

SynUSA Synthetic Health Insurance Analytic Files

Developed by Steve Parente, PhD

Copyright, 2018

The SynUSA Synthetic Health Insurance Analytic File System (SHIAF) was created with the goal of providing a realistic set of claims data for all insured populations in the United States. The purpose of SynUSA is a training database for students of health insurance claims-based analytics.

At present SynUSA is a Beta test database being made available to educational institutions. The database is available in CSV and SAS formats. It is a large database with two years of data (2014-2015) with linked person specific information.

Other features include:

- Five major insurance types covering almost all insured in the US including:
 - Employer Sponsored Insurance (ESI)
 - Medicaid (under 65)
 - Medicare Fee for Service (including dual eligible over 65)
 - Medicare Advantage (including dual eligible over 65)
 - Non-Group: Both ACA exchange and non-exchange
- Person-level statistical weights to generate national count statistics
- State and 4 census region variables
- Major age categories
- Linkage to the ARCOLA and MARCOLA health insurance demand microsimulation model populations (See: Parente, S.T., Feldman, R., Spetz, J. et al. "Wage Growth for the Health Care Workforce: Projecting the Affordable Care Act Impact. Health Services Research. 2016 May 3. doi: 10.1111/1475-6773.12497)

Downloadable Files

- [Data Schema / Dictionary \(via Proc Contents\)](#)
- SynUSA 2014-15 Membership/Beneficiary Files SAS & CSV Files (Zipped)
- SynUSA 2014 Claims SAS File (Zipped)
- SynUSA 2014 Claims CSV File (Zipped)
- SynUSA 2015 Claims SAS File (Zipped)
- SynUSA 2015 Claims CSV File (Zipped)