

Comparing denominator sources for real-time disease incidence modeling: American Community Survey and WorldPop

Abstract

Across the United States public health community in 2020, in the midst of a pandemic and increased concern regarding racial/ethnic health disparities, there is widespread concern about our ability to accurately estimate small-area disease incidence rates due to the absence of a recent census to obtain reliable population denominators. 2010 decennial census data are likely outdated, and intercensal population estimates from the Census Bureau, which are less temporally misaligned with real-time disease incidence data, are not recommended for use with small areas. Machine learning-based population estimates are an attractive option but have not been validated for use in epidemiologic studies. Treating 2010 decennial census counts as a “ground truth”, we conduct a case study to compare the performance of alternative small-area population denominator estimates from surrounding years for modelling real-time disease incidence rates. Our case study focuses on modelling health disparities in census tract incidence rates in Massachusetts, using population size estimates from the American Community Survey (ACS), the most commonly-used intercensal small-area population data in epidemiology, and WorldPop, a machine learning model for high-resolution population size estimation. Through simulation studies and an analysis of real premature mortality data, we evaluate whether WorldPop denominators can provide improved performance relative to ACS for quantifying disparities using both census tract-aggregate and race-stratified modeling approaches. We find that biases induced in parameter estimates due to temporally incompatible incidence and denominator data tend to be larger for race-stratified models than for area-aggregate models. In most scenarios considered here, WorldPop denominators lead to greater bias in estimates of health disparities than ACS denominators. These insights will assist researchers in intercensal years to select appropriate population size estimates for modelling disparities in real-time disease incidence. We highlight implications for health disparity studies in the coming decade, as 2020 census counts may introduce new sources of error.

Keywords: Population denominators; Real-time incidence modeling; Health disparities

1 Introduction

Public health researchers regularly confront the need for near real-time, small area population estimates to serve as population denominators in epidemiologic studies. These studies often utilize present day disease incidence information, and seek to create incidence rates over small areas to properly characterize spatial patterns and dynamics of disease incidence. The calculation of reliable small-area incidence rates requires accurate estimates of the at-risk population size for areas and time periods coinciding with the incidence data. To investigate disparities in disease incidence across groups, demographically-stratified population size data are also needed. Public health research during the COVID-19 crisis has exemplified these challenges. In the United States in 2020, COVID-19 incidence data are available in near real time, but this is not the case for the population data, particularly for small areas.

US Census Bureau (USCB) products are the default for public health researchers seeking intercensal population estimates. Products from the USCB have the advantage of coming from a credentialed government agency, being aggregated by default to commonly-used geographies, and having detailed stratification (e.g., by age, racial/ethnic group, and sex). Systematic undercount of some racial and ethnic minority groups in USCB products is well-documented (Robinson et al., 1993; Robinson, 2011) and poses threats to the accuracy of disparity analyses employing population estimates stratified by racial/ethnic group. However, because more reliable data are rarely available for this purpose, USCB products remain the primary source of population denominator data for these analyses.

Likely the most commonly-used source of intercensal small-area population estimates today is the USCB’s American Community Survey (ACS) 5-year estimates, which are based on a rolling 5-year sample survey. ACS 5-year estimates are appealing due to their high spatial resolution— they are available for all census geographies larger than census blocks— and their rich demographic stratification. However, ACS data are released on a substantial delay, so that the most recent estimates available at any given time typically represent population sizes for the 5-year interval ending two years prior. The USCB also provides yearly population size estimates through its Population Estimates Program (PEP), which combines the decennial census with birth, death, and migration data (United States Census Bureau, 2020). Annual PEP estimates are available by racial/ethnic group, sex, and age at the national, state, and county level. For minor civil divisions and incorporated places, PEP provides total population estimates only.

Formally, the USCB recommends the use of PEP or decennial census counts as population size estimates in intercensal years, while using ACS for information about changing socioeco-

nomic and demographic features (United States Census Bureau, 2019). However, decennial census counts are often outdated, and PEP estimates are unavailable for commonly-used small census geographies such as census tracts and census block groups. In many epidemiologic contexts, such as studies of health disparities, a focus on these smaller geographies has been shown to be advantageous (Krieger et al., 2003). This creates confusion about the appropriate source of population estimates to use, and in practice many studies rely on ACS small-area estimates (Millett et al., 2020; Krieger et al., 2020a).

Private companies and academic groups have also begun to produce high-resolution gridded population estimates. These estimates are based on machine learning models that often combine census, remote sensing, land use, and other information to estimate population sizes at very small geographies in near real-time. One of the most popular products of this type is WorldPop, which utilizes an open-source algorithm and provides yearly global high-resolution population estimates (Stevens et al., 2015; Tatem, 2017). Advantages of WorldPop include its contemporaneous nature (available for the current year) and high spatial resolution. However, their estimates have undergone little validation, have not received wide acceptance in epidemiology, and are only stratified by age and sex.

Each source of population size data brings its own types of bias and uncertainty. To our knowledge, no studies have yet compared the performance of these data sources as denominators in the context of real-time statistical modeling of disease rates. In this paper, we describe the methods used to produce ACS and WorldPop population estimates relevant to disease mapping models. Then, we conduct a study in which 2010 small-area disease incidence rates are created using 2010 decennial census population counts, and we evaluate the bias induced when ACS and WorldPop data from 2010 and from one to two years prior are used instead as the denominator in classic health disparity models. These potential biases are illustrated using both a simulation study and a real case study of racial disparities in premature mortality in Massachusetts (MA). Throughout this paper, we aim to evaluate the performance of population denominator data sources when employed as we have observed them being used in epidemiologic practice (or as we anticipate they would be used in practice), and we do not intend to recommend that the data be used in each of the ways considered here. Our findings may assist epidemiologic researchers in assessing the biases that arise due to imperfectly measured population denominators and in determining whether novel high-resolution population estimate products like WorldPop can improve on default USCB products for modeling real-time small-area disease incidence rates.

2 Methods

2.1 ACS methodology

Since 2005 when ACS replaced the long-form decennial census, ACS continually collects US demographic data using a sophisticated sampling survey design. Briefly, within each county, ACS samples housing facilities from census blocks, the smallest unit of geography used by the USCB (600-3,000 residents). Each block is assigned a sampling rate that is inversely related to its population size. Sampling is structured so that no address will be selected more than once in a 5-year period (United States Census Bureau, 2014). For more detail, see Section S.1.

ACS employs four modes of data collection: internet, mail, telephone, and in-person visit. A complex survey weighting scheme is applied to adjust for sampling rates, to make estimates representative of larger area demographic characteristics (i.e., to conform to PEP estimates), to account for differential non-response by demographic group, to adjust for differences by data collection mode and seasonal population shifts, and more (Spielman et al., 2014). On the basis of these weighted data, ACS compiles 1-year, 3-year, and 5-year estimates. To achieve reliable estimates for small-areas, ACS must pool multiple years of data (Spielman et al., 2014). 1-year and 3-year population estimates are available only for geographies with populations of 65,000+ and 20,000+, respectively. The 3-year estimates were only produced from 2007 to 2013. The 5-year estimates are produced for all census geographies except census blocks, and they are available beginning with the 2005-2009 interval (United States Census Bureau, 2014). ACS 5-year population estimates are stratified by age, sex, and racial/ethnic group (see Section S.2 for more detail). ACS population estimates are also accompanied by margins of error. Following current practice in the disease mapping literature, we do not utilize these uncertainty measures in our analyses. We focus instead on the biases arising from the issue of temporal mismatch of incidence and denominator data in standard epidemiologic models.

The USCB references each ACS 5-year release by the final year covered by the interval, e.g., ACS 2008-2012 estimates are referred to as the ACS 2012 release. The USCB discourages the use of 5-year ACS estimates to represent the population size in the center year of the interval, yet in practice this is routinely done in the epidemiologic literature (Leas et al., 2019; Krieger et al., 2020b). Although we recognize this disclaimer by the USCB, throughout this paper we refer to ACS 5-year releases by their center year, to correspond to how they are largely used in practice, e.g., ACS 2008-2012 will be referred to as ACS 2010. USCB also

cautions against comparing ACS 5-year estimates that contain overlapping years (United States Census Bureau, 2019). In spite of this guidance, USCB does provide documentation explaining how to assess trends over time using overlapping ACS data (with caveats) (United States Census Bureau, 2009). Moreover, we have also noted in longitudinal studies that ACS population size estimates from overlapping years are used to track population changes over time (Hunt and Hurlbert, 2016; Mooney et al., 2018; Krieger et al., 2020b). Thus we will investigate the utility of time trends in population size estimates captured by overlapping consecutive ACS releases.

2.2 WorldPop methodology

WorldPop was created to generate accurate population counts that could be used to track the 2015 United Nations Sustainable Development Goals, with a focus on countries without timely and comprehensive census counts. WorldPop population size estimates are created by combining census counts with geospatial covariates (e.g. land cover and night lights) for a given year to produce gridded population estimates that are at a finer scale than typical administrative units (Stevens et al., 2015; Tatem, 2017). This process is often called “down-scaling” of census counts (Mennis, 2009). For years 2010, 2015 and 2020, WorldPop uses random forest algorithms to weight and disaggregate the census data onto a roughly 100m grid (referred to as a “top-down” modeling approach). For other years, gridded population counts are interpolated from these three years by applying a linear growth rate (Stevens et al., 2015). For countries without recent or reliable census data, they also conduct a second modeling step that incorporates reliable survey data (referred to as a “bottom-up” approach). For our analyses, we use US WorldPop estimates produced using a top-down, unconstrained estimation method. More details and justification are provided in Section S.3.

More detail about the geospatial covariates used in the US WorldPop models is provided in Section S.3. WorldPop also releases age- and sex-specific population counts produced by applying census age and sex proportions from larger geographies to the population estimate in each nested grid (for the US, WorldPop utilizes county-level age and sex proportions). WorldPop population datasets are available annually beginning in 2000 through the current year (Tatem, 2017; Stevens et al., 2015; WorldPop, 2020).

2.3 Data collection and alignment

Using the `tidycensus` R package (Walker, 2020), we extract census tract (CT) population counts stratified by age, sex, and racial/ethnic group from the 2010 decennial census for the

state of MA. Throughout our analyses, we consider these to be “ground truth” population counts, while acknowledging concerns about differential biases leading to both under- and over-counts (see Section 4). We collect ACS 5-year population size estimates stratified by age, sex and racial/ethnic group, for each CT for each year 2008-2010 (center years of ACS).

WorldPop population size estimates are available on a three arc-second grid (corresponding to 100m grids at the equator). We extract age- and sex-stratified gridded estimates for MA in each year 2008-2010 (WorldPop, 2018). We generate CT-aggregate population counts for each age and sex stratum by assigning each grid cell to a CT determined by the location of its centroid, and summing the stratum-specific population counts across grids within a CT. WorldPop does not produce population estimates stratified by racial/ethnic group, therefore we engineer them using ACS group proportions. For each age and sex stratum within each CT, we (1) compute the proportion of its population belonging to each racial/ethnic group based on the ACS and (2) multiply the racial/ethnic group proportions by the WorldPop stratum-specific estimates. This procedure yields CT-level WorldPop estimates stratified by racial/ethnic group, age, and sex. We use visualizations and summary statistics to compare decennial census, ACS, and WorldPop population size estimates cross-sectionally and with respect to temporal trends, with an eye towards understanding the effectiveness of using ACS and WorldPop estimates as a proxy for census counts.

2.4 Case study of premature mortality rate modeling

Our outcome of interest is premature mortality (death before age 65), a common choice for studying health disparities. We investigate disparities in premature mortality by socioeconomic status and racial/ethnic group. We primarily focus on comparisons of risk in Black and non-Hispanic White populations, motivated by longstanding health inequities between these groups and sufficiently large population sizes in MA to support the analyses (Krieger et al., 2003, 2020b). We obtain 2010 mortality data from the MA Department of Public Health. For each death certificate, the age, sex, racial/ethnic group, and location of residence for the deceased individual are recorded. We geocode the addresses and compute stratified premature mortality counts by CT (see Krieger et al. (2021) for more detail).

Using the 2010 decennial census counts and each set of ACS and WorldPop estimates (years 2008-2010) for ages 0-64, we compute age- and sex-standardized 2010 premature mortality ratios (SMR) for each MA CT using the indirect standardization method (Chen, 2013). The SMRs are calculated using the CT observed count in the numerator and an expected count for the CT, based on its population size and age and sex distribution, in the

denominator. This results in seven different sets of CT-level SMRs, each corresponding to a different set of denominator data. To study the impact of the different denominators in public health practice, we implement two modeling approaches that are commonly used to assess racial and socioeconomic disparities in premature mortality.

First, we fit CT-aggregate models to examine the association of CT SMRs (outcome) with CT racial composition and socioeconomic status. The explanatory variables, proportion of CT residents identifying as Black (PropBlack) and proportion below the poverty line (PropPov), are both taken from ACS 2010 across all models. Using each of the seven sets of SMRs described above (separately), we fit a Poisson regression model with spatial and independent random effects, following Besag et al. (1991).

As a second modeling approach, we consider a commonly-used variant of this model that stratifies by racial/ethnic group (hereafter referred to as race-stratified models for simplicity). This approach allows us to directly examine how within-CT Black and non-Hispanic White premature mortality rates differ, on average, as opposed to simply how the racial/ethnic composition of a CT is associated with CT premature mortality. Prior to model fitting, we separately age- and sex-standardize premature mortality within the non-Hispanic White and Black population within each CT, doing so with each of the seven sets of population size data. We retain these race-stratified SMRs as the outcomes in the modeling, and we analyze how the SMRs differ by racial/ethnic group (by including an indicator of Black vs. non-Hispanic White, $I(\text{Black})$, in the model) and how the SMRs are associated with the CT-aggregate proportion in poverty (PropPov). Again, the seven sets of SMRs are modeled separately, using a multi-level variant of the spatial Poisson regression model, following Leroux et al. (2000).

The mathematical formulation of each set of models is provided in Section S.4. All models are fit using a Bayesian approach implemented in the `CARBayes` package in R (Lee, 2013). We report incidence rate ratio (IRR) estimates based on the relevant exponentiated coefficient parameter from each model.

2.5 Simulation study

The aim of our simulation study is to assess the magnitude of biases induced in estimates of health disparities by using ACS/WorldPop denominators in standard models when the outcomes are generated using 2010 decennial census denominators. We conduct two sets of simulations, one using the age- and sex-standardized denominators described above and another using the crude population counts as denominators. Under each scenario, we sim-

ulate data from both a CT-aggregate model and a race-stratified model. The details of the data generation are provided in Section S.5. Briefly, we employ the same real MA CT-level covariates used in the real data analysis, and the 2010 decennial census CT denominators, to generate synthetic incidence data from a spatial Poisson model with known parameters. We randomly generate 100 such synthetic datasets for each of the CT-aggregate and race-stratified variants. We then fit models to each synthetic dataset, which are correctly specified except that the ACS/WorldPop denominators from 2008, 2009, and 2010 are used in place of the “true” 2010 census denominator. Bias in the resulting disparity estimates is assessed.

3 Results

3.1 Cross-sectional comparison of denominators

In Figure 1, we map the MA CT total population sizes from ACS and WorldPop for each year, alongside the 2010 decennial census counts (for Boston alone, see Figure S.1). Across both space and time, the distributions of the ACS and WorldPop total CT population size estimates appear highly similar to the decennial census. Using year 2010 data, there is a correlation of > 0.99 between ACS and census and between WorldPop and census (see Figures S.2 and S.3 for scatterplots). Figure S.4 shows the distribution of the difference in CT WorldPop and ACS estimates separately by population density quantile (per the 2010 census), and finds that, as CT urbanicity increases, the mean difference in WorldPop and ACS estimates increases (indicating higher estimates from WorldPop), as does the standard deviation.

Because accurate age distributions are critical in disease mapping, we also investigate how CT age-stratified ACS and WorldPop estimates compare to those from the decennial census. As shown in Figures S.5 and S.6, there is considerably more noise in the relationships between the age-stratified ACS/WorldPop estimates (year 2010) and the decennial census than was observed for the total population estimates. While the relationship between ACS and census is stable and linear for all age groups, the relationship between WorldPop and census estimates for certain age strata is erratic and possibly non-linear. This inconsistency likely occurs due to WorldPop’s method of obtaining age-stratified estimates, whereby age distributions for larger areas are projected onto the nested gridded high-resolution total population estimates.

More problematic for disparity modeling, errors in WorldPop’s age-stratified estimates appear to be associated with key measures of disadvantage. To illustrate this phenomenon,

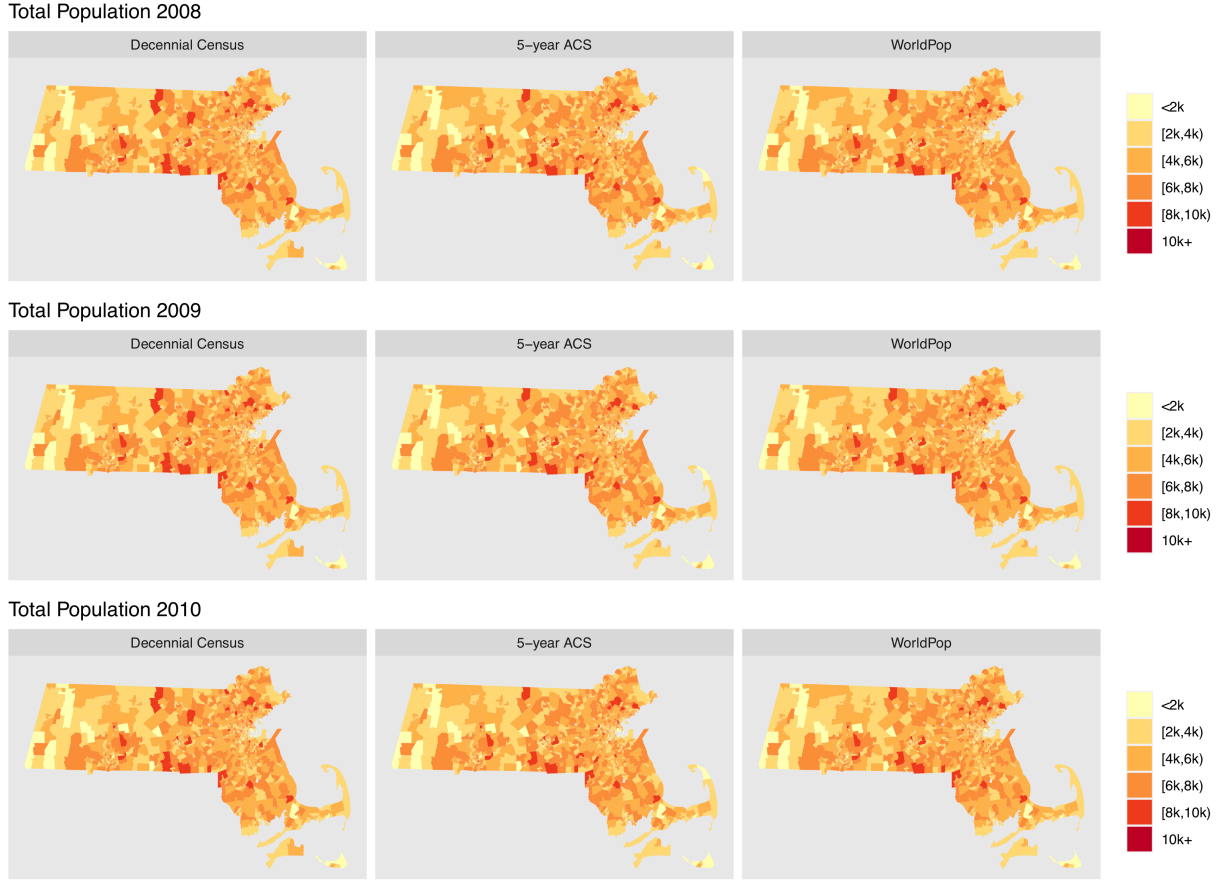


Figure 1: Spatial distribution of census tract-level 2010 decennial census population counts compared to 5-year ACS and WorldPop population estimates for years 2008-2010, Massachusetts, USA.

in Figure 2 we show the difference in decennial census and WorldPop 2010 age-stratified CT population estimates plotted against estimates of the percent of the CT in poverty (from the 2010 ACS). Roughly speaking, for a given age group, a positive relationship between these variables indicates that WorldPop tends to over-estimate population size (relative to census) for high-income CTs. Conversely, a negative relationship indicates that WorldPop tends to under-estimate population size for high-income CTs. From this Figure, it is clear that in high-income CTs, WorldPop tends to over-estimate the number of young people and under-estimate the number of older people. To understand why this might happen, consider projecting the age distribution of a large and diverse county, e.g., a county containing a major city, onto each nested CT, which is WorldPop's approach. Because higher income neighborhoods tend to have fewer children and more older people (and vice-versa for low

income neighborhoods), applying the age distributions of the whole city to a high income neighborhood would generally lead to over-estimation of the number of children in the CT and under-estimation of the number of older people. No such systematic biases (relative to the census) appear to exist in the ACS, as shown in Figure S.7.

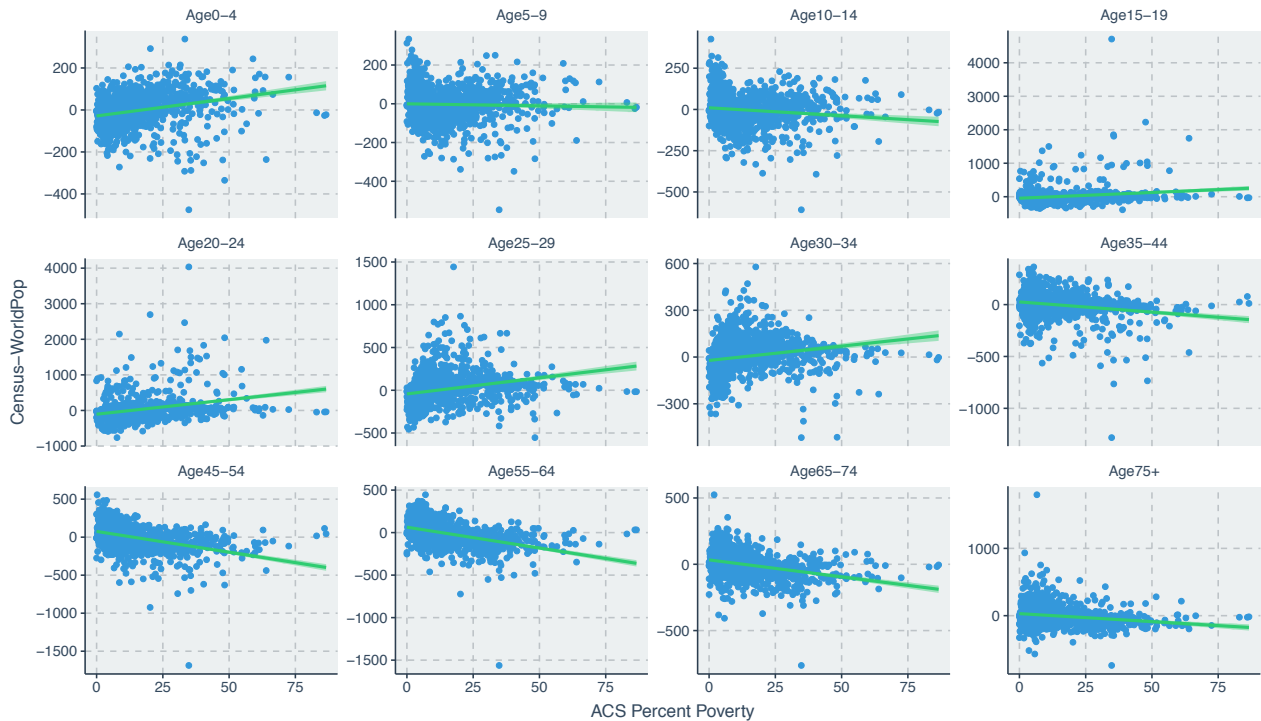


Figure 2: Scatterplots of difference in WorldPop 2010 and decennial census age-stratified CT population estimates vs. percent of the CT in poverty.

3.2 Temporal trends

Figure 3 shows the temporal changes in ACS CT population sizes, for both the total population and the Black population, grouped by county over the years 2008-2010. Figure S.8 presents the same plot for non-Hispanic Whites. The population size estimates for each year are shown as a proportion of the 2010 decennial census population count. For the total CT population, most of the ACS changes over time are within 10% of the 2010 census population size. Moreover, discrepancies in ACS and 2010 census population sizes do not, on the whole, appear to diminish as the ACS year approaches 2010. Many of the largest relative changes over time are in CTs with small populations. This suggests that ACS changes over time in the total CT population size may be attributable largely to sampling variability. For CT-

level Black population size estimates, we tend to see (a) much larger relative discrepancies in ACS and decennial counts and (b) much larger relative changes over time in ACS. Again, the largest relative changes tend to occur in CTs with small Black populations.

Using the procedures recommended by USCB for investigating time trends using overlapping ACS estimates (United States Census Bureau, 2009), we compute the proportion of CTs with statistically significant changes in Black or non-Hispanic White population sizes for each combination of years. The results are shown in Table S.6. From 2008 to 2010, the ACS Black population size estimates changed significantly for 22% of CTs and the non-Hispanic White population size estimates changed significantly for 29% of CTs.

Figure S.9 shows the time trends in WorldPop CT population sizes relative to the decennial census. Within each MA county, the CT population size estimates over time nearly all follow a common linear trend. This reflects WorldPop’s use of linear interpolation of the 2010 estimates to produce population size estimates for prior years. Naturally, this procedure leads to estimates that are highly model-dependent at fine levels of spatial and temporal granularity.

3.3 Premature mortality modeling results

IRR estimates and 95% Bayesian credible intervals from the CT-aggregate and race-stratified models are presented in Table 1. Because our aim is to understand the impact of different population denominators in real applications, we focus on the discrepancies in IRR estimates across models. For convenience, we refer to the IRR for the racial/ethnic group covariate in the models as the “race IRR” and to the IRR for the percent poverty covariate as the “poverty IRR”. For a discussion of the implications of our findings with regards to racial/ethnic and socioeconomic disparities in premature mortality, see Section S.4.3.

In the CT-aggregate models, we observe minimal changes in the IRR estimates across years for a given denominator source. This suggests that temporal mismatch has little impact on the IRR estimates in this context. The use of ACS denominators induces little bias relative to the decennial census, with only a slight upward bias in the poverty IRR estimate. WorldPop denominators lead to a substantial downward bias in the poverty IRR estimate relative to the census, attributable to the differential error in WorldPop’s age-stratified population size estimates for neighborhoods with different poverty levels. Underestimation of the number of older people and overestimation of the number of young people in high income CTs results in expected premature mortality counts (denominators) that are artificially low, and accordingly erroneously high SMRs in high income CTs. The opposite occurs for low

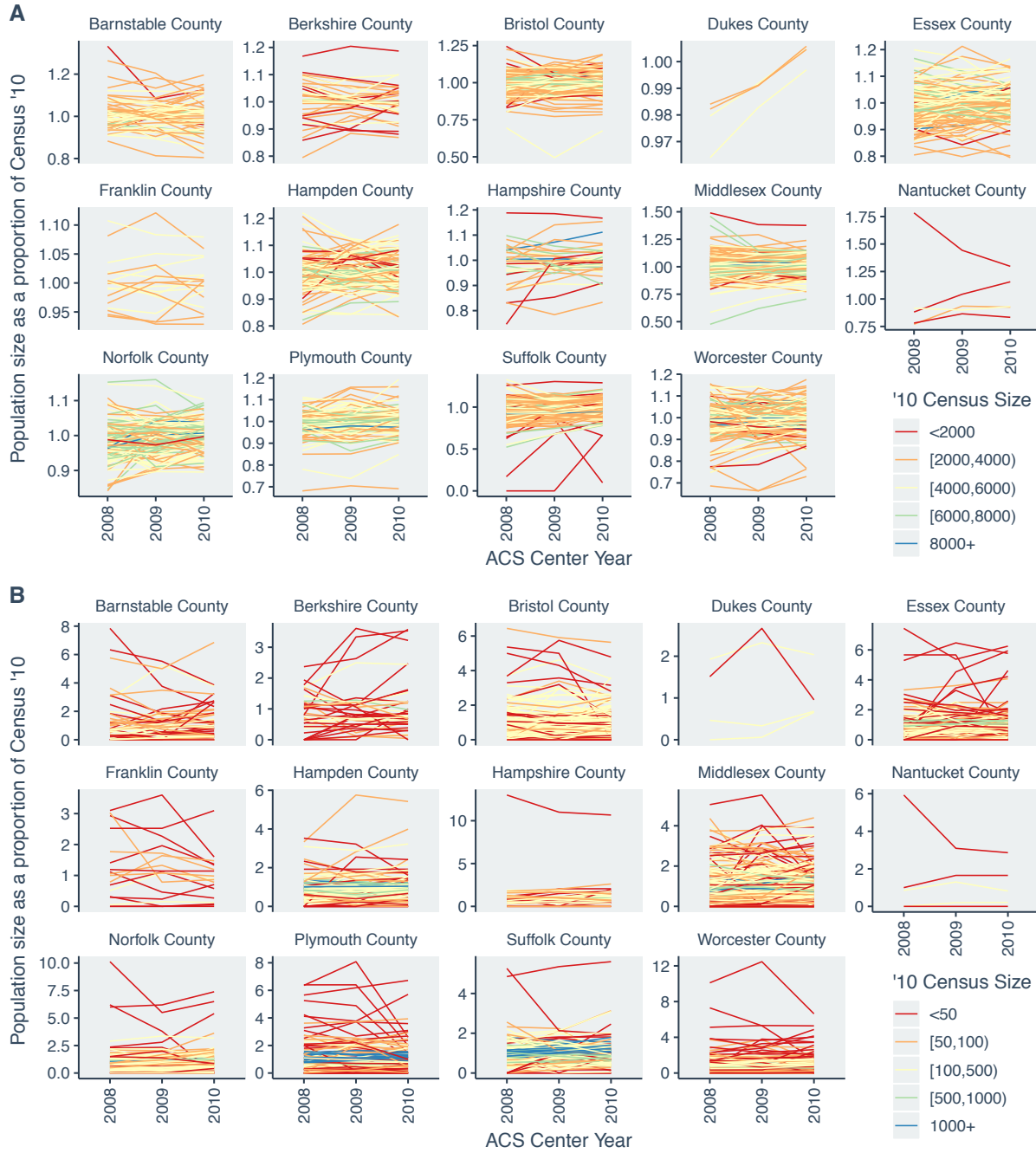


Figure 3: MA census tract ACS total population size estimates (A) and Black population size estimates (B) over time as a proportion of 2010 decennial census population size. Colors represent 2010 decennial census population size bins.

income CTs, and this produces a downward bias in the poverty IRR estimate.

In the race-stratified models, temporal mismatch of incidence data and ACS/WorldPop

denominators may lead to greater bias in disparity estimates. Recall that WorldPop population sizes stratified by racial/ethnic group are derived from the racial/ethnic distributions in ACS, so we would expect to see similar trends in the race IRR estimates across denominator sources. Indeed, for ACS and WorldPop, the race IRR estimates in a given year are highly similar. Relative to the 2010 census results, the 2008 ACS/WorldPop race IRR estimates are the most biased, with bias diminishing as the year of denominator data approaches 2010. As in the CT-aggregate models, the poverty IRR estimate is slightly upwardly biased when using ACS denominators and substantially downwardly biased when using WorldPop denominators.

Table 1: Incidence rate ratio estimates (95% credible intervals) from premature mortality models. In the CT-aggregate models, the Race variable is an ecologic variable (CT proportion Black), while in the race-stratified models, the Race variable is a group-level binary indicator of Black (versus non-Hispanic White). In both models, the Poverty variable is CT-aggregate proportion in poverty. Continuous covariates are centered and scaled.

		CT-aggregate			Race-stratified		
		ACS	WP	Census	ACS	WP	Census
2008	Intercept	0.97 (0.95,0.99)	0.95 (0.93,0.97)	-	1.05 (1.03,1.08)	1.02 (1.00,1.05)	-
	Race	1.09 (1.05,1.14)	1.10 (1.06,1.15)	-	0.95 (0.88,1.03)	0.97 (0.89,1.05)	-
	Poverty	1.22 (1.17,1.27)	1.10 (1.06,1.14)	-	1.34 (1.29,1.39)	1.19 (1.15,1.24)	-
2009	Intercept	0.97 (0.95,0.99)	0.95 (0.93,0.97)	-	1.06 (1.03,1.08)	1.03 (1.01,1.06)	-
	Race	1.09 (1.05,1.13)	1.10 (1.05,1.14)	-	0.92 (0.84,1.00)	0.92 (0.84,1.01)	-
	Poverty	1.22 (1.18,1.26)	1.10 (1.06,1.14)	-	1.34 (1.29,1.39)	1.20 (1.16,1.24)	-
2010	Intercept	0.97 (0.95,0.99)	0.95 (0.93,0.97)	0.97 (0.95,0.99)	1.07 (1.04,1.10)	1.04 (1.02,1.07)	1.07 (1.05,1.09)
	Race	1.09 (1.05,1.14)	1.10 (1.05,1.14)	1.09 (1.05,1.14)	0.88 (0.82,0.95)	0.88 (0.81,0.97)	0.89 (0.81,0.96)
	Poverty	1.22 (1.18,1.26)	1.11 (1.07,1.15)	1.19 (1.15,1.23)	1.35 (1.31,1.40)	1.21 (1.17,1.26)	1.32 (1.28,1.36)

3.4 Simulation results

Results for the CT-aggregate and race-stratified simulations using standardized denominators are shown in panels A and B, respectively, of Figure 4. Figure S.10 shows analogous results from the simulations using crude denominators. The coefficient estimates from the 100 simulated datasets are summarized in boxplots, with the true parameter values indicated by a black horizontal line. All biases in the coefficient estimates are assumed to be solely due to misspecification of denominators (covariates are correctly specified).

The simulation results are consistent with the findings of the real data analysis. In the CT-aggregate models, the use of ACS or WorldPop denominators induces little bias in coefficient estimates, with the exception of the poverty coefficient. Moreover, the bias does not consistently diminish as the year represented by the denominator data approaches

2010. This suggests that, over relatively short periods, the time trends in ACS denominators representing the total CT population may be attributable more to sampling variability than to real population changes. Moreover, the linear interpolation of WorldPop estimates over time appears to have little impact on health disparity estimates in this context.

In the race-stratified simulations, we generally observe more severe bias in all coefficient estimates. In addition to bias in the poverty coefficient (primarily for WorldPop), here we also observe substantial bias in the race coefficients in years prior to 2010, i.e., an upward bias of about 50% for ACS 2008 and even higher for WorldPop. This suggests that differential biases in ACS population size estimates by racial/ethnic group may be more problematic for modeling disparities than the bias in total CT population sizes. However, bias in the race coefficient estimates tends to decrease as the year of the data approaches 2010. This improved performance suggests that changes in the racial/ethnic composition of a CT in ACS, even over short periods, may represent meaningful changes rather than sampling variability. This agrees with our finding above that ACS racial/ethnic group-specific population size estimates changed significantly for roughly 25% of CTs between 2008 and 2010.

In simulations using the crude denominators (Figure S.10), ACS and WorldPop denominators generally perform comparably well and yield estimates with little bias. An exception is the race coefficient in the race-stratified models, which is heavily biased for both ACS and WorldPop and across all years. This is likely a product of the small Black populations in many MA CTs, i.e., the very small crude denominators may cause instability in the race coefficient. In this simulation scenario, because we do not impose differing disease risks by age stratum and do not use standardized denominators, the issue of bias in the poverty coefficient is eliminated.

4 Discussion

In this paper, we explore the impact of using imperfect and temporally mismatched population size estimates to model small-area, real-time disease incidence rates for epidemiologic studies. In a case study of MA CT-level data, we found no evident advantages of using WorldPop population size estimates, either alongside or in place of the default ACS estimates, to create small-area incidence rates for real-time epidemiologic modeling. This is in spite of the fact that, in practice, WorldPop estimates may be available contemporaneously while ACS estimates are from prior years. WorldPop’s method of obtaining high-resolution age-stratified population size estimates, i.e., by projecting county-level age distributions onto small area total population size estimates, proves particularly problematic for modeling

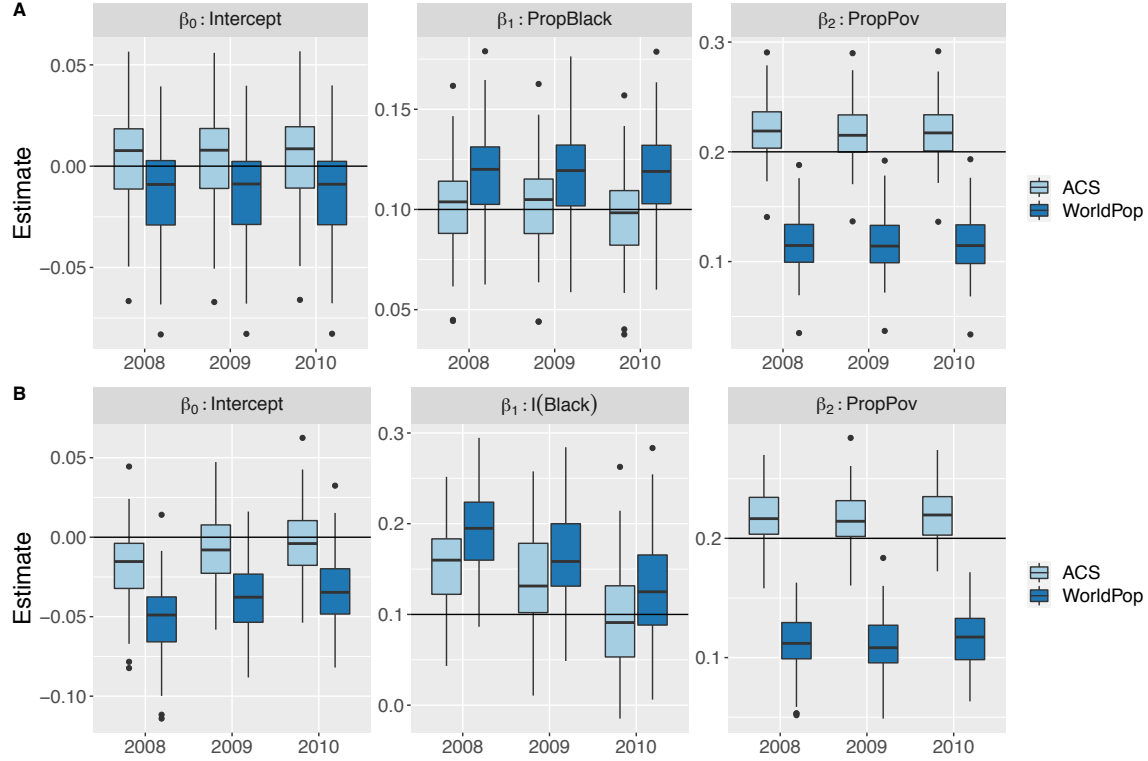


Figure 4: Simulation results using standardized denominators. Parameter estimates using ACS and Worldpop population size estimates in CT-aggregate (A) and race-stratified (B) models. Data are generated using 2010 decennial census population sizes. True values of each parameter are denoted by the black horizontal lines.

health disparities. This approach induces error in WorldPop’s age-stratified estimates that is associated with key measures of disadvantage, such as poverty, leading to bias in health disparity estimates when using age-standardized WorldPop denominators.

By comparing analyses using ACS denominators and “ground truth” 2010 decennial census denominators, we demonstrate that the impact of the inevitable temporal incompatibility of ACS denominators for quantifying real-time disparities in disease incidence depends on the modeling approach being used. Over short time periods (several years), changes in total CT population size as measured by ACS may be attributable primarily to sampling variability, and this variability may have little to no impact on health disparity estimates. On the other hand, changes over time in CT racial composition measured by ACS likely represent meaningful changes, and using outdated population estimates in race-stratified disparity models may induce substantial bias in disparity estimates that could lead to incorrect inference. These findings are consistent with ACS’s documentation, which states that ACS should

be used primarily for tracking changes in the characteristics of areas (racial composition, socioeconomic status) rather than tracking changes in population sizes.

To our knowledge this is the first study that compares the impact of biases in commonly-used population size estimates for real-time, small-area disease modeling. The primary limitation of our study is its narrow temporal and geographic scope. The quality of both ACS and WorldPop population estimates varies across space and time, limiting the generalizability of our work. WorldPop relies on remote sensing data, the quality of which differs by geographic region (Elvidge et al., 2017). ACS data quality is known to be poorer for low-income and African American populations (Spielman et al., 2014), leading to regional differences in data quality. Thus, our findings in a single Northeastern state may not be representative of other states and regions. We are also utilizing the decennial census population counts as the ground truth population sizes for 2010. Differential undercounts in the decennial census, primarily impacting low-income and minority communities, are well-documented (Robinson et al., 1993; Robinson, 2011). In the 2010 census, the USCB found an overall 2.1% undercount of the Black population and a 0.8% overcount of the non-Hispanic White population (United States Census Bureau, 2012). These issues may be further exacerbated for small-areas, which should be taken into consideration when interpreting our results. However, census counts serve as a standard referent population in the US.

In spite of numerous longstanding USCB programs and policies aimed at reducing undercounts, the USCB has struggled to overcome the many and complex challenges that lead to undercounting of certain disadvantaged groups and overcounting of certain privileged groups (O’Hare, 2019). These systematic errors could lead to either over-estimation or under-estimation of health disparities, depending on how the magnitude of error varies across age and sex groups. Initiatives like the Census Post-Enumeration Survey (also called the Accuracy and Coverage Evaluation or the Census Coverage Measurement Survey) and the Demographic Analysis program provide nation-level estimates of census under/overcount of various demographic groups. However, because the nature and magnitude of under/overcounting varies across space, simply scaling small-area counts to adjust for national-level under/overcount estimates is unlikely to correct the problem. Moreover, our analyses with WorldPop have revealed how attempts to produce ultra high-resolution population size estimates, by downscaling census counts and making strong assumptions about homogeneity of demographic features across space, can induce further errors that severely bias health disparity estimates.

The collection of exceptional events affecting 2020 census data collection— including the

COVID-19 pandemic, increasing concern about law enforcement practices, the consideration of the addition of a citizenship question on the census, and the move to more online census data collection— are likely to introduce new sources of counting error (The White House, 2021; Jarmin, 2021). Moreover, USCB will, for the first time, apply differential privacy procedures to public-release 2020 census data, the consequences of which are still not fully understood for small area disease modeling applications (Krieger et al., 2021). These challenges will complicate future epidemiologic studies using decennial census counts as a “gold standard” and will necessitate novel approaches for addressing differential under/overcounts to accurately estimate health disparities. For populations age 65+, Medicare enrollment data, which are made available to researchers by the US Centers for Medicare and Medicaid Services (CMS), may provide a more accurate and representative source of small-area intercensal denominator information than USCB products. Over 96% of Americans age 65+ are enrolled in Medicare (Di et al., 2017), and the data provided by CMS contain individual-level age, sex, racial/ethnic group, and zipcode identifiers for enrollees. While Medicaid enrollment data are not similarly representative of a common population due to heterogeneity in state-level programs, these data could likely be used to create small-area-specific scaling factors to adjust for undercounts of low-income populations (including young people).

Given the unprecedented quantity and richness of data available today, the future of high-resolution population size estimation may lie in combining information from USCB products and other data sources. It is essential that these efforts, in addition to improving accuracy in total population size estimates, consider how the integration of other data sources can address census under-representation of disadvantaged groups. Finally, the development and application of statistical methods that can better account for population size uncertainties/measurement error in disease mapping models are critical to address these issues.

References

- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.
- Chen, J. T. (2013). Multilevel and hierarchical models for disease mapping. *In: Boscoe, FP (ed). Geographic Health Data: Fundamental Techniques for Analysis*, pages 183–208.
- Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., and Schwartz, J. D. (2017). Air pollution and mortality in the Medicare population. *New England Journal of Medicine*, 376(26):2513–2522.
- Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., and Ghosh, T. (2017). Viirs night-time lights. *International Journal of Remote Sensing*, 38(21):5860–5879.
- Hunt, B. R. and Hurlbert, M. S. (2016). Black:white disparities in breast cancer mortality in the 50 largest cities in the United States, 2005–2014. *The International Journal of Cancer Epidemiology, Detection, and Prevention*, 45:169–173.
- Jarmin, R. (2021). *2020 Census Processing Updates*. <https://www.census.gov/newsroom/blogs/director/2021/02/2020-census-processing-updates.html>.
- Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., and Subramanian, S. (2003). Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—The Public Health Disparities Geocoding Project. *American Journal of Public Health*, 93(10):1655–1671.
- Krieger, N., Nethery, R. C., Chen, J. T., Waterman, P. D., Wright, E., Rushovich, T., and Coull, B. (2021). Impact of differential privacy and census tract data source (decennial census versus American Community Survey) for monitoring health inequities. *American Journal of Public Health*, 111(2):265–268.
- Krieger, N., Waterman, P. D., and Chen, J. T. (2020a). COVID-19 and overall mortality inequities in the surge in death rates by ZIP Code characteristics: Massachusetts, January 1 to May 19, 2020. *American Journal of Public Health*, 110(12):1850–1852.
- Krieger, N., Wright, E., Chen, J. T., Waterman, P. D., Huntley, E. R., and Arcaya, M. (2020b). Cancer stage at diagnosis, historical redlining, and current neighborhood characteristics: Breast, cervical, lung, and colorectal cancers, Massachusetts, 2001–2015. *American Journal of Epidemiology*, 189(10):1065–1075.

- Leas, E. C., Schleicher, N. C., Prochaska, J. J., and Henriksen, L. (2019). Place-based inequity in smoking prevalence in the largest cities in the United States. *JAMA Internal Medicine*, 179(3):442–444.
- Lee, D. (2013). CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, 55(13):1–24. <http://www.jstatsoft.org/v55/i13/>.
- Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer.
- Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass*, 3(2):727–745.
- Millett, G. A., Jones, A. T., Benkeser, D., Bara, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S., Sherwood, J., and Sullivan, P. S. (2020). Assessing differential impacts of COVID-19 on Black communities. *Annals of Epidemiology*, 47:37–44.
- Mooney, A. C., Giannella, E., Glymour, M. M., Neilands, T. B., Morris, M. D., Tulskey, J., and Sudhinaraset, M. (2018). Racial/Ethnic Disparities in Arrests for Drug Possession After California Proposition 47, 2011–2016. *American Journal of Public Health*, 108(8):987–993.
- O’Hare, W. P. (2019). *Differential undercounts in the US census: who is missed?* Springer Nature.
- Robinson, J. G. (2011). *Coverage of population in Census 2000 based on demographic analysis: The history behind the numbers*. US Census Bureau, Population Division.
- Robinson, J. G., Ahmed, B., Gupta, P. D., and Woodrow, K. A. (1993). Estimation of population coverage in the 1990 United States Census based on demographic analysis. *Journal of the American Statistical Association*, 88(423):1061–1071.
- Spielman, S. E., Folch, D., and Nagle, N. (2014). Patterns and causes of uncertainty in the American Community Survey. *Applied Geography*, 46:147–157.

- Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one*, 10(2):e0107042.
- Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, 4(1):1–4.
- The White House (2021). *Executive Order on Ensuring a Lawful and Accurate Enumeration and Apportionment Pursuant to the Decennial Census*. <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-ensuring-a-lawful-and-accurate-enumeration-and-apportionment-pursuant>
- United States Census Bureau (2009). *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*. <https://www.census.gov/content/dam/Census/library/publications/2009/acs/ACSResearch.pdf>.
- United States Census Bureau (2012). *Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census*. https://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html.
- United States Census Bureau (2014). *American Community Survey Design and Methodology (January 2014)*.
- United States Census Bureau (2019). *Comparing ACS Data*. <https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html>.
- United States Census Bureau (2020). *Methodology for the United States Population Estimates: Vintage 2019*.
- Walker, K. (2020). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. R package version 0.9.9.2.
- WorldPop (2018). *Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076)*. www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur and Center for International Earth Science Information Network (CIESIN), Columbia University. <https://dx.doi.org/10.5258/SOTON/WP00670>.
- WorldPop (2020). *WorldPop gridded population estimate datasets and tools. How are they different and which should I use?* <https://www.worldpop.org/methods/populations>.

WorldPop (2021). *Top-down estimation modelling: Constrained vs. Unconstrained*. https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained.