International Journal of Population Data Science

Journal Website: www.ijpds.org

Reconciling Parent-Child Relationships across US Administrative Datasets

O'Hara, A¹, Genadek, K², Medalia, C², and Alexander, T³

¹Stanford University

²US Census Bureau

³ICPSR, University of Michigan

Introduction

Population data capture children, parents, relatives, and others moving in and out of households. The U.S. has seen falling marriage rates, and increases in multigenerational households and complex families, young children living with grandparents, and adult children living with parents. Robust parent-child linkages are critical to understand these demographic shifts.

Objectives and Approach

We construct and validate parent-child linkages over a century to observe how U.S. households are changing over time. The three largest person-based datafiles in the U.S. are the decennial censuses, the Social Security Administration transaction file, and individual tax returns from the Internal Revenue Service. These sources operationalize relationships differently, capture data at various frequencies, and gather the data for unique purposes. We use probabilistic matching to observe and reconcile parent-child relationships across these sources. The data include a variety of personal identifiers including name, date of birth, parents' names, address, and place of birth that support matching and validation.

Results

We find that understanding the content, consistency, and coverage of the files before matching is critical for high quality linkages. The representativeness of the parent-child relationship file improves over time, with the weakest coverage for the Greatest Generation and the strongest coverage for Millennials. Coverage varies by source: tax data underrepresent non-white children and have duplicate records for SSNs, while names and dates of birth are missing from Census data. Multiple match rates differ among demographic groups and over time. In the matching process, the blocking variables rely on common variables across the population datasets. Our approach provides robust entity resolution for women, despite married-maiden name changes. We describe challenges due to data problems in old census records and validation changes in social security data.

Conclusion/Implications

We conduct a successful reconciliation of parent-child relationships in U.S. population level files. The project supports operational and research uses, such as the 2020 Census. We will extend this work using graph matching and will expand the method to validate other relationship links including spouses and siblings.