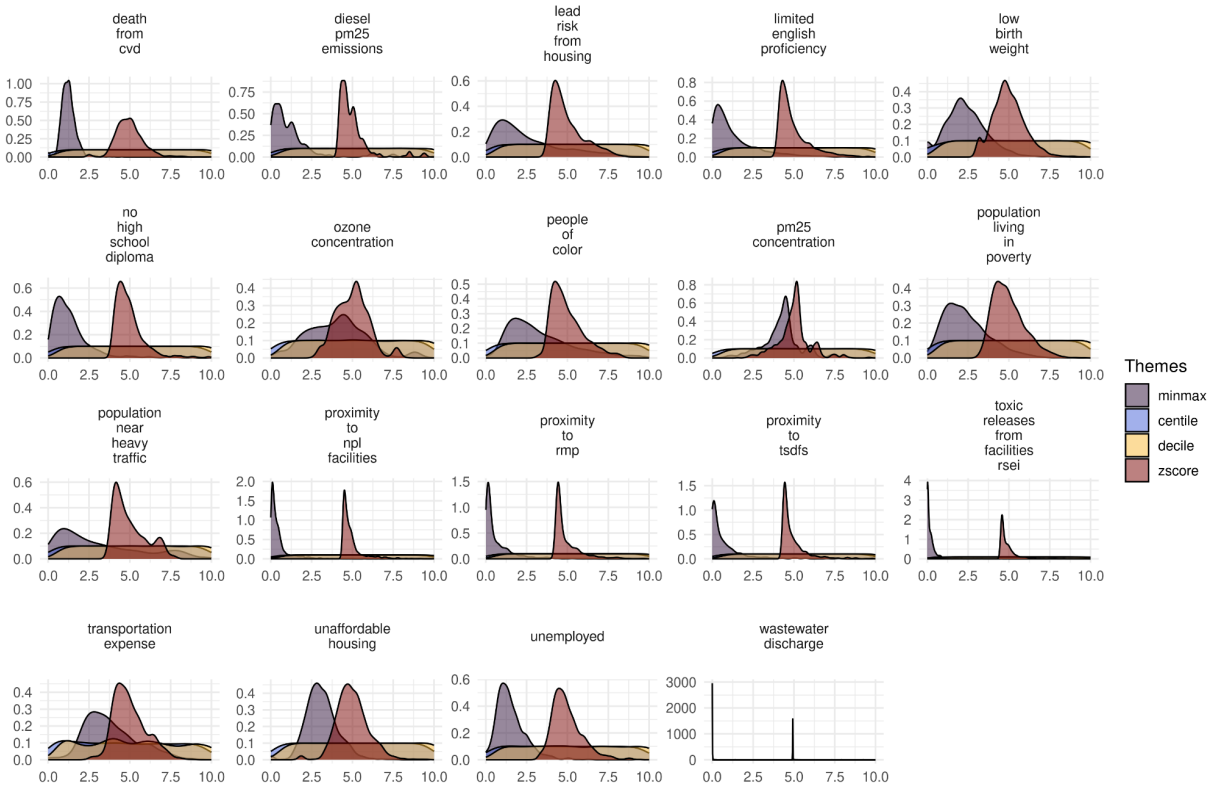
The raw data was observed to exhibit considerable positive skewness for the majority of indicators selected for the EHD map (*figure 1a)*. The Environmental Effects indicator grouping was universally skewed to the right (group skew of 6.3), with wastewater discharge being the most impacted (skew of 15.6). Environmental Exposures data was also largely skewed (group skew of 3.2), though ozone (0.4) and PM2.5 (0.4) concentrations were the most normally distributed indicators in this dataset. The Sensitive Populations and Socioeconomic Factors data was also more normally distributed where both themes had a group skew of 1.4, and all of the indicators for rates of low birth weight, population living in poverty, population near heavy traffic, transportation expense, and unaffordable housing had skew of less than one.

Figure 1: Distributions of indicator data under different normalization functions

a)



b)

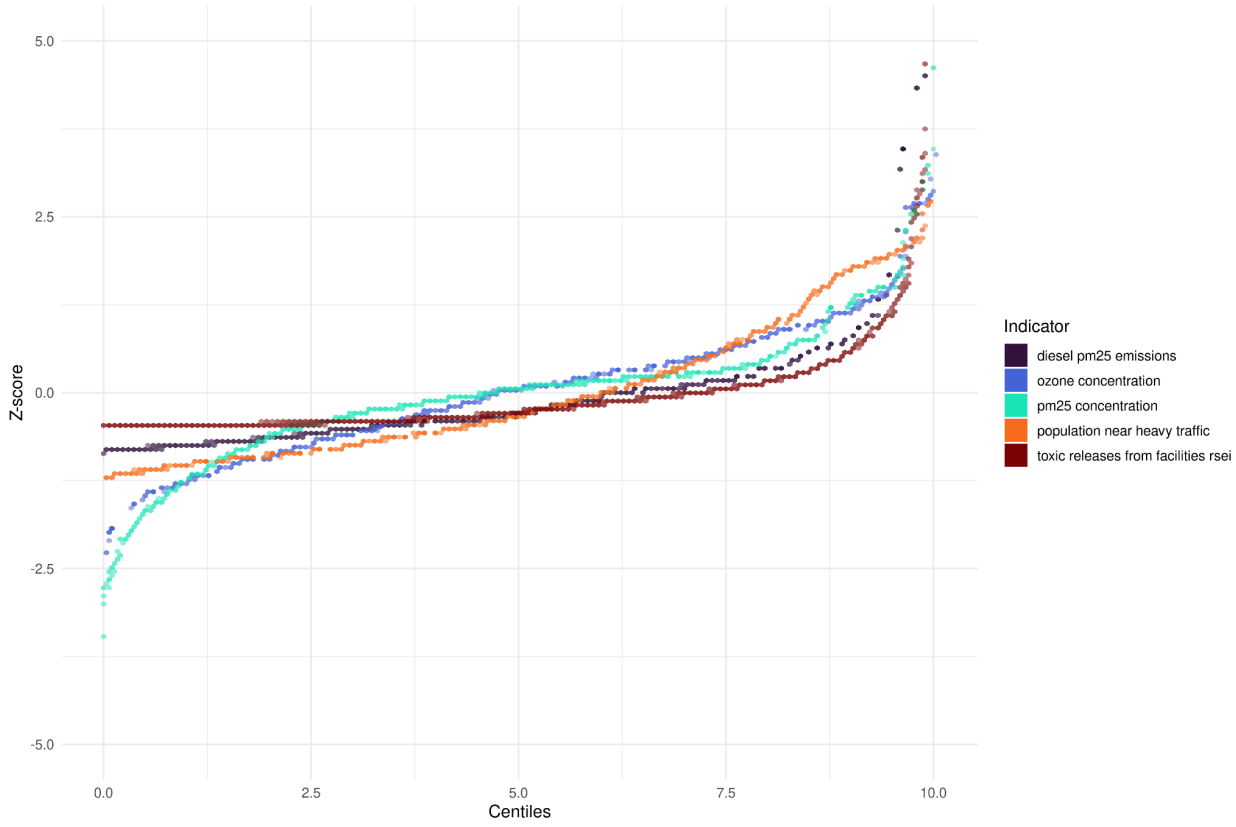


The probability distributions of the 19 indicators as raw data (**a)** and the data after transformation in four different normalization schemes (**b**) are plotted. The distributions are colored based on their theme grouping as raw data and colored based on the normalization scheme being used when transformed.

We observe that rank-standardization methods of decile or percentile ranking inherently impose a uniform distribution onto the indicator data after normalization, while the linear scaling functions of z-scoring or min-max normalization tend to preserve the shape and statistical characteristics of the input distributions (*figure 1b*). When comparing centiles to min-max estimates across all of the input indicators (*figure 2)*, we observe two important nonlinear transformations. First, at the upper bound of the distribution, centile ranking introduces a compression effect and variation is heavily reduced when units that have a large absolute distance in values are binned into the top percentiles and given the same score. Second, the much denser center of the probability distribution is stretched across multiple percentile groupings as arbitrary

divisions are created to appropriately size the bins. This compression effect is more notable for skewed indicators and the curves for the Environmental Effects data are clustered, while the stretching effect tends towards impacting the normally-distributed indicators, such as the Sensitive Populations theme. Both effects are most extreme for the wastewater discharge data, which is heavily right skewed and with substantial non-random missingness. As expected, the min-maxed estimates for this indicator are highly subject to the impact of outliers, while the ranks produced by centile normalization will introduce more artificial variation by forcibly dividing the mode of this distribution to equally-sized bins.

Figure 2: Nonlinear scaling effects



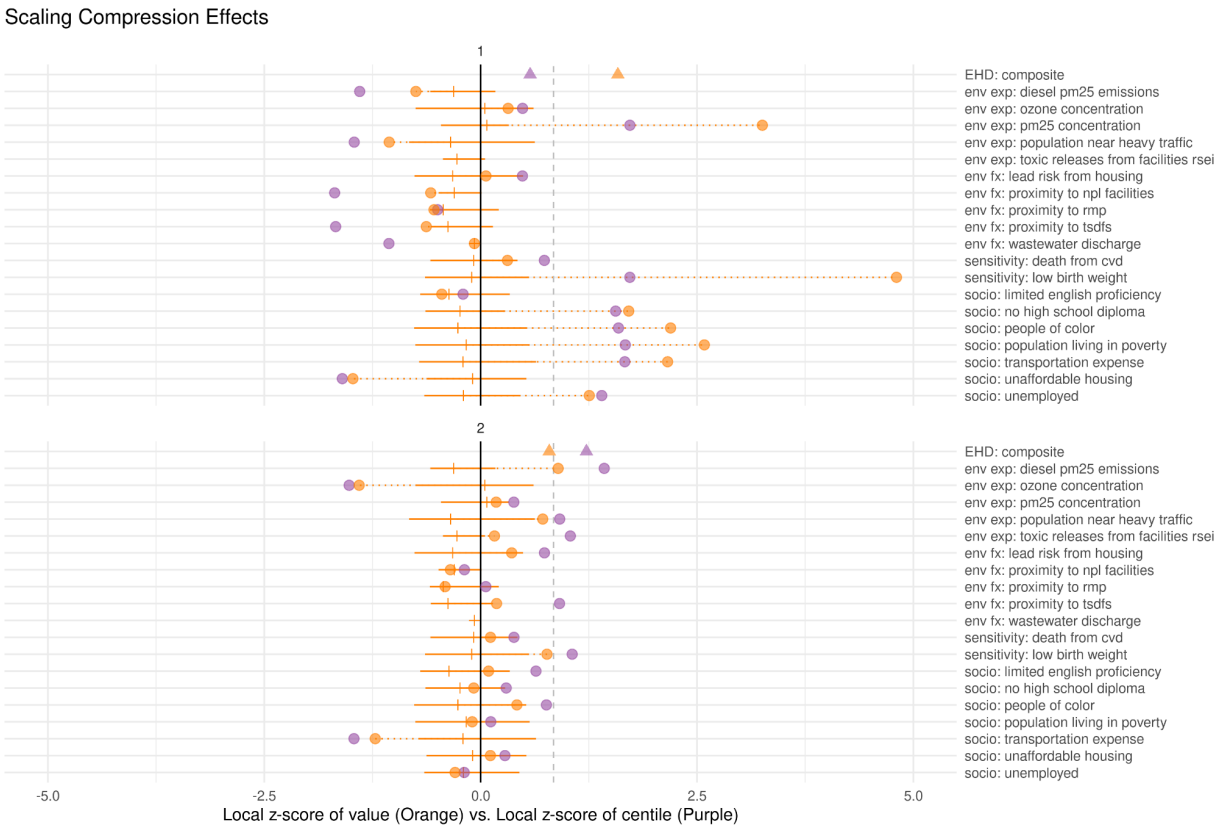
Plotting the indicator values for each census tract when scaled nonlinearly using centile ranking (x-axis) against their values when scaled linearly by z-score standardization (y-axis) demonstrates the transformation that across the range of the probability distribution in the former method of normalization. The data is plotted separately for each of the environmental exposures indicators and the hexagons are colored according to the indicator.

To examine the influence of linear versus nonlinear scaling methodology on the classification of highly impacted census tracts, we analyzed two case studies of census tracts with the largest difference in ranking between the baseline EHD index and a version that used z-score normalization instead (*figure 3)*. The first tract, located in the Greenwood neighborhood of north Seattle, ranks in the 85th percentile of the baseline EHD index and in the 73rd percentile of a z-scored index. Examining the scores for each indicator and comparing them to the distribution of the rest of the state, it has high values for diesel PM2.5 emissions, rates of low birthweight and low values for ozone concentration and transportation expense. All other metrics are within the IQR for the state and many are close to the median value. We observe that 14/19 indicators are above the median value for the state, and as such the centile rankings for these indicators are similar, ranging from the 50th percentile to the 79th.

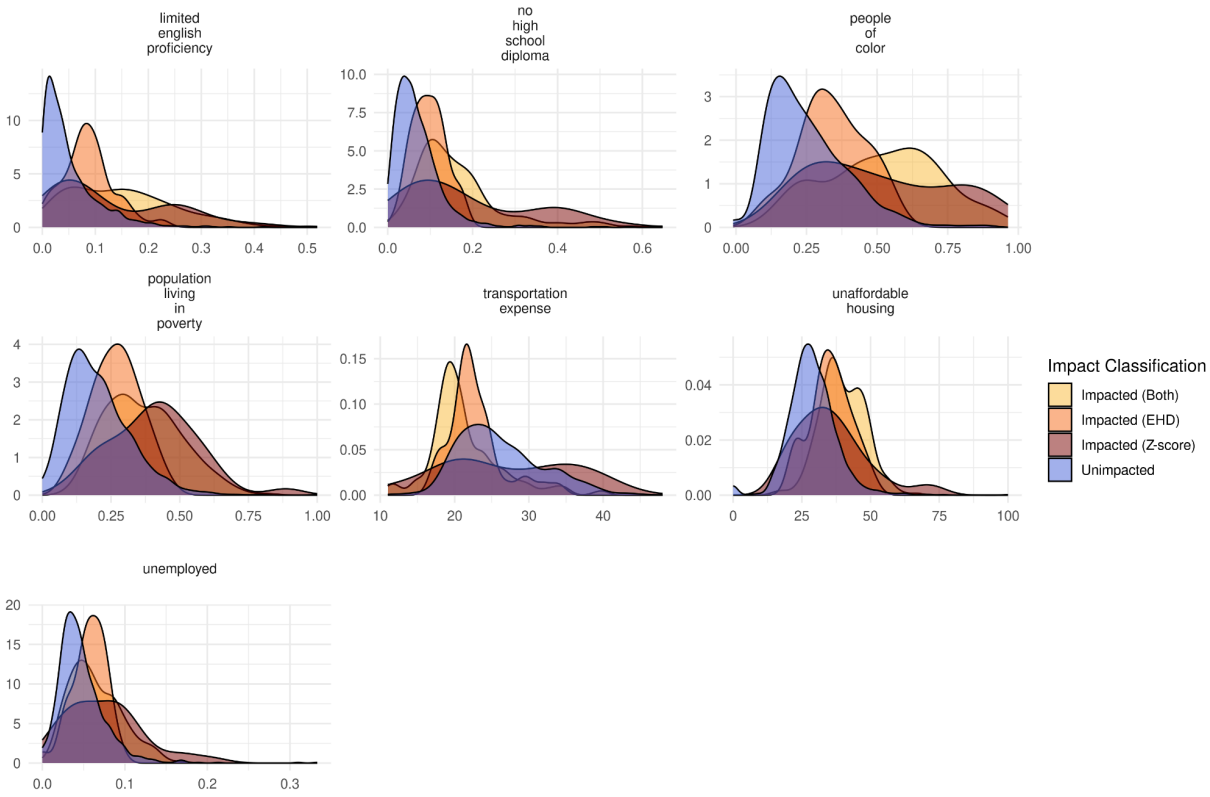
The second tract, located in rural Okanogan County, ranks in the 66th percentile of the baseline EHD index and the 96th percentile of an index using z-score normalization. We observe severe outliers for rates of low birthweight, PM2.5 concentration, population living in poverty, transportation expense, and proportions of population that are people of color, lacking a high school diploma, and unemployed. There are also outliers on the left side of the distribution for unaffordable housing and population near heavy traffic. All other metrics are within the state-wide IQR, and only 10/19 indicators are above the state median. Centile-based normalization is observed to favor higher ranks for tracts that are above average in the majority of indicators, while reducing the impact of outliers on the final ranking. Z-score normalization tends to highly score the environmental health effects in tracts where there are extreme values in a smaller number of metrics across the dataset. Comparing the raw values for tracts that were classified as highly impacted under the baseline EHD to those that were classified as highly impacted with Z-score normalization, we observe that the latter tend to be less wealthy, less educated, and have less English proficiency and more people of color (*figure 3c).* While there is disagreement between the two methods statewide, the largest clusters of census tracts that were classified as highly impacted when normalized by Z-scores but not in the baseline EHD index are located in rural areas in the center of Washington state.

FIgure 3: Case studies of scaling compression effects

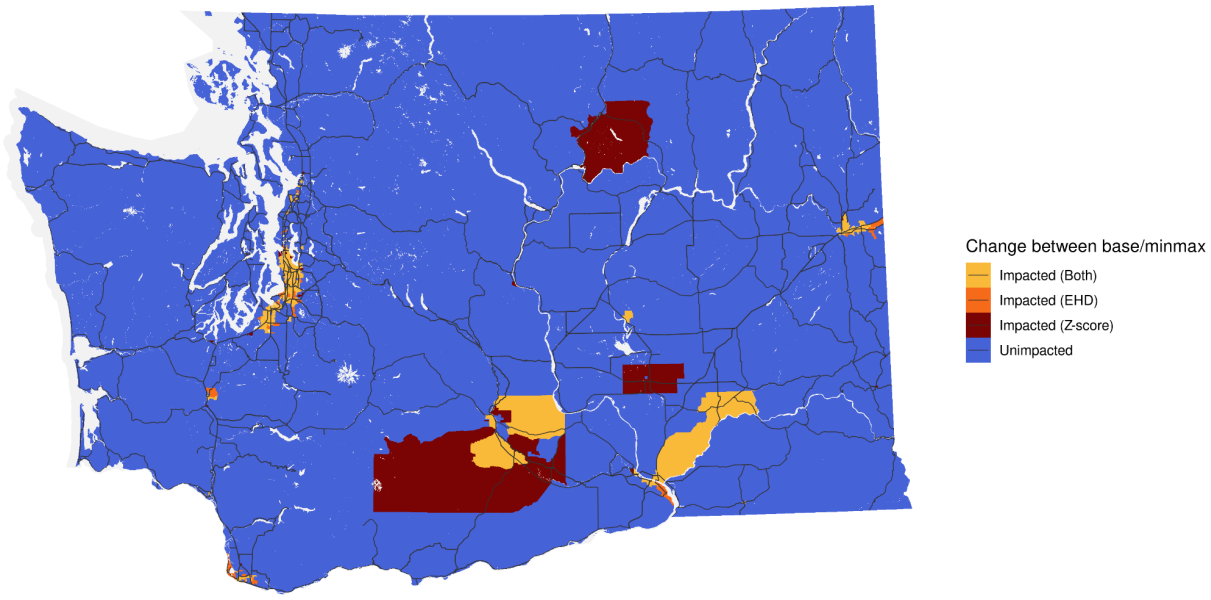
a)



b)


limited english proficiency
no high school diploma
people of color
population living in poverty
transportation expense
unaffordable housing
unemployed
Impact Classification
Impacted (Both)
Impacted (EHD)
Impacted (Z-score)
Unimpacted

c)

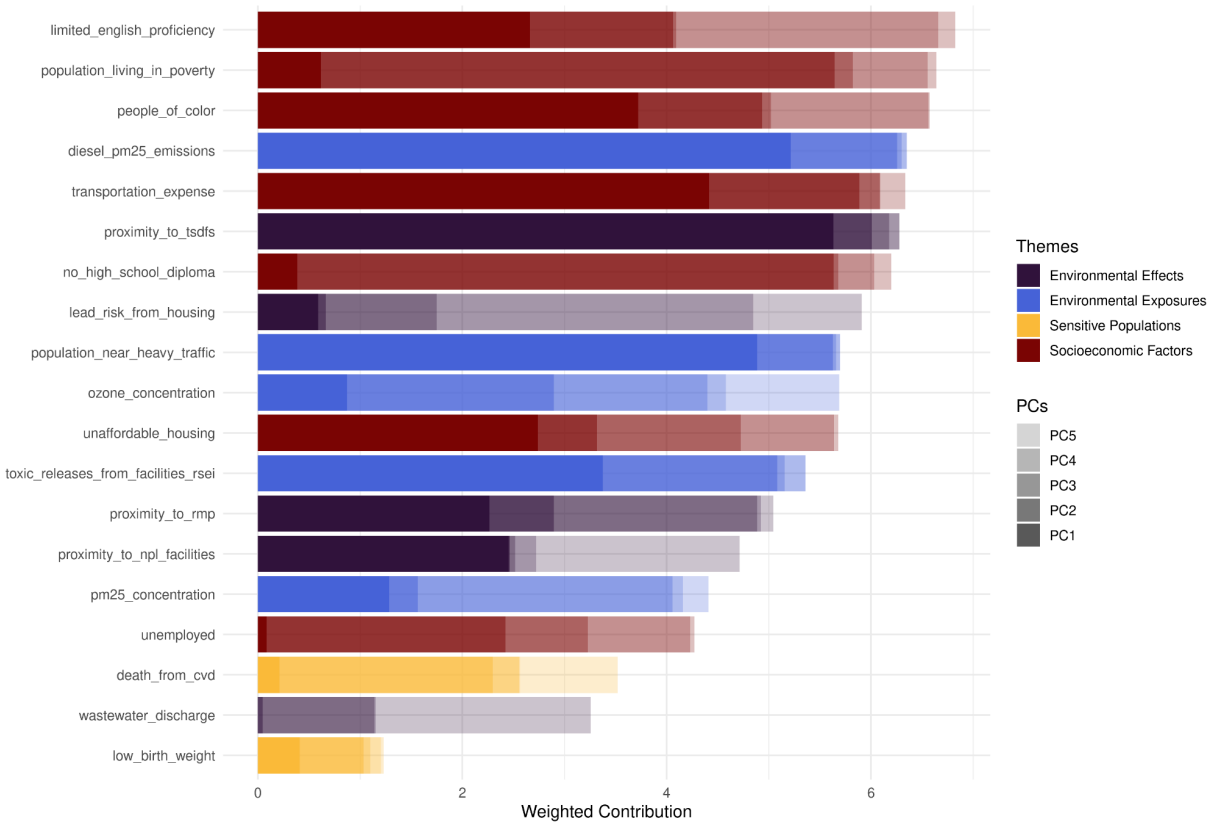


We examine two case studies where the difference in ranking and impact classification between z-score normalization and centile normalization is greatest **(b)**. The first case study is a tract in rural Okanogan county where the overall EHD ranking is in the 66th percentile when centile transformed and in the 97th percentile when z-scored, while the second is a tract in North Seattle that was overall in the 85th percentile by centile ranking and in the 73rd percentile when z-scored. For both tracts, the aggregated ranking (triangle) and each of the 19 indicators are displayed after z-score transformation (orange) and after centile ranking (purple). To aid visualization, the centile values are also z-score transformed to standardize such that the distance from the state-wide mean (vertical black line) is apparent. The dashed grey vertical line represents the threshold for high impact (top 20%). The median and IQR of the state-wide distribution of z-scored values is also displayed as an orange boxplot to highlight the relationship of the transformed values to the rest of the state and the skewness of the distribution, where medians that are further left of the mean are more positively skewed.

We also generated a version of the EHD map using PCA in order to assess the differences between inductive and deductive aggregation methodologies. After running PCA on the 19 indicators and filtering the resulting principal components according to the Kaiser criterion, we were left with five components that explained 67% of the total variation. Socioeconomic indicators tended to have the largest contribution to the final PCA score, with the percentages of limited English proficiency, people living below the

poverty line, and people of color being the top three contributors (*figure 3*). Diesel PM2.5 emissions, transportation expense, proximity to TSDFs, percent without a high school diploma, lead risk from housing, population counts near heavy traffic areas, and ozone concentration rounded out the rest of the top ten - suggesting that the socioeconomic factors (five) explain the most variation, followed by environmental exposures (three) and environmental effects (two). The sensitive population indicators had the lowest contribution, with the bottom three contributors being rates of low birth weight, wastewater discharge, and rates of cardiovascular disease. The first two components explained substantially more variation (45% in total, 28% and 18% respectively) than the remaining three, so we focused additional analysis on examining those components to understand the correlation structure of the underlying indicators (*figure 4a)*.

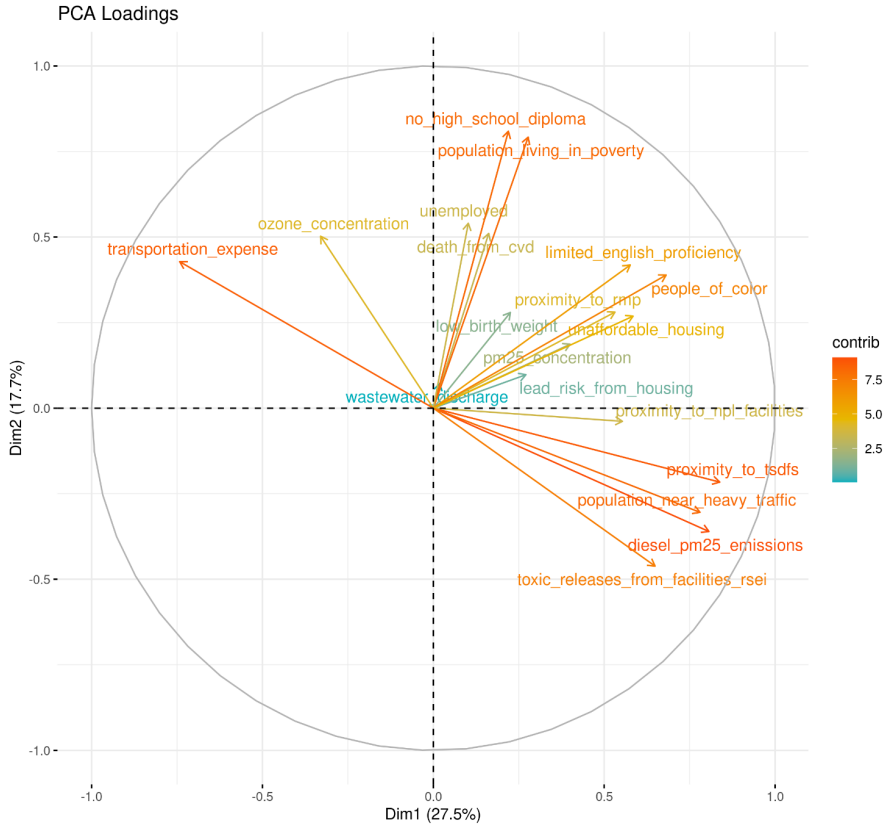
Figure 4a: PCA Contribution



The weighted contribution of each of the 19 indicators to the oveall PCA value are displayed, where the transparency of the bar represents the variance explained by that component. The bars are colored by the theme category grouping and ordered by their overall contribution to the PCA.

We observed that the first component (28% of the total variation) captures an urban environmental degradation effect, with high contribution from indicators that measure traffic and industrial pollution, with large loadings from proximity to TSDF (0.37), RMP (0.23), and NPL (0.24) facilities, population near heavy traffic (0.34), diesel PM2.5 emissions (0.35), toxic releases from RSEI facilities (0.28), and percent people of color (0.29). PC1 also had a large negative loading from transportation expense (-0.32), which exhibits inverse correlation with urbanicity[16] and tends to be highest in rural census tracts. Rates of low birthweight, death from cardiovascular disease, and proportion of population without a high school degree or employment all had loading values less than 0.10 for the first component. The second component, which captured an additional 18% of the data variation, suggests an underlying factor of poverty and socioeconomic deprivation. Loadings were observed to be high for the proportion of population without a high school diploma (0.44), living in poverty (0.43), unemployed (0.29), rates of death from cardiovascular disease (0.28) and ozone concentration (0.27). Lead risk from housing and proximity to NPL facilities both had loading values that were below 0.10 for the second component.

Figure 4b: PCA Biplot



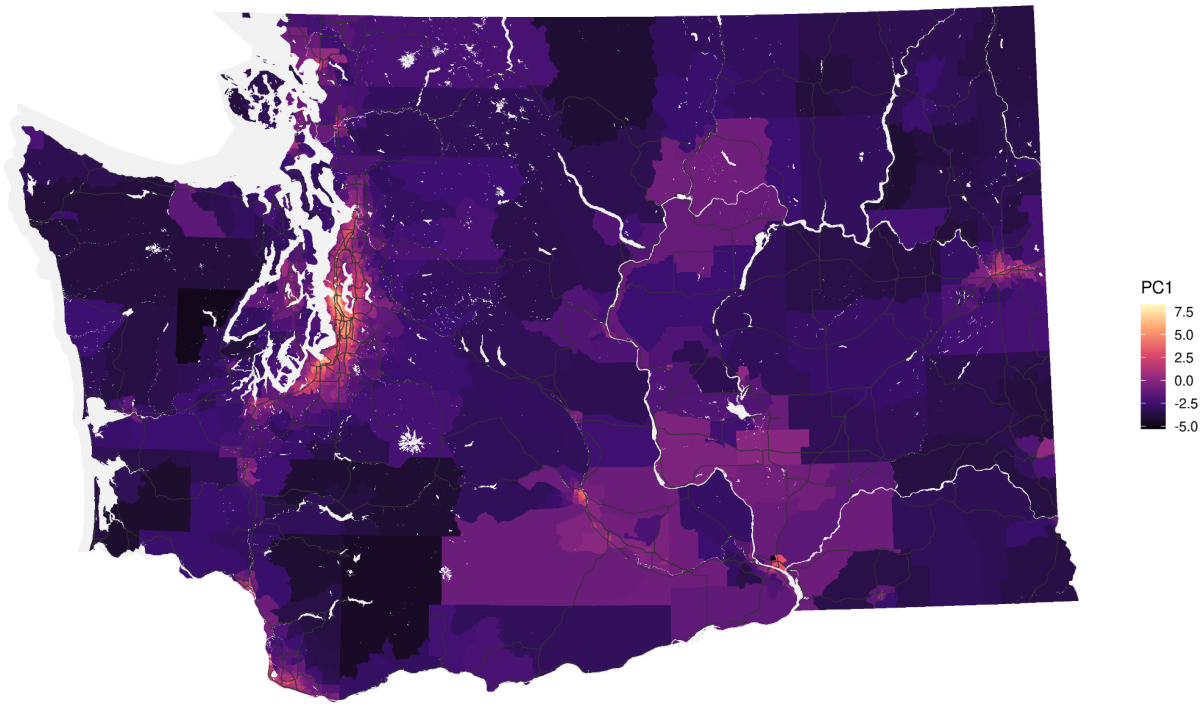
The makeup of the first two principal components are displayed as a biplot where the first component is the y-axis and the second is the x. Each indicator is displayed as vector in that space, where distance from either axis represents the contribution to that component and the direction of the vector represents the sign of that relationship. Indicators that are closer to an axis are contributing less to that axis, while indicators that are located near the 45 degree line are contributing relatively equally. We observe the upper right quadrant to be capturing a combined gradient of urban degradation and socioeconomic deprivation, where a number of indicators such as limited English proficiency and poverty contribute to both, while the lower left quadrant would then represent a less socioeconomically deprived tract that is still impacted by urban degradation and association pollution burden.

Wastewater discharge had the lowest combined contribution to the first two components, with neither loading exceeding 0.05, while the proportion of population with low English proficiency and people of color were the only two indicators that had loadings of greater than 0.20 for both. Mapping the first component at the census tract level (*figure 5a)* illustrates the gradient of urban and industrial pollution being captured in this transformation of the underlying data, with the highest values concentrated in King and Pierce counties. PC2 (*figure 5b)* is less spatially homogenuous, with hot spots

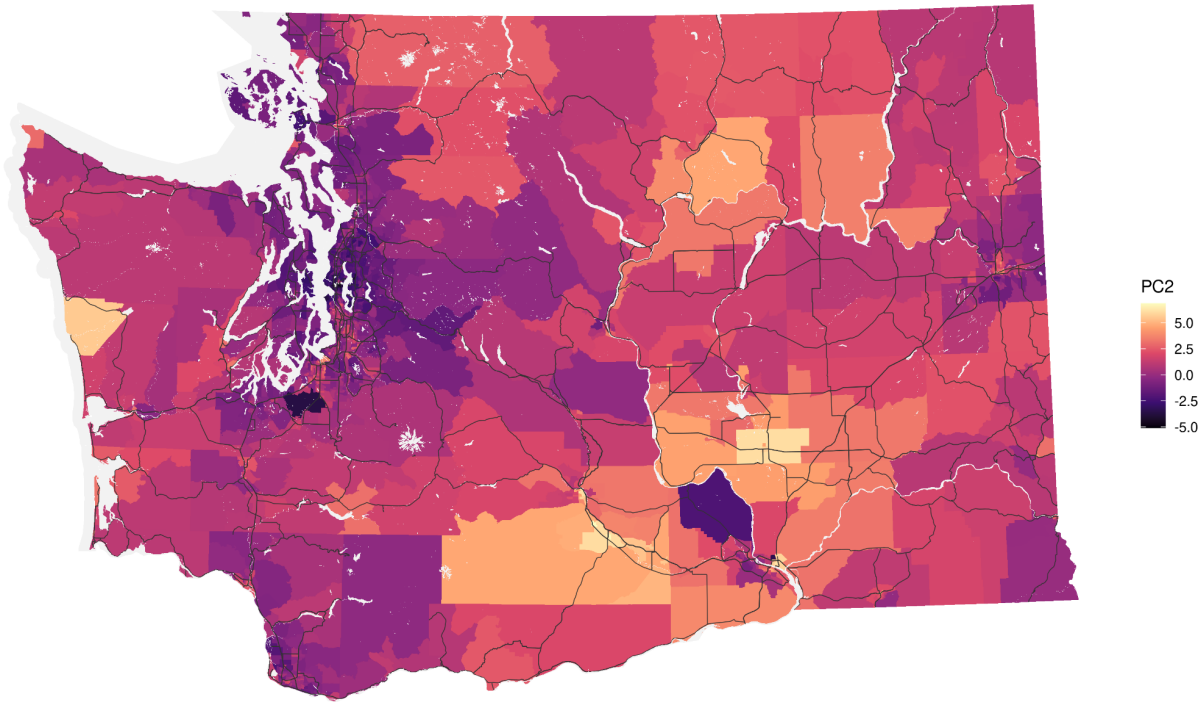
throughout rural areas in the center of the state, with values that are particularly high for Yakima county. We observe higher values for PC2 in some urban areas as well, mostly near the I-5 corridor south of Seattle. In comparing deciles of environmental impact created by this PCA to the current version of the EHD map, we observe that on average the agreement was close, with a population-weighted deviation in census tract decile ranking is 0.03 (*figure 5d).* However, there are census tracts in Wahkiakum, Spokane, Benton, and Garfield counties where the PCA index was considerably lower, while it was notably higher in census tracts on the Olympic Peninsula and in Island, Pacific, and Adams counties.

Figure 5: PCA Maps

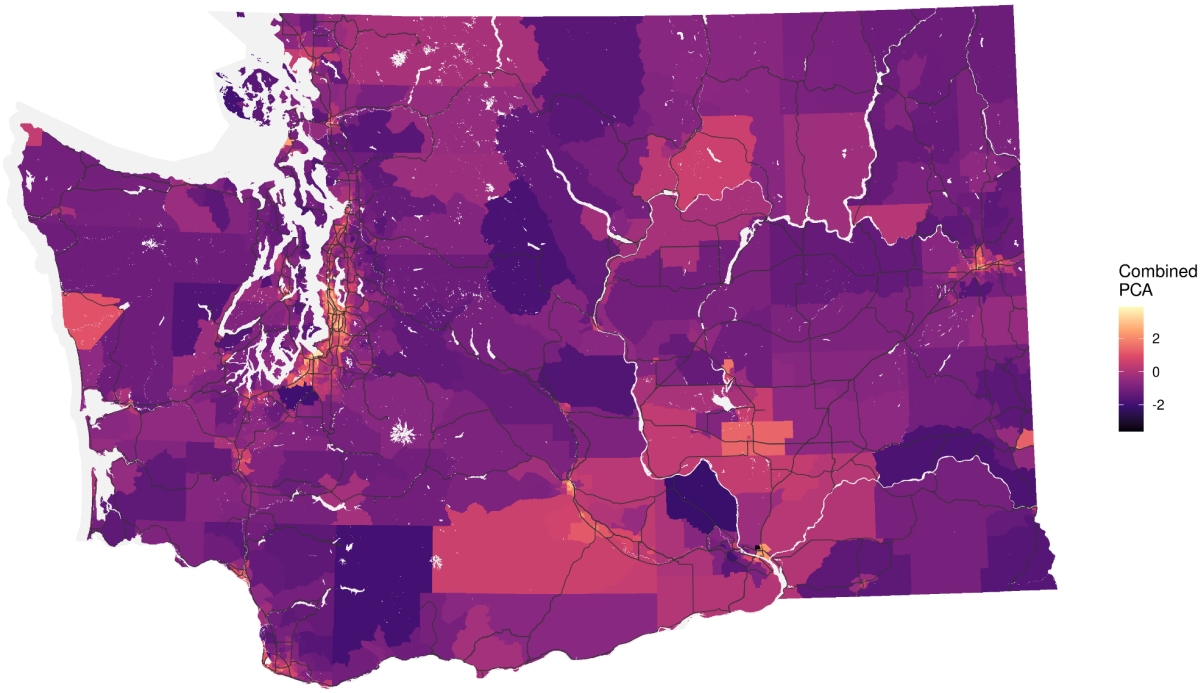
a)



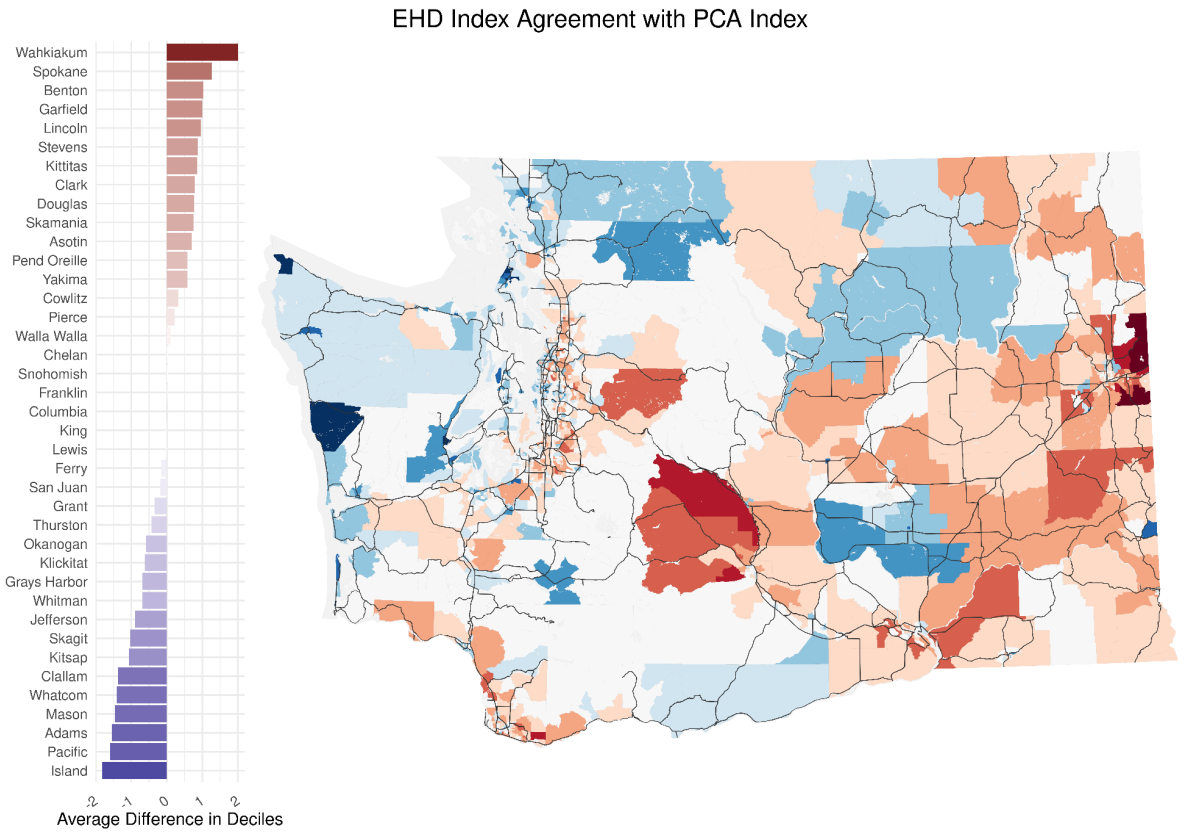
b)



c)



d)

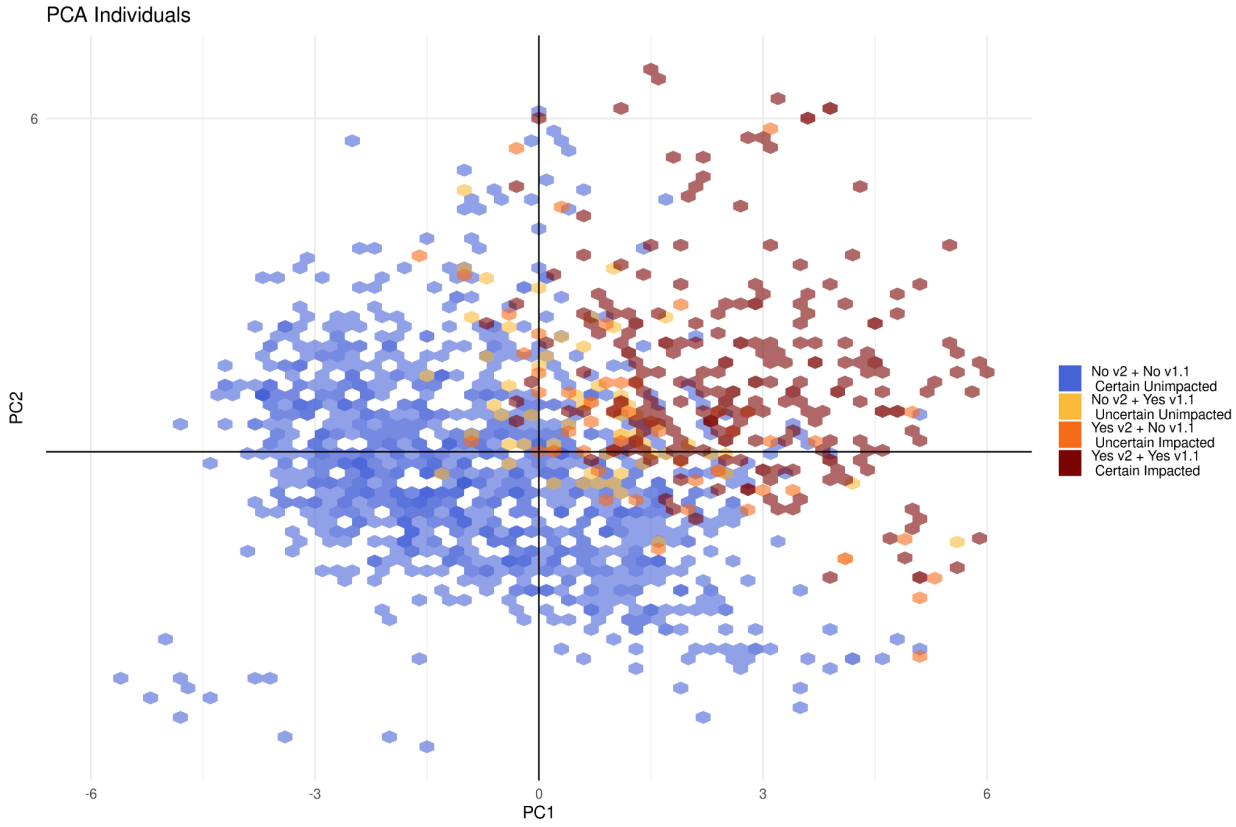


Elements of the PCA are mapped at the census tract level, including the values for the first component (**a**), second component (**b),** the contribution-weighted overall score (**c**) and the agreement between the baseline EHD and decile rankings of the contribution-weighted overall score (**d).** The agreement plot was created by subtracting the PCA deciles from the baseline score, so positive values (red) represent tracts where the baseline EHD values were higher and vice versa for negative values (blue).

We generated a metric for the certainty of classifying highly impacted tracts (the upper 20% of the rankings) by estimating agreement between the first two versions of the EHD index. Census tracts classified as highly impacted in both versions were understood to be more certain than those that moved in or out of the impact classification based on methodological update. We observed a strong relationship between the first two components of this PCA and the classifications for highly impacted tracts produced by the baseline EHD ranking system (*figure 4b)*. Of the tracts classified as highly impacted with greater certainty, 77% had positive values for both components. Only 7% of the census tracts that were found in the bottom 80% of environmental

impact by both maps were also located in that upper right quadrant of the PCA biplot. Nearly 98% of the tracts highly impacted with greater certainty had positive values for the first component, suggesting that the urban degradation gradient identified by the PCA is more predictive of tracts being classified as highly impacted in the current formulation of the EHD map.

Figure 6: Classification of highly impacted tracts according to the first two principal components



The impact classification of each census tract in the baseline EHD index is plotted as a function of the first two principal components, where the transparency of the plotted hexagons represents the density of census tracts in that plot location. The colors of the hexagons where generated by comparing the impact classification of the first two versions of the EHD index, where blue and red represent agreement in the classification of impact between the two versions. The yellow and orange points mean that a given tract either dropped out or was shifted into the top 20% in the updated version of the EHD index, suggesting less certainty of its classification status. We observe that the first two principal components classify with high accuracy the tracts that were more certain to be highly impacted, as 77% of those tracts are located in the upper right quadrant (socioeconomically deprived and urban) of the biplot.

Discussion

The majority of the indicators selected to form the EHD rankings exhibit considerable positive skewness and the transformed shape of these abnormal distributions are highly influenced by the choice of normalization methodology in the index construction. Nonlinear scaling by way of decile or centile rankings inherently assumes uniformity, and will both stretch the central mass of the distribution into a number of equally spaced quantiles and compress outlying values on the right side of the distribution into the upper ranks. We observe that these effects impact the rankings and classification of highly impacted tracts that are produced in indicators using different types of normalization, where rank scaling selects for tracts that are above average in the majority of metrics, while linear transformations like z-score standardization favors tracts with severe outliers in a fewer number of indicators. Inductively aggregating the index using PCA produces rankings that generally agree with the baseline EHD index, and only two principal components, explaining close to 50% of the total data variation, can form an accurate classifier for the highly impacted census tracts in the baseline EHD index. The structure of these first two components indicates underlying gradients of urban environmental degradation and socioeconomic deprivation are driving variation in environmental justice in Washington state.

Composite indicators have established sensitivity to the choice of normalization methodology, and this study helps to illustrate the mechanism through which different methods of transformation will influence results[20,21]. Rank normalization is often employed in statistical analysis to reduce the influence of outliers, and we observe this effect in the baseline EHD index as the comparison z-score index tends to rank more highly tracts that have substantial outliers in a fewer number of the indicators. The baseline EHD index favors tracts that are above average in most indicators, while not necessarily having any values that are excessive when compared to the state-wide distribution. This functionality is aligned with the philosophical directive of a cumulative impact score, and is supported by evidence that the effects of environmental risks and population-level vulnerabilities are synergistic and exhibit effect modification that can multiplicatively increase their impact on health[22–24]. Furthermore, several of the measured environmental pollutant exposures have been observed to have supralinear

dose-response curves, suggesting that disease risk rises sharply with low levels of exposure and later plateaus[25–27]. As such, the compression effect for positively skewed distributions of environmental risk data may more accurately reflect the health effect of those pollutants.

However, the shape of the association between socioeconomic vulnerabilities and environmental exposures is less clear, and we observe that an index generated from linear scaling transformations of this dataset tends to rank more highly the impact for census tracts located in areas that are more socioeconomically deprived. Clearly there is an important environmental justice consideration that must be understood when deciding between these normalization alternatives. Being above average in many indicators and exhibiting severe outliers in a few both represent critical threats to the health of a given community, and perhaps an alternative formulation of the index that takes the maximum of both rankings would be preferable. This method would capture both kinds of cumulative impact, and ensure that neither community is left behind during policymaking and funding allocation.

Our case study for estimating the index based on the results of a PCA illustrates some of the fundamental differences between using inductive and deductive aggregation strategies to generate environmental composite indicators. First, the strong performance of PCA as a classifier of highly impacted tracts suggests that more data-driven approaches are feasible for future iterations of environmental justice indicators that reduce the influence of methodological assumptions on the end result. However, we observed that while both approaches generally agreed on the most highly impacted units, there was spatial heterogeneity in the differences in rankings across the state for these two systems. This finding aligns with previous work suggesting that the rankings of the EHD index exhibit considerable sensitivity to analyst decisions, but that the highest impact areas tend to be robustly estimated regardless of the index formulation. Furthermore, this analysis demonstrates that factors generated based on underlying variation in this dataset have important differences in their inherent structure when compared to the theoretically derived theme groupings in the EHD map and suggests that the processes that drive environmental disparities in Washington state are complex, exhibiting substantial intercorrelation.

While the two observed gradients of urban degradation and socioeconomic deprivation are not contradictory with the Pollution Burden and Sensitive Populations aggregate themes of the EHD map, we also see that the underlying indicators have overlap considerably between and within those groups in the PCA. The high explanatory power of the urban component in the highly impacted tracts supports community concerns that the current indicator selection for the EHD index does not do enough in measuring environmental health issues that are relevant to rural communities, and addition of new metrics such as wildfire smoke and pesticide usage in the next iteration will help to address that data gap. Our finding that indicators of population with limited English proficiency and people of color contribute so heavily to both of the first two principal components aligns with a long history of BIPOC communities suffering disproportionate impact from environmental injustice, and supports previous evidence that these disparities are not explained by differences in income[2,28]. Finally, the notable lack of variation explained by the wastewater discharge indicator points to issues with data quality that should be examined in future versions of the EHD index.

There are several inherent limitations to PCA that must be considered when interpreting these results. PCA is a linear process of data transformation, which may be inapplicable in data with non-linear structures. Furthermore, there are several different criteria for which components to retain from the initial estimation of PCA, where the Kaiser criterion is the standard - filtering out any factors with an eigenvalue of less than one. This method is now understood to be an overly conservative threshold, retaining too many inconsequential factors and leading to the development of Horn's parallel analysis method[19]. Parallel analysis employs a Monte Carlo simulation to run PCA on simulated versions of the dataset and then compare the resulting distributions of eigenvalues to the observed values from running PCA on the original dataset. Factors with observed eigenvalues that are greater than the simulated ones are retained, typically with lower retention compared to the Kaiser criterion[29]. As such, it's likely that permuting the threshold at which components were filtered from the analysis would reveal that our PCA results are sensitive to that analytical assumption and alternative formulations of the PCA would have fewer components. Finally, since PCA is driven

purely by variation in the data, indicators are considered to be fundamentally equal, which may not reflect the true association with these indicators and environmental health risk. Cumulative risk based approaches can more accurately capture both the underlying risk-response functions of these indicators as well as their interaction effects and propagate uncertainties in those effects throughout the estimation process. Future iterations of this analysis should incorporate the current body of knowledge in environmental risk assessment to reflect the strength of evidence for the constituent indicators and weight them accordingly.

Conclusion

Populations in Washington state are not distributed equally across the spectrum of exposure to environmental health risks, and accurately quantifying these longstanding disparities is paramount to the design of policies that proactively manifest environmental justice. The derivation of composite indicators that produce cumulative impact scores from an array of relevant environmental and socioeconomic indicators has been demonstrated to be sensitive to analyst assumptions, and this analysis illustrates the pathways through which the most sensitive parameters influence the results. Future iterations of these metrics must carefully consider the ramifications of their architecture in order to assure accuracy, validity and fair access to remediation funding opportunities. Misunderstanding the biases introduced by methodological assumptions can inadvertently perpetuate inequality, skew results, and misinform strategic decisions, all of which can lead to deleterious societal implications. Further, such biases can distort the representation and interpretation of reality, rendering policies less effective or even harmful. Recognizing and addressing these biases can foster more equitable, reliable, and robust policies. Hence, when data-driven policies are shaped with an informed awareness of potential biases, they are not only scientifically rigorous, but also more ethically sound, reinforcing a broader goal of societal welfare and justice.

References

1. Shaffer, R. M. *et al.* Improving and Expanding Estimates of the Global Burden of Disease Due to Environmental Health Risk Factors. *Environ. Health Perspect.* **127**, 105001.
2. DOWNEY, L. & HAWKINS, B. RACE, INCOME, AND ENVIRONMENTAL INEQUALITY IN THE UNITED STATES. *Sociol. Perspect. SP Off. Publ. Pac. Sociol. Assoc.* **51**, 759–781 (2008).
3. Ingram, C., Min, E., Seto, E., Cummings, B. & Farquhar, S. Cumulative Impacts and COVID-19: Implications for Low-Income, Minoritized, and Health-Compromised Communities in King County, WA. *J. Racial Ethn. Health Disparities* **9**, 1210–1224 (2022).
4. Min, E. *et al.* The Washington State Environmental Health Disparities Map: Development of a Community-Responsive Cumulative Impacts Assessment Tool. *Int. J. Environ. Res. Public. Health* **16**, 4470 (2019).
5. Huynh, B. Q. *et al.* Potential for allocative harm in an environmental justice data tool. Preprint at https://doi.org/10.48550/arXiv.2304.05603 (2023).
6. Ott, W. R. *Environmental Statistics and Data Analysis*. (CRC Press, 1994).
7. Walter, S. D. The ecologic method in the study of environmental health. II. Methodologic issues and feasibility. *Environ. Health Perspect.* **94**, 67–73 (1991).
8. Abel, T. D. & White, J. Skewed Riskscapes and Gentrified Inequities: Environmental Exposure Disparities in Seattle, Washington. *Am. J. Public Health* **101**, S246 (2011).
9. Elliott, P. & Wartenberg, D. Spatial Epidemiology: Current Approaches and Future Challenges. *Environ. Health Perspect.* **112**, 998–1006 (2004).
10. Pollesch, N. L. & Dale, V. H. Normalization in sustainability assessment: Methods and implications. *Ecol. Econ.* **130**, 195–208 (2016).
11. Handbook on constructing composite indicators: methodology and user guide - OECD. https://www.oecd.org/els/soc/handbookonconstructingcompositeindicatorsmethodologyanduserguide.htm.

12. Santeramo, F. G. On the Composite Indicators for Food Security: Decisions Matter! *Food Rev. Int.* **31**, 63–73 (2015).

13. Asadzadeh, A., Kötter, T., Salehi, P. & Birkmann, J. Operationalizing a concept: The systematic review of composite indicator building for measuring community disaster resilience. *Int. J. Disaster Risk Reduct.* **25**, 147–162 (2017).

14. Tate, E. Uncertainty Analysis for a Social Vulnerability Index. *Ann. Assoc. Am. Geogr.* **103**, 526–543 (2013).

15. Cutter, S. L., Boruff, B. J. & Shirley, W. L. Social Vulnerability to Environmental Hazards*. *Soc. Sci. Q.* **84**, 242–261 (2003).

16. University of Washington Department of Environmental & Occupational Health Sciences and Washington State Department of Health. Washington Environmental Health Disparities Map: Cumulative Impacts of Environmental Health Risk Factors Across Communities of Washington State: Technical Report Version 2.0. (2022).

17. Data Mining: Concepts and Techniques - 3rd Edition. https://www.elsevier.com/books/data-mining-concepts-and-techniques/han/978-0-12-381479-1.

18. Yeomans, K. A. & Golder, P. A. The Guttman-Kaiser Criterion as a Predictor of the Number of Common Factors. *J. R. Stat. Soc. Ser. Stat.* **31**, 221–229 (1982).

19. Schmidtlein, M. C., Deutsch, R. C., Piegorsch, W. W. & Cutter, S. L. A Sensitivity Analysis of the Social Vulnerability Index. *Risk Anal.* **28**, 1099–1114 (2008).

20. Talukder, B., W. Hipel, K. & W. vanLoon, G. Developing Composite Indicators for Agricultural Sustainability Assessment: Effect of Normalization and Aggregation Techniques. *Resources* **6**, 66 (2017).

21. Cherchye, L., Moesen, W., Rogge, N. & Puyenbroeck, T. V. An Introduction to ‘Benefit of the Doubt’ Composite Indicators. *Soc. Indic. Res.* **82**, 111–145 (2007).

22. Solomon, G. M., Morello-Frosch, R., Zeise, L. & Faust, J. B. Cumulative Environmental

Impacts: Science and Policy to Protect Communities. *Annu. Rev. Public Health* **37**, 83–96 (2016).

23. Park, S. K. *et al.* Air Pollution and Heart Rate Variability. *Epidemiol. Camb. Mass* **19**, 111–120 (2008).

24. Charafeddine, R. & Boden, L. I. Does income inequality modify the association between air pollution and health? *Environ. Res.* **106**, 81–88 (2008).

25. Burnett, R. *et al.* Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci.* **115**, 9592–9597 (2018).

26. Hazucha, M. J. & Lefohn, A. S. Nonlinearity in human health response to ozone: Experimental laboratory considerations. *Atmos. Environ.* **41**, 4559–4570 (2007).

27. Zeise, L., Wilson, R. & Crouch, E. A. Dose-response relationships for carcinogens: a review. *Environ. Health Perspect.* **73**, 259–306 (1987).

28. Bullard, R. D. Race and Environmental Justice in the United States. *Yale J. Int. Law* **18**, 319 (1993).

29. Tate, E. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Nat. Hazards* **63**, 325–347 (2012).