



APPENDIX AVAILABLE ON THE HEI WEBSITE

Research Report 211

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods

Francesca Dominici et al.

Appendix A. Supplemental Tables and Figures

This Appendix was reviewed solely for spelling, grammar, and cross-references to the main text. It has not been formatted or fully edited by HEI. This document was part of the HEI Low-Exposure Epidemiology Studies Review Panel's review process.

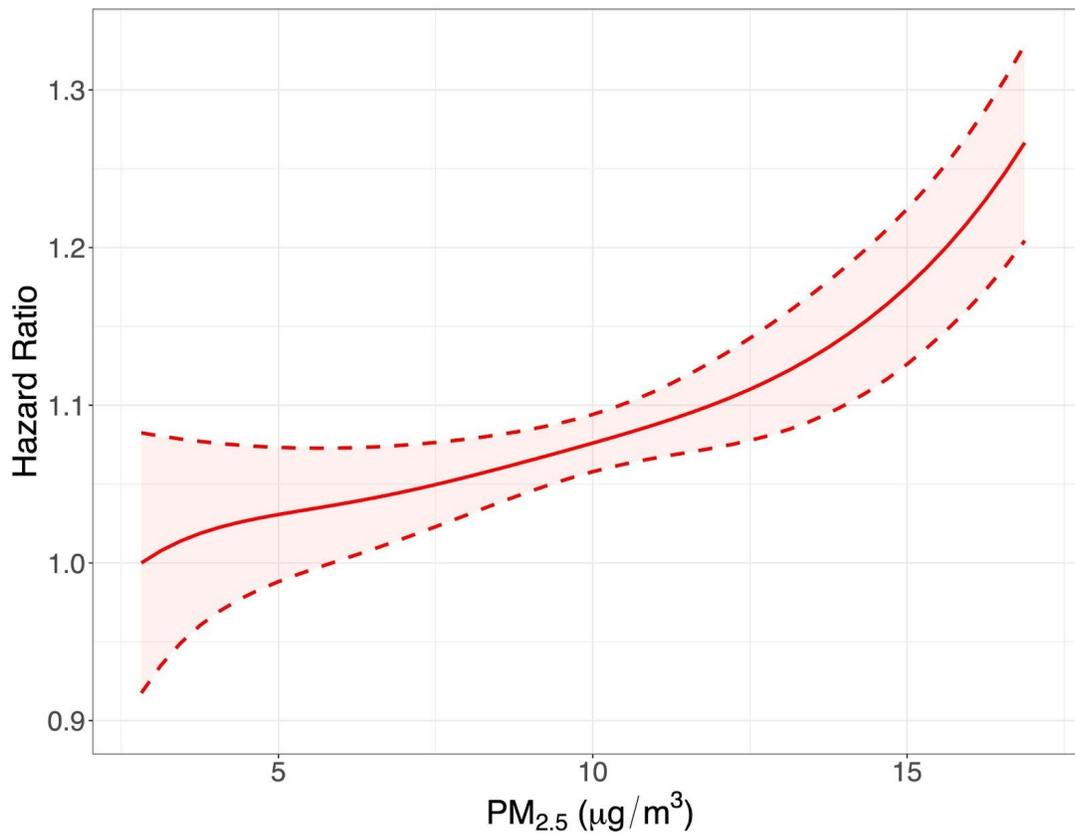
Correspondence may be addressed to Dr. Francesca Dominici, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; e-mail: fdominic@hsph.harvard.edu.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR-83467701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

© 2022 Health Effects Institute, 75 Federal Street, Suite 1400, Boston, MA 02110

HEI Research Report 211 Dominici Appendix, Available on the HEI Website

APPENDIX A. SUPPLEMENTAL TABLES AND FIGURES



Appendix Figure 1. Estimated causal ERC relating PM_{2.5} to all-cause mortality among Medicare enrollees (2000–2016) with associated 95% confidence bands obtained via bootstrap, only adjusting for one pollutant (NO₂) as a potential confounder. We define the baseline rate as the estimated HR corresponding to an exposure level set at the 1st percentile of the distribution of each pollutant. To avoid extrapolation at the support boundaries, we excluded the highest 1% and lowest 1% of pollutants exposures.

Appendix Table 1. HRs and 95% CIs Relating PM_{2.5}, NO₂, and O₃ to All-Cause Mortality in Medicare Beneficiaries (2000–2016)^a

| Pollutants | Models | GPS Matching | Poisson |
|-------------------|--|----------------------|----------------------|
| PM _{2.5} | Adjusted for NO ₂ and O ₃ | 1.036 (1.023, 1.065) | 1.056 (1.049, 1.063) |
| | Unadjusted for NO ₂ and O ₃ | 1.063 (1.050, 1.077) | 1.067 (1.060, 1.075) |
| | Adjusted for NO ₂ only | 1.044 (1.031, 1.057) | 1.055 (1.048, 1.062) |
| NO ₂ | Adjusted for PM _{2.5} and O ₃ | 0.997 (0.992, 1.001) | 1.009 (1.006, 1.012) |
| | Unadjusted for PM _{2.5} and O ₃ | 0.996 (0.992, 1.001) | 1.017 (1.014, 1.020) |
| O ₃ | Adjusted for PM _{2.5} and NO ₂ | 1.004 (0.995, 1.012) | 0.994 (0.990, 0.998) |
| | Unadjusted for PM _{2.5} and NO ₂ | 1.007 (0.999, 1.015) | 0.996 (0.992, 1.000) |

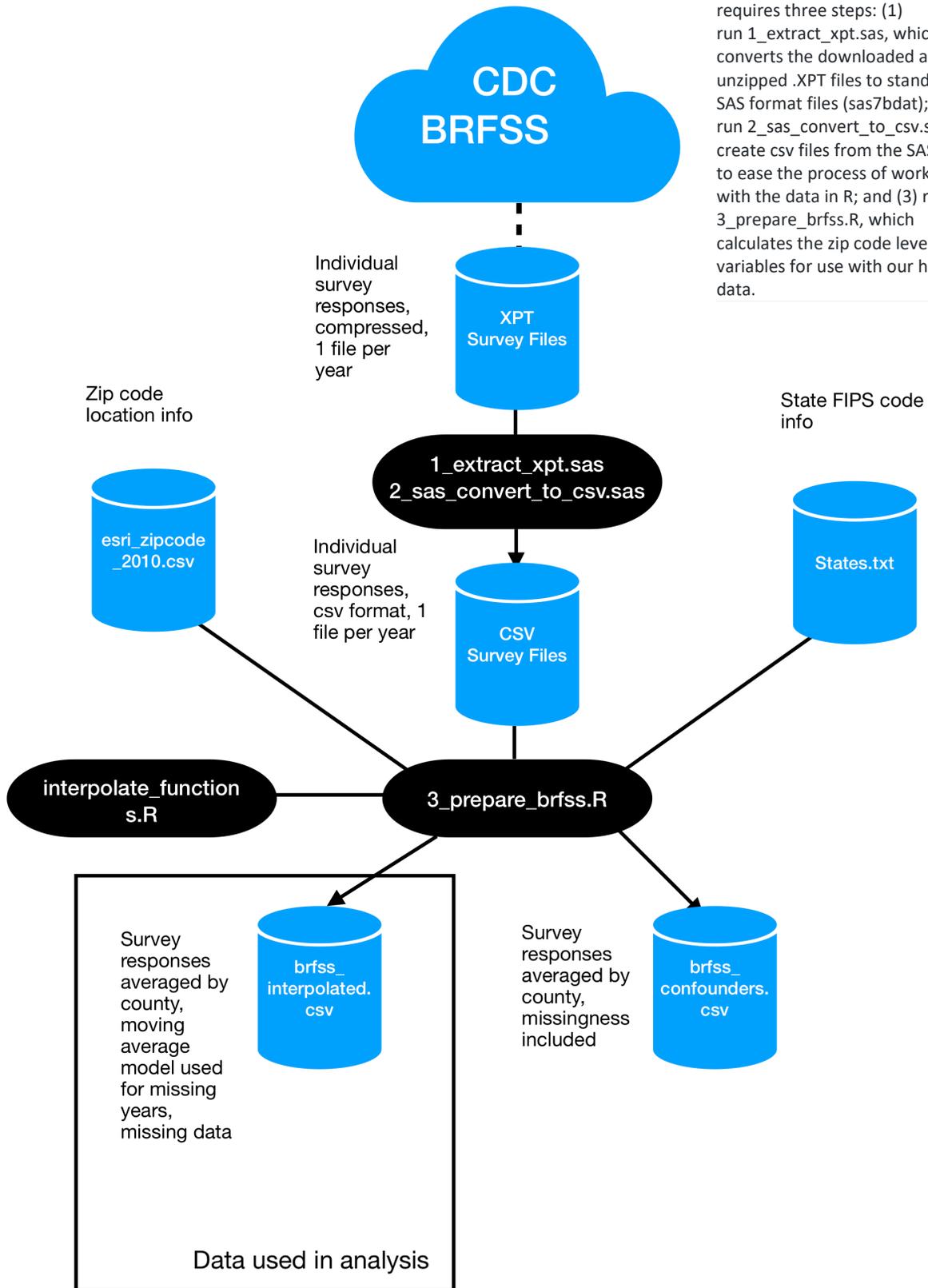
^aThe HRs relate three air pollutants to all-cause mortality among Medicare beneficiaries (2000–2016). These estimated HRs are obtained using both the GPS matching method and multivariate Poisson regression method under the assumption of a constant HR.

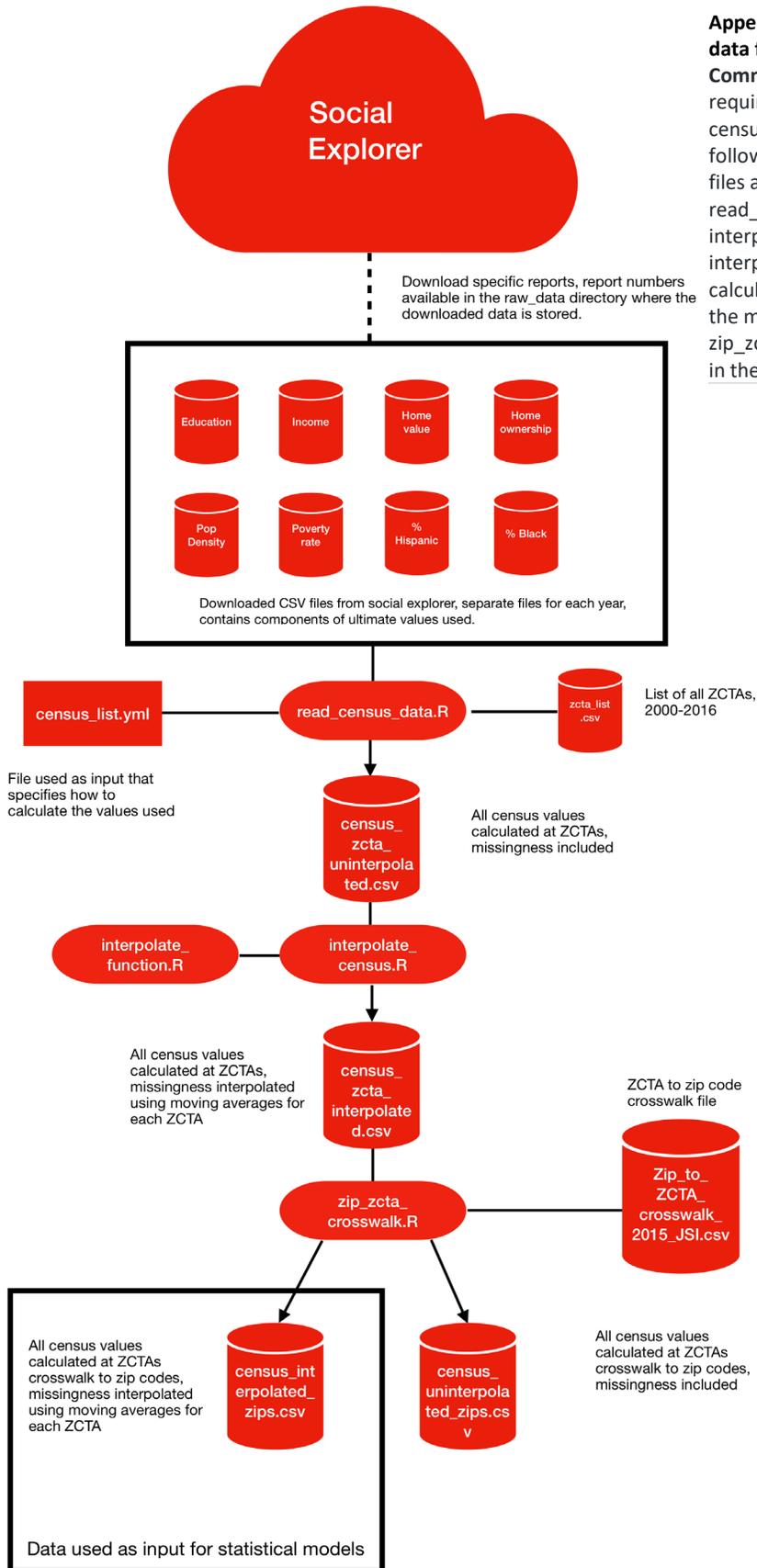
Appendix Table 2. Sensitivity Analysis Including Point Estimates and 95% CIs of the HRs Using Penalized Splines of Each Measured Potential Confounder^a

| Methods | Main Analysis |
|--|----------------------|
| GPS Matching | 1.261 (1.233, 1.289) |
| GPS Weighting | 1.268 (1.237, 1.300) |
| GPS Adjustment | 1.231 (1.180, 1.284) |
| Cox with Linear Adjustment | 1.369 (1.340, 1.399) |
| Poisson with Linear Adjustment | 1.347 (1.320, 1.375) |
| Poisson with Penalized Spline Adjustment | 1.293 (1.267, 1.320) |

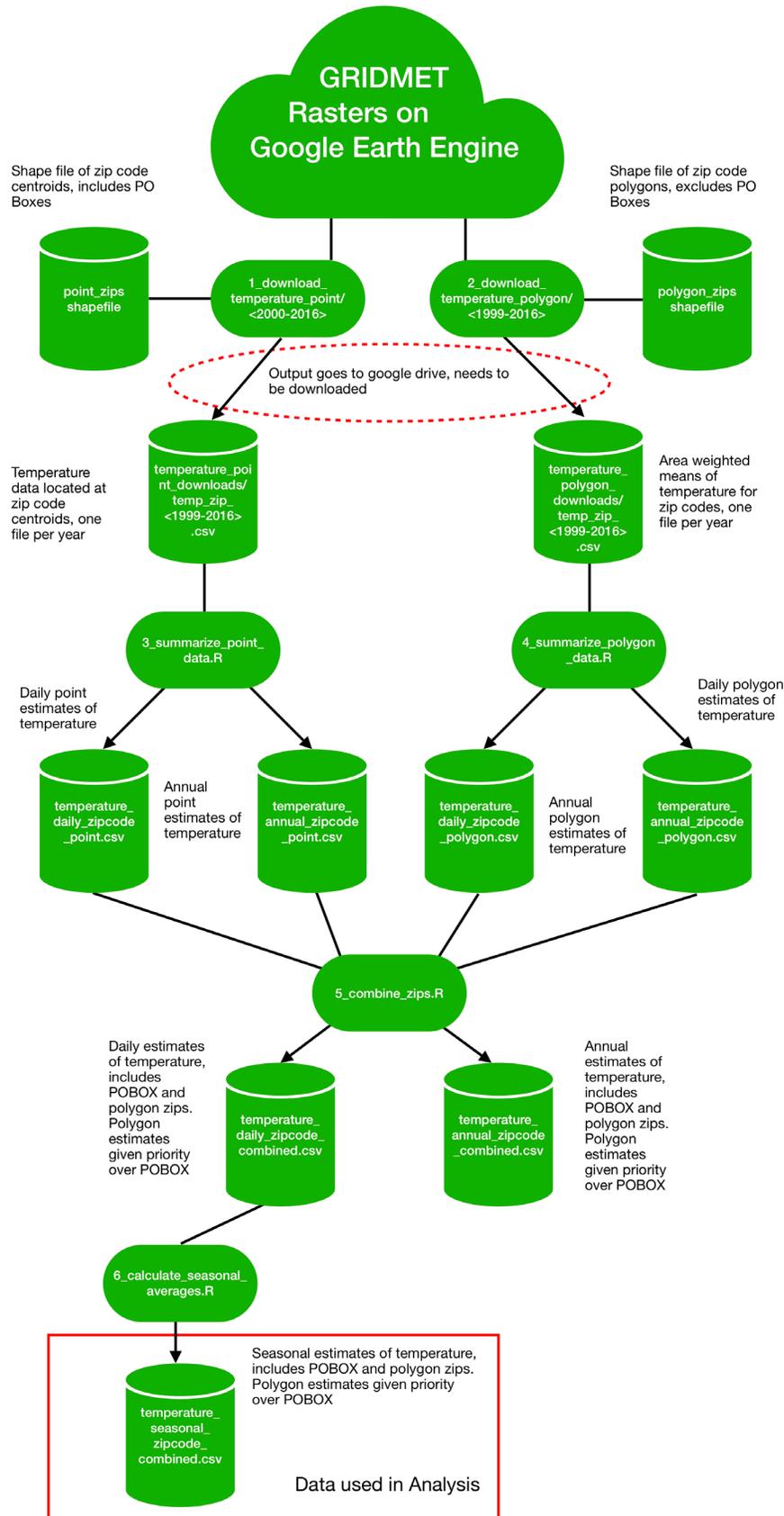
^a These estimated HRs are obtained for the analysis of long-term exposure to PM_{2.5} and all-cause mortality on the low-level cohort from 2000 to 2016.

Appendix Figure 2. Pipeline for processing data from CDC's BRFSS. Creating the final data requires three steps: (1) run `1_extract_xpt.sas`, which converts the downloaded and unzipped .XPT files to standard SAS format files (`sas7bdat`); (2) run `2_sas_convert_to_csv.sas` to create `csv` files from the SAS files, to ease the process of working with the data in R; and (3) run `3_prepare_bfss.R`, which calculates the zip code level variables for use with our health data.

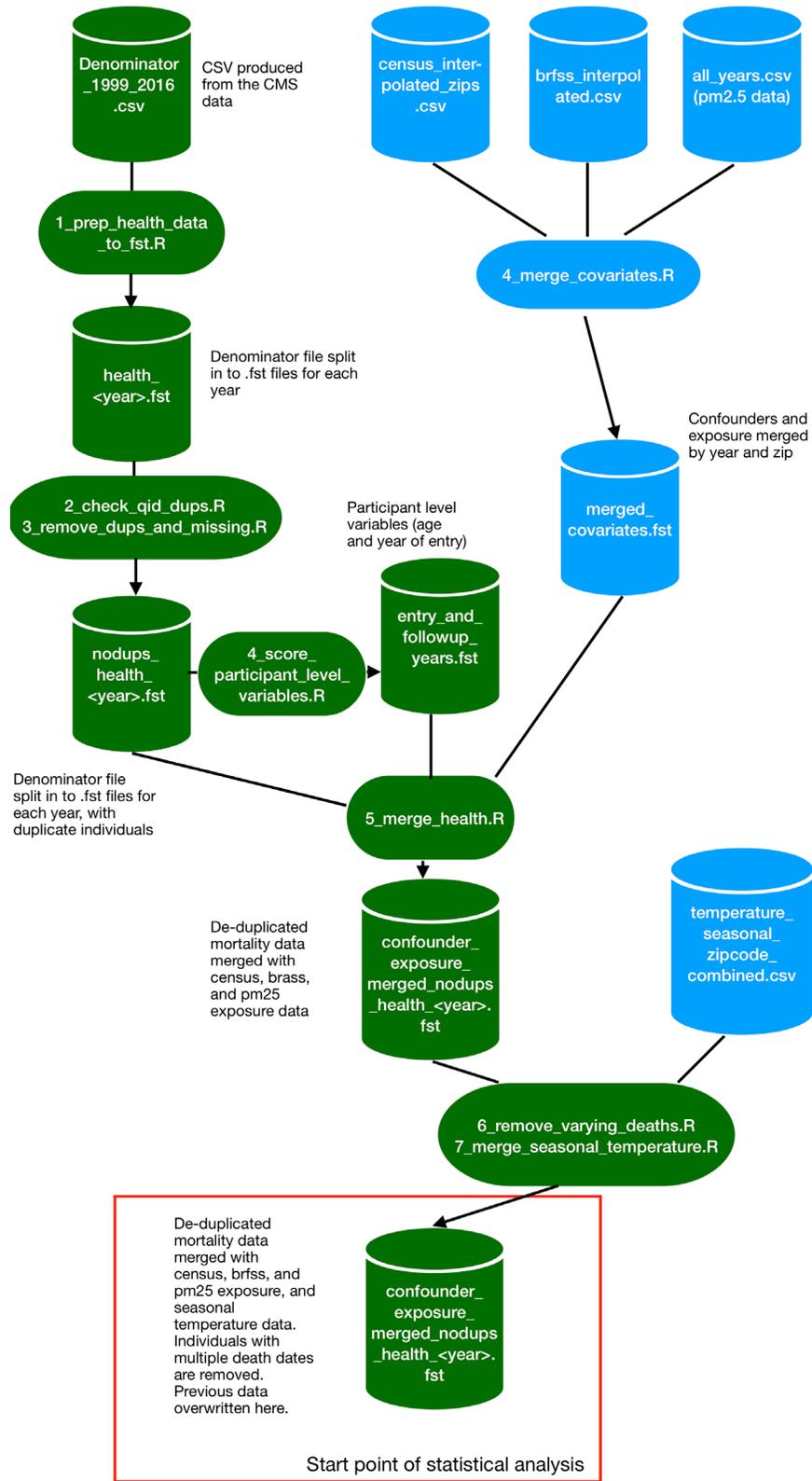




Appendix Figure 3. Pipeline for processing data from the U.S. Census and American Community Survey. Creating the final data requires running make.R, which creates census_interpolated_zips.csv using the following steps: (1) variables across multiple files are combined using read_census_data.R; (2) temporal interpolation is performed using interpolate_census.R. A moving average is calculated within each zcta and used to fill in the missing years; and (3) a crosswalk file, zip_zcta_crosswalk.R, is used to link the zctas in the dataset to zip codes.



Appendix Figure 4. Pipeline for processing data from the Google Earth Engine. Creating the final data requires two major steps: (1) using Google Earth Engine to aggregate the source data to the zip code level and downloading the data; and (2) combining the downloaded files into large single files and calculating seasonal and annual averages.



Appendix Figure 5. Pipeline for cleaning and merging exposure, covariate, and health data to produce combined datasets covering the period 1999–2016. Creating the final data requires the following steps: (1) run 1_prep_health_data_to_fst.R to create fst files for every year; (2) run check_qid_dups.R and 3_remove_dups_and_missing.R to find and remove duplicates; (3) run 4_merge_covariates.R to prepare covariates and exposure data; (4) run 4_score_participant_level_variables to create additional individual level variables; (5) run 5_merge_health.R to merge exposure, confounders, and health data; (6) run 6_remove_varying_deaths.R to remove individuals with multiple death days; and (7) run 7_merge_seasonal_temperature.R to merge temperature variables into the dataset.