

International Journal of Population Data Science

Journal Website: www.ijpds.org



Generating Cohort-specific Earnings through the Secure Query System

Amy O'Hara¹, Stephanie Straus¹, and Jake Pasner¹

¹Georgetown University, Washington DC, USA

Objectives

In collaboration with the U.S. Internal Revenue Service, we are developing a Secure Query System (SQS) to produce earnings and tax statistics for lists of individuals supplied by government agencies or academic institutions. Without the SQS solution, these organizations cannot observe outcomes for their program participants or learners.

challenges, with additional products in development to measure earnings outcomes for randomized controlled trials and businesses.

Methods

The SQS provides a pre-defined table of aggregate statistics for the individuals provided by the SQS clients who match to tax records. The SQS solution has four main components: (1) A data validation tool to ensure client data align with IRS schema; (2) Person-level linkages using Splink; (3) Tabulation of statistics for client file matches to either wage records or individual income tax returns; and (4) An automated disclosure avoidance review that suppresses and coarsens percentages and dollar values.

Results

The SQS minimum viable product includes a user-friendly tool to check the completeness and value sets of input variables, ensuring that files pulled by IRS are optimized for matching. The Splink matching module allows for a variety of identifiers to be used alone or in combination to obtain accurate matches among the hundreds of millions of tax returns at IRS. Blocking strategies help achieve efficient matching. Returns for matching taxpayers are processed, for instance combining wages across multiple employers or setting flags for credits claimed. All processing is done by IRS employees on IRS systems (Georgetown acts as intermediary but does not view or process any microdata). Results for each taxpayers are aggregated back to the cohort level to produce the output statistics.

Conclusion

The SQS solution is designed to reduce burden on IRS staff while meeting the measurement needs of state, local, and non-profit organizations. It is versatile solution to measurement

