Uncertainty in Life Expectancy in Small Areas When Using ACS Population Estimates

Corey S. Sparks
University of Texas at San Antonio
Department of Demography

May 19, 2021

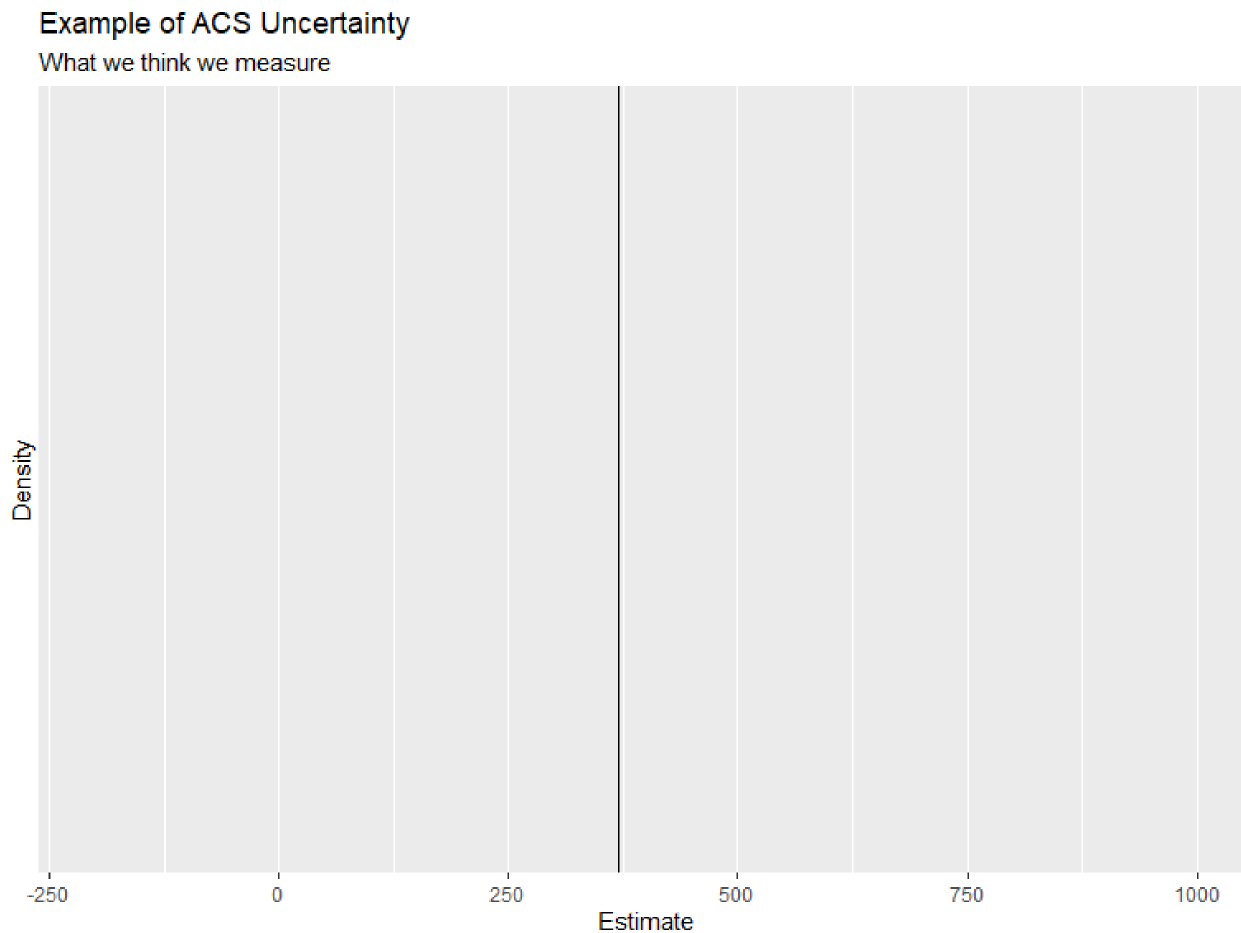
American Community Survey

- Began in the mid 2000's to replace the US Census Bureau's socioeconomic sample data (long form)
  - Based on a 1 in 20 sample of the US population
- Summary file data are available for all but the smallest Census geographies
  - The estimates at the smaller geographies (tracts and block groups) are often based on very small sample sizes
- High-variance estimates for these areas, but also for larger areas when segmenting the population (e.g. Spielman et al., 2014, 2015)
  - e.g. poverty rates for minority groups in small geographies

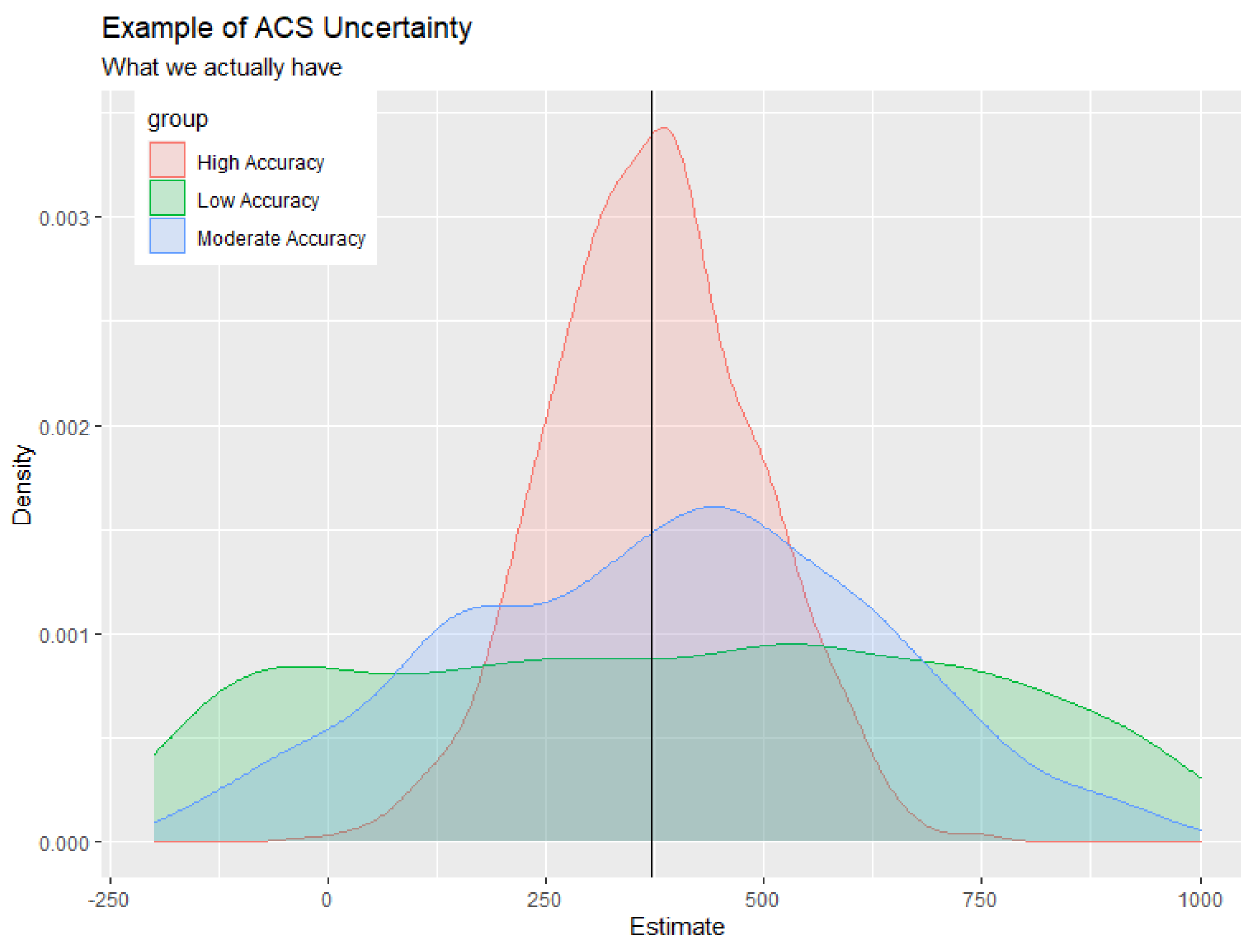
American Community Survey - Margins of Error

- Census Bureau publishes point estimates and margins of error for all estimates
- Census ***strongly encourages*** users to be mindful of the margins of error
- Researchers in many fields use these estimates for all manner of analyses
  - Measures of socioeconomic conditions, housing, environmental variables
  - Hundreds of studies have used these data
- Napierala and Denton 2017 use ACS uncertainty to bootstrap MSA segregation indices.
- Orndahl and Wheeler epidemiological study includes measurement error in the ACS variables.

Measurement error example -Under Age 5 in an **“Average”** Tract



Measurement error example



Measurement Error Models

- Models for measurement error have been around for a while
- Berkson error model (Berkson, 1950)
- Classical error model (Various authors)

Measurement Error in Rates

- Statistical models for rates often use the Poisson or Negative Binomial distributions
- These models are based on the observed count being the rate numerator, and the denominator, modeled as an offset term, comes from a population estimate.
- When the ACS is used as the population estimate, say for small area rate estimates, then **the error must be considered, as it introduces uncertainty in the rates.**

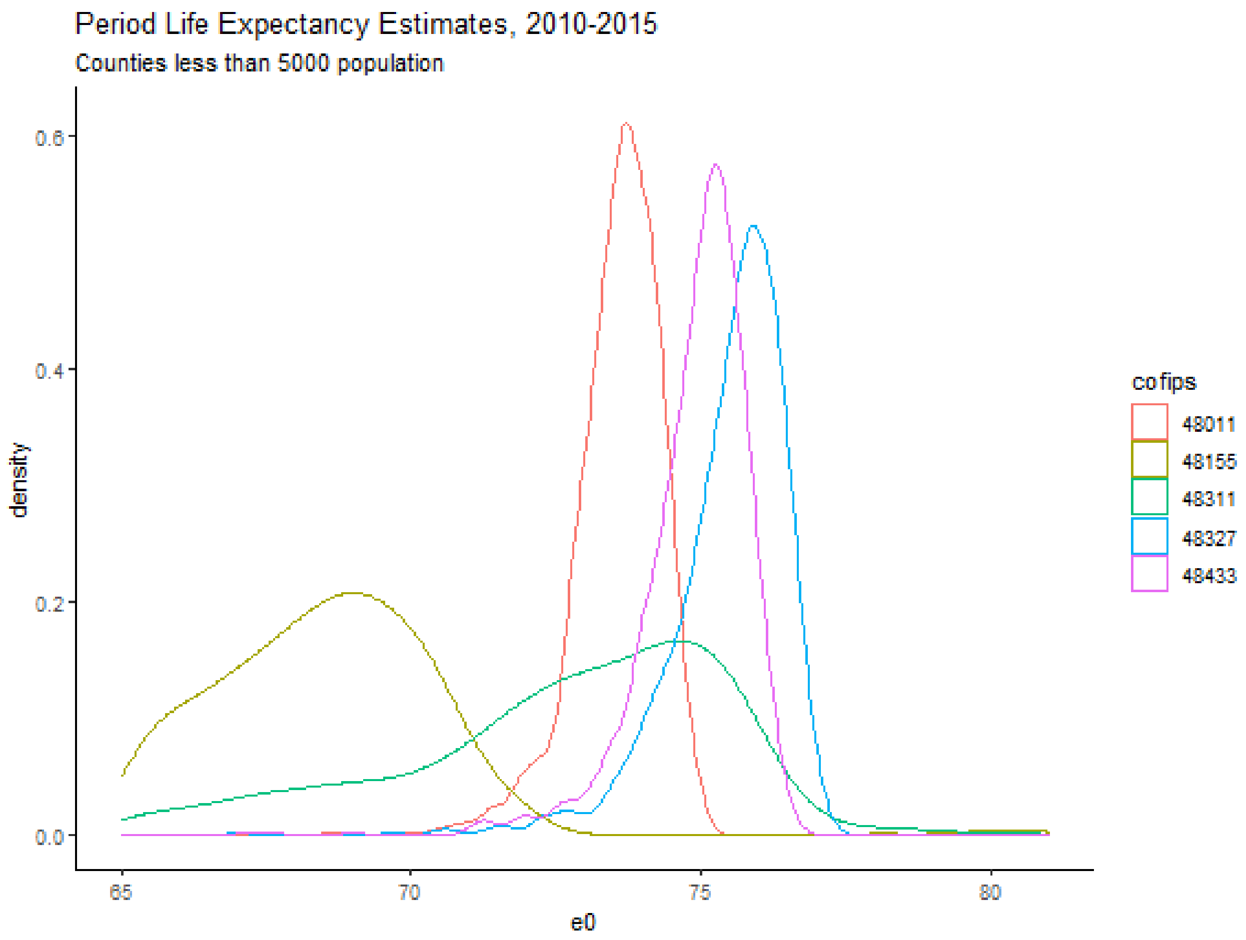
Methods

– Since the classical error model is basically a *latent variable model*, with Z observed only partially, then the Bayesian modeling paradigm is very useful
– ASMRs are estimated, with the number of deaths being the numerator, and the uncertain population size as the denominator.
– If we model each of the denominators as
– $Z \sim Normal(X, \sigma_x)$
– Sample directly from the implied sampling distribution for the population estimate.
– Obtain a distribution for the ASMR for each geography/age combination
– Using these Monte Carlo estimates, standard life tables are estimated

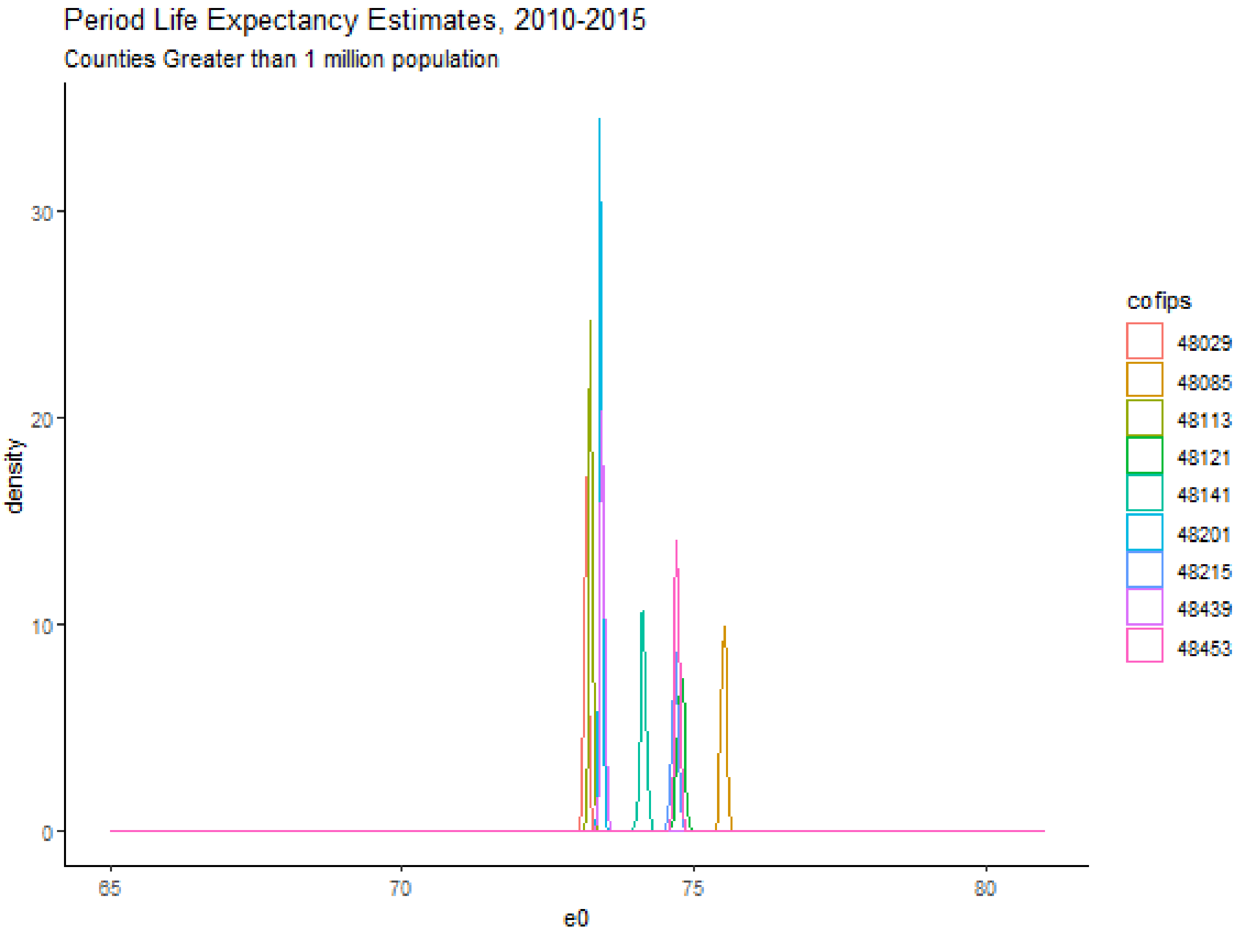
Data

- 5 years of individual death certificates from Texas Department of State Health Services (IRB 18-045)
  - Geocoded to counties for the current analysis, aggregated by age (10 year groups)
  - ~ 840,000 deaths in the period
- Future analysis will also be done at ZIP code and Census tract levels.

Some results - Life Expectancy estimates



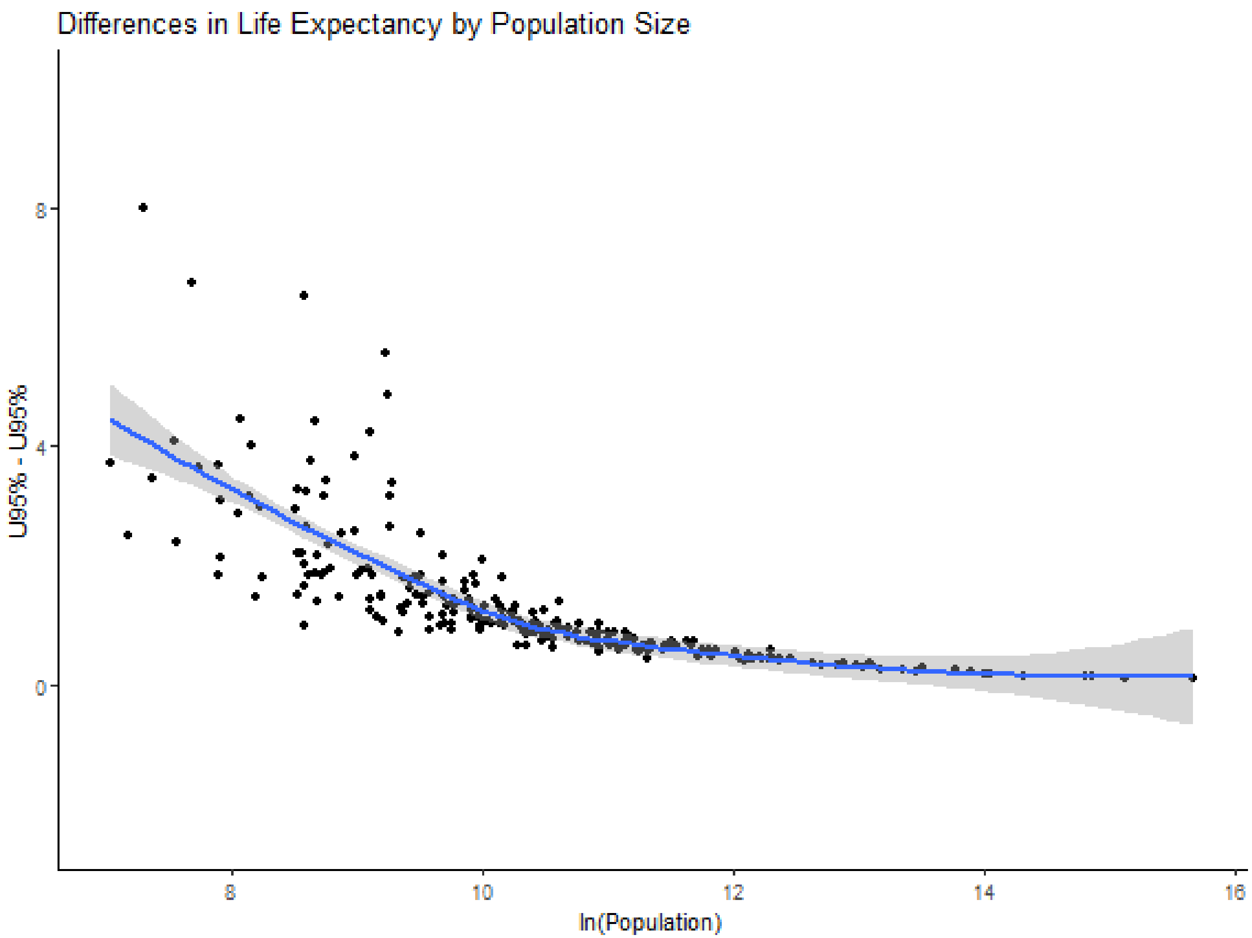
Some results - Life Expectancy estimates



Variance by size

- <5,000 - 14.6 years on average
- 5,000 to 10,000 - 7.1 years on average
- 10,000 to 50,000 - 4.6 years on average
- 50,000 to 100,000 - 4.3 years on average
- 100,000+ - 4.3 years on average

Some results - errors and population size



Comparison to other estimates

– <5,000 - 14.6 years on average
– IHME 3.0 years on average
– 5,000 to 10,000 - 7.1 years on average
– IHME 5.8 years on average
– 10,000 to 50,000 - 4.6 years on average
– IHME 3.3 years on average
– 50,000 to 100,000 - 4.3 years on average
– IHME 1.7 years on average
– 100,000+ - 4.3 years on average
– IHME .8 years on average

Thank you!

– Corey Sparks, Ph.D.
– UTSA Department of Demography
– corey.sparks@utsa.edu |
– @Coreysparks1