

Linking the 1980 Enumeration Sample to the 1980 Census

Kelsey J. Drotning and Katie R. Genadek

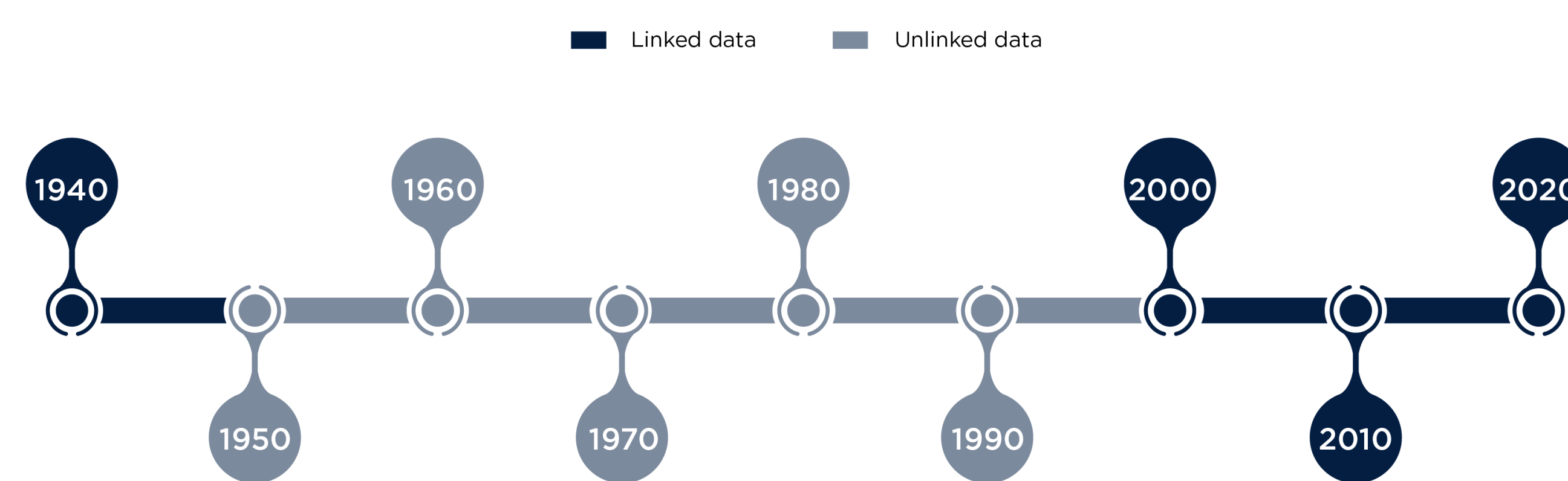
Project Goals

1. Describe the Decennial Census Digitization and Linkage project
2. Document the 1980 Enumeration Sample (E-Sample)
3. Use two record linkage methods to link the 1980 E-Sample to the 1980 Decennial Census

1. The Decennial Census and Digitization Linkage Project (DCDL)

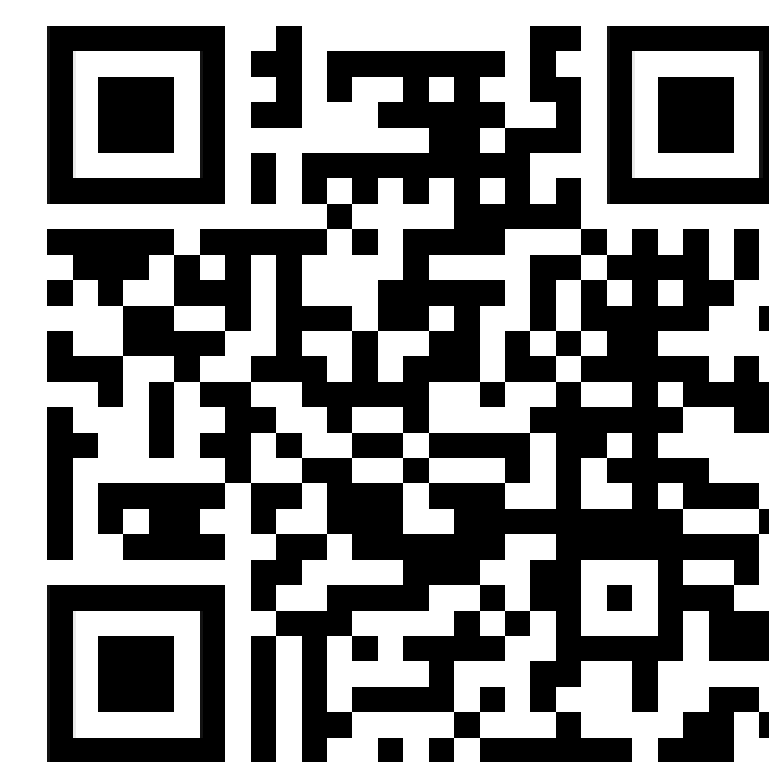
DCDL is an initiative to produce linked restricted microdata files from the 1960-1990 censuses. Currently, the 1940 Census can only be linked to the 2000, 2010, and 2020 censuses. This initiative, in addition to current efforts to digitize and link the 1950 Census, will complete the U.S. census longitudinal infrastructure.

Figure 1. Linked Decennial Census Data



Project Phases:

1. Digitally scan the microfilm containing the original census images from 1960 – 1990 censuses.
2. Recover handwritten respondent names from the digital images using handwriting recognition technology
3. **Attach respondent names to otherwise complete microdata files held at the Census Bureau**
4. Assign Protected Identification Keys (PIKs) to the data so that individuals can be linked across files following the removal of PII
5. Disseminate linked data for Census Bureau use and to researchers in a restricted environment



Follow our progress on the DCDL Website.



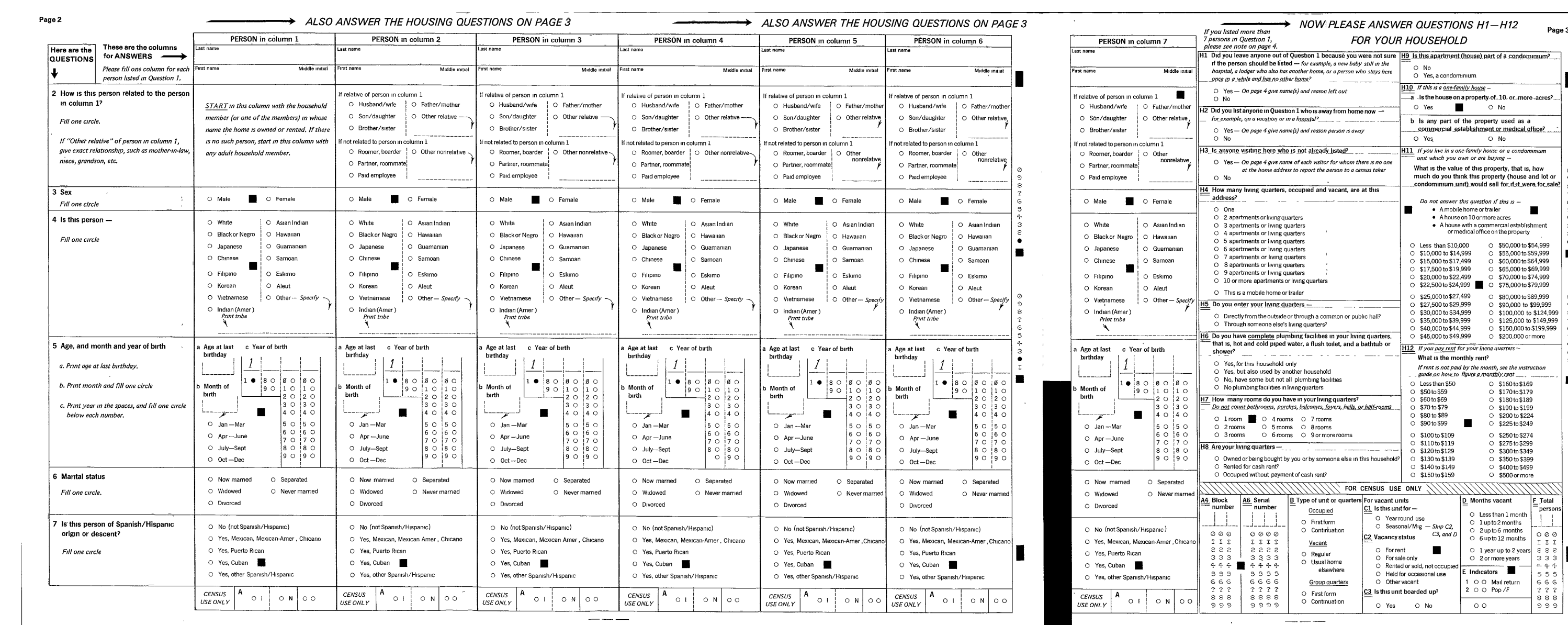
2. The 1980 Enumeration Sample

Purpose: to estimate erroneous enumeration in the Census, such as when a person or household is counted at the wrong address or counted more than once.

Sample: **110,000 housing units** drawn from the 1980 Census in clusters of 10 housing units per enumeration district. Eligible enumeration districts were restricted to Primary Sampling Units (PSUs) in the 1980 April Current Population Survey.

Method: Census staff are sent to interview all current residents in the selected housing unit about where they lived on census day and if there were any other locations where they might have been recorded. If a household member was not home, information was sought from present household members or the post office.

Figure 2. 1980 Census Short Form



3. Record Linkage

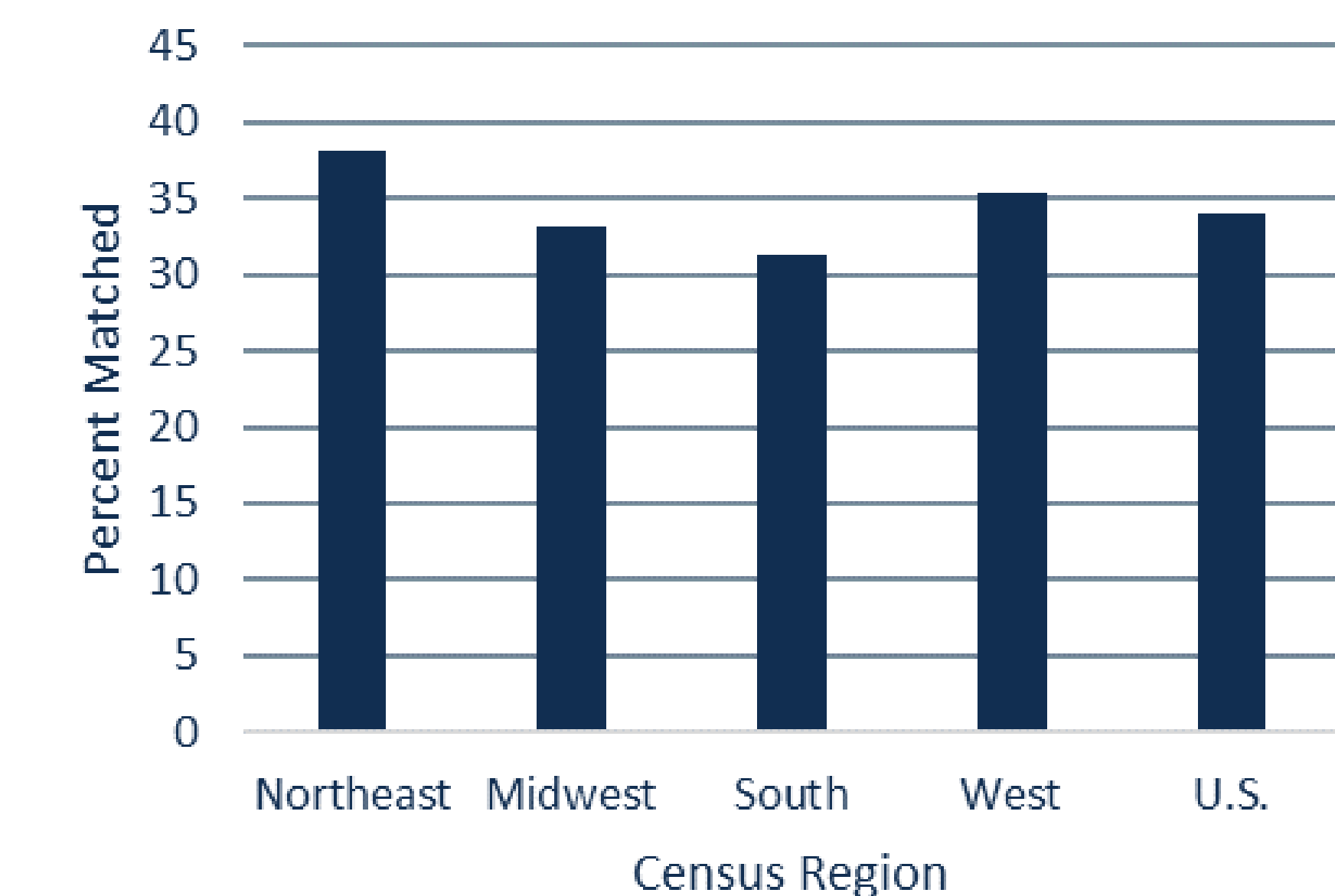
We are interested in linking residents within housing units (person records) included in the E-Sample (which has names) with their corresponding person records in the 1980 Decennial Census (which does not have names). While the E-Sample was drawn from the 1980 Census, unique identifiers were not retained so record linkage is not straight forward. We used **deterministic record linkage**:

- Deterministic record linkage relies on either a linkage key or a set of identifier variables common to both datasets to evaluate whether two records match.
- We linked records using exact matching on both geographic variables (state FIPS, district office, tract, enumeration district) and demographic variables (sex, race, age, marital status, relationship to household head).

Results

About **34 percent** of person records from the E-Sample were matched to the Census using the deterministic record linkage algorithm.

Figure 3. Percentage of 1980 E-Sample Person Records Matched to the 1980 Decennial Census by Linkage Method



Lessons Learned

This work contributes to Phase 3 of the DCDL project. Linking the E-Sample names to Census microdata that is missing names provides a test case for developing scalable record linkage algorithms that will be used to link newly captured names to existing microdata.

- Record linkage of persons (residents) within households using historical data is challenging due to the lack of string variables (i.e., first name, last name) available for use in matching.
- Deterministic record linkage in this case was relatively ineffective and additional methods should be tried.

Next Steps

- Next we will try using **probabilistic record linkage**. This method uses window blocking and fuzzy matching to predict the likelihood two records match.
- We will use exact matching on both geographic variables (state FIPS, district office, tract, enumeration district) and demographic variables (sex, race, marital status, relationship to household head). We will also use +1/-1 window blocking for age and concatenate the age variable across the household for fuzzy matching.

References:

1. Cowan and Bettin (1982)
2. 1980 Census Population and Housing History – Chapter 9

