OBJECTIVE: Pharmacy datasets are useful for evaluating drug costs, patient utilization, and patient adherence. Few have information on sociodemographic variables including race, education, income, and urban/rural designation of the patients’ neighborhoods. We undertook to estimate those variables by linking USA census 2000 data with one such database. METHODS: We obtained census data tables for race, median income, education, and percentage of urban/rural residences by Zip Improvement Plan (ZIP) codes. Linking to the RxAmerica pharmacy claims database by 5-digit ZIP, we estimated sociodemographic variables for a cohort of adult new users of lipid-lowering therapy with ≥18-months continuous eligibility. For patients without a ZIP, pharmacy ZIP was used as a proxy. Four variables were used to generate race, median income, education level, and urban/rural designation of the patients’ neighborhoods.

RESULTS: A total of 29,667 patients met the inclusion criteria. Of these, 28,293 (95.4%) had a valid 5-digit ZIP that linked with one in the census tables; 97.0% of the ZIPs were defined from the patients’ addresses. Among a sample of 19,458 patients, there were 4662 unique ZIPs; number of patients within each ZIP ranged from 1–125 (SD 11.9). The mean observation for income based on the median income variable reported in the census table was $45,924 (SD $15,965). The median observation for education was at least a high-school diploma; 25% of subjects had an observation corresponding to at least some college. Most of the patients had observations for race indicating they were from predominantly white neighborhoods (63.5% from neighborhoods that were ≥75% white). Most of the patients also had observations indicating they were from highly urbanized areas (74.0% from neighborhoods that were ≥75% urban). CONCLUSION: Census tables may be useful for estimating sociodemographic variables for pharmacy claims analyses. Future work should focus on the validity of these variables for estimating sociodemographic variables.