

Linking the 1980 Post-Enumeration Program Samples to the 1980 Decennial Census

Kelsey J. Drotning and Katie R. Genadek
2023 FCSM Conference



Any conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau (Data Management System number: P- 7522572, Disclosure Review Board (DRB) approval number: CBDRB-FY23-ADEP001-006.

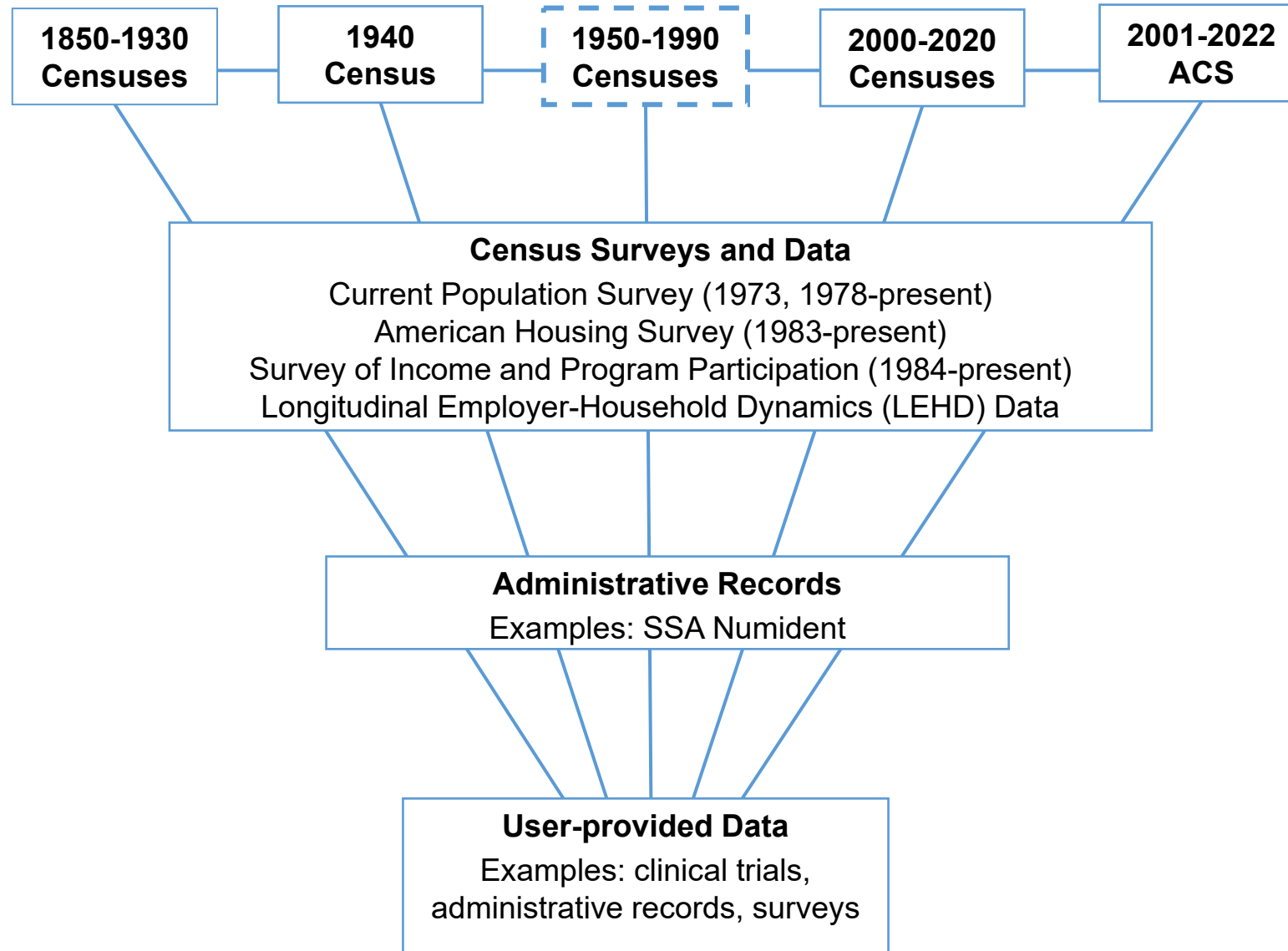


How the Census Bureau Links Records

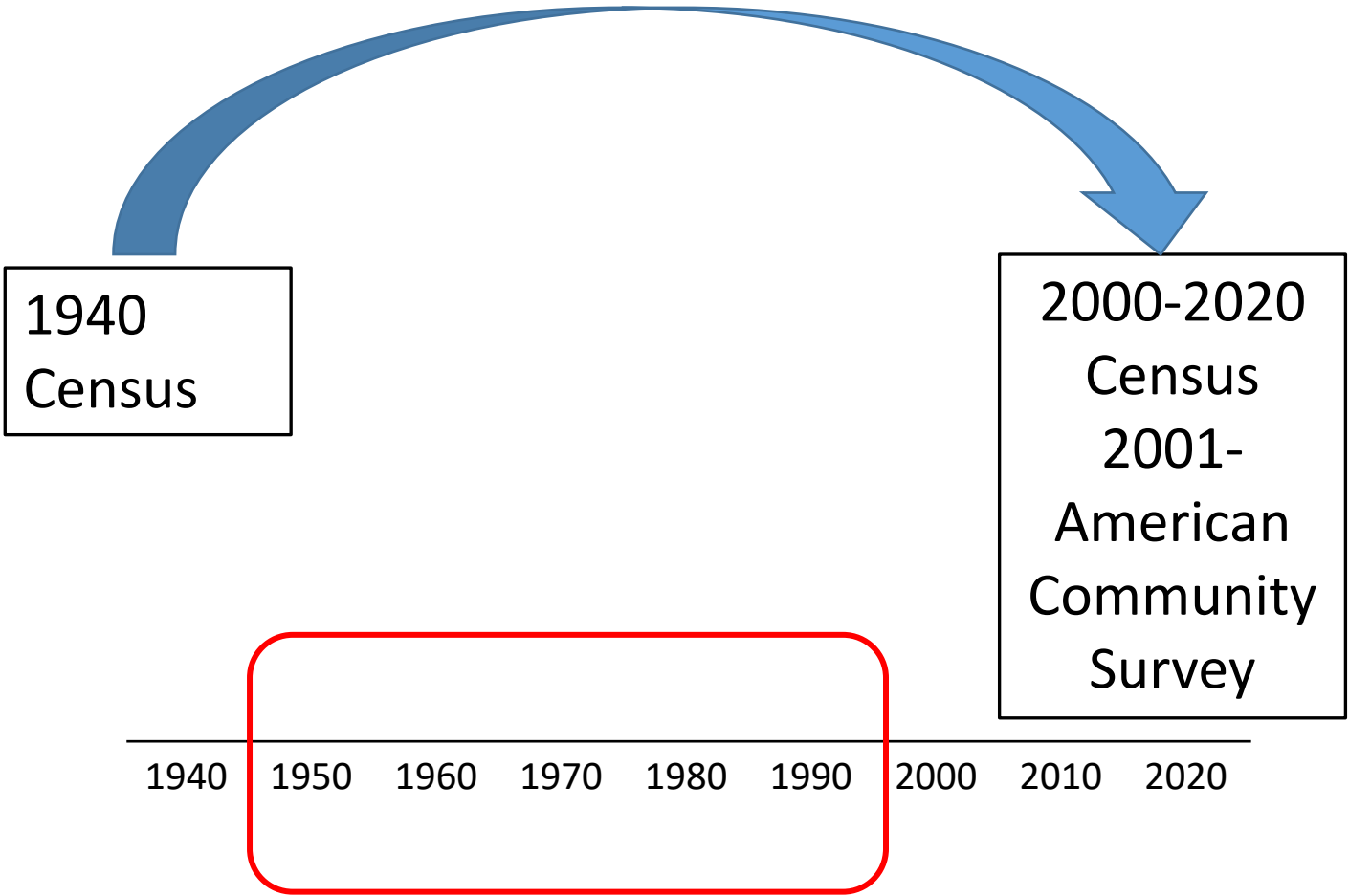
All cases linked to a common “composite file” of administrative records

- Most common linkage keys: name, date of birth, sex, place of residence
- Uses standard techniques of data cleaning and probabilistic matching
- Matched cases are assigned a Protected Identification Key (PIK)
- Data assigned PIKs can be linked to any other data with PIKs

Data Linkage Infrastructure



Existing linked Decennial Census of Populations



Why is this difficult?

- Microdata files for 1960-1990 exist but do not include names
- Names were handwritten on census forms
- Census forms are stored on 250,000 microfilm reels
- Forms and data are highly restricted to protect respondent confidentiality

The Decennial Census Digitization and Linkage (DCDL) Project

- Initiative to produce linked microdata files from the 1960 – 1990 Censuses
- 1950 Census is currently being digitized and will be linked into infrastructure

Project Phases

1. Digitization: Scanning the microfilm
2. Capture: Recovering handwritten respondent names from scans
3. Merge: Attaching names to otherwise complete microdata files
4. Linkage: Assigning Protected Identification Keys (PIKs) to the data
5. Dissemination: making linked data available to researchers in a restricted environment

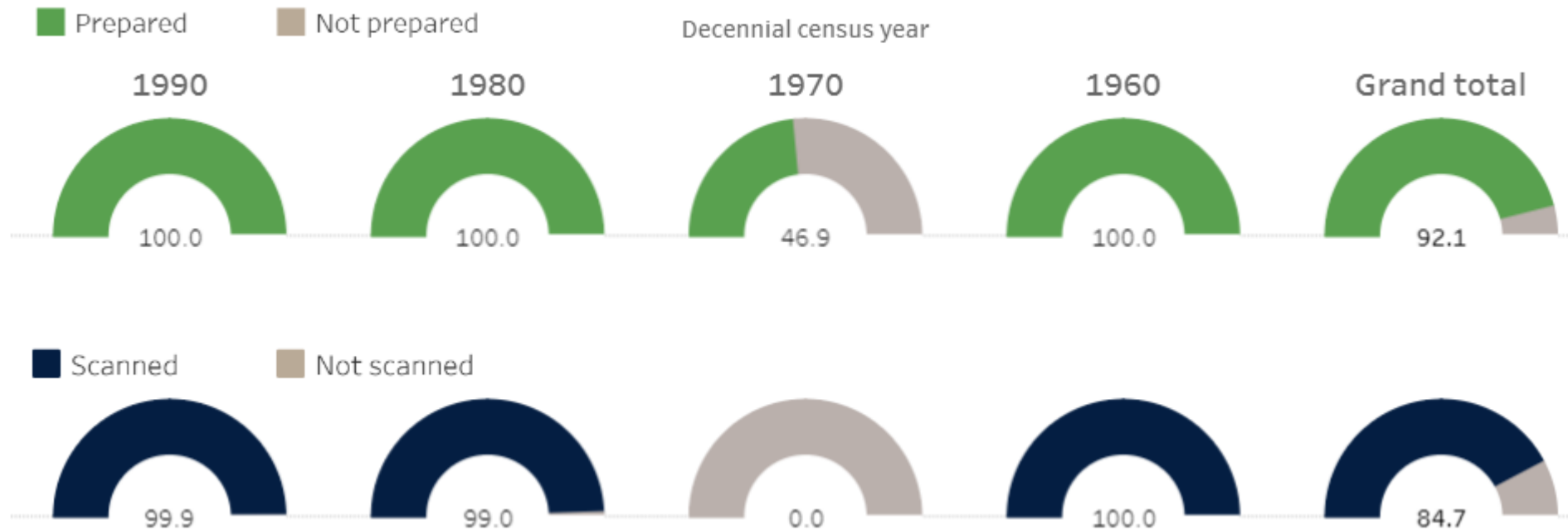
Digitization: Microfilm Scanning Ongoing



Digitization: Project Status

- Follow our progress on the DCDL website: www.census.gov/dcdl

Percent Completion of Reel Preparation and Scanning



Capture: Optical Character Recognition (OCR) and Optical Mark Recognition (OMR)

- Partnering with a team from Information Sciences Institute at University of Southern California
- Performed best on 1990 Census Pilot in 2015
- 2020 Census Bureau Data Capture team working on custom OMR
- Production OCR and OMR to begin in 2024

Linkage: PIK Assignment

- Assign PIKs by adapting standard Census Bureau linkage methods
- Make use for 1969-1989 IRS 1040 full universe files
- Harmonized addresses in IRS data to Census geography
- Recovering 1970, 1980 Census address lists
- Adapt linkage strategies on samples of census data with names

The 1980 Enumeration Sample (E-Sample)

- **Original Purpose:** To estimate erroneous enumeration in the census
- **Sample:** 110,000 housing units from 1980 Census
- **Method:** Enumerator interviews
- **Capture:** Keyed in Name respondents

Here are the QUESTIONS ↓	These are the columns for ANSWERS → Please fill one column for each person listed in Question 1.	PERSON in column 1	PERSON in column 2
		Last name First name Middle initial	Last name First name Middle initial
<p>2 How is this person related to the person in column 1?</p> <p>Fill one circle.</p> <p>If "Other relative" of person in column 1, give exact relationship, such as mother-in-law, niece, grandson, etc.</p>	<p><i>START in this column with the household member (or one of the members) in whose name the home is owned or rented. If there is no such person, start in this column with any adult household member.</i></p>	<p>If relative of person in column 1</p> <p><input type="radio"/> Husband/wife <input type="radio"/> Father/mother</p> <p><input type="radio"/> Son/daughter <input type="radio"/> Other relative</p> <p><input type="radio"/> Brother/sister</p> <p>If not related to person in column 1</p> <p><input type="radio"/> Roomer, boarder <input type="radio"/> Other nonrelative</p> <p><input type="radio"/> Partner, roommate</p> <p><input type="radio"/> Paid employee</p>	
<p>3 Sex</p> <p>Fill one circle</p>	<p><input type="radio"/> Male <input checked="" type="checkbox"/> <input type="radio"/> Female</p>	<p><input type="radio"/> Male <input checked="" type="checkbox"/> <input type="radio"/> Female</p>	
<p>4 Is this person —</p> <p>Fill one circle</p>	<p><input type="radio"/> White <input type="radio"/> Asian Indian</p> <p><input type="radio"/> Black or Negro <input type="radio"/> Hawaiian</p> <p><input type="radio"/> Japanese <input type="radio"/> Guamanian</p> <p><input type="radio"/> Chinese <input checked="" type="checkbox"/> <input type="radio"/> Samoan</p> <p><input type="radio"/> Filipino <input type="radio"/> Eskimo</p> <p><input type="radio"/> Korean <input type="radio"/> Aleut</p> <p><input type="radio"/> Vietnamese <input type="radio"/> Other — Specify</p> <p><input type="radio"/> Indian (Amer) <i>Print tribe</i></p>	<p><input type="radio"/> White <input type="radio"/> Asian Indian</p> <p><input type="radio"/> Black or Negro <input type="radio"/> Hawaiian</p> <p><input type="radio"/> Japanese <input type="radio"/> Guamanian</p> <p><input type="radio"/> Chinese <input checked="" type="checkbox"/> <input type="radio"/> Samoan</p> <p><input type="radio"/> Filipino <input type="radio"/> Eskimo</p> <p><input type="radio"/> Korean <input type="radio"/> Aleut</p> <p><input type="radio"/> Vietnamese <input type="radio"/> Other — Specify</p> <p><input type="radio"/> Indian (Amer) <i>Print tribe</i></p>	
<p>5 Age, and month and year of birth</p> <p>a. Print age at last birthday.</p> <p>b. Print month and fill one circle</p> <p>c. Print year in the spaces, and fill one circle below each number.</p>	<p>a Age at last birthday c Year of birth</p> <p>1</p> <p>b Month of birth</p> <p>1 ● 8 ○ 9 ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○</p> <p>5 ○ 6 ○ 7 ○ 8 ○ 9 ○</p> <p><input type="radio"/> Jan—Mar <input checked="" type="checkbox"/> <input type="radio"/> 5 ○ 6 ○ 7 ○ 8 ○ 9 ○</p> <p><input type="radio"/> Apr—June</p> <p><input type="radio"/> July—Sept</p> <p><input type="radio"/> Oct—Dec</p>	<p>a Age at last birthday c Year of birth</p> <p>1</p> <p>b Month of birth</p> <p>1 ● 8 ○ 9 ○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○</p> <p>5 ○ 6 ○ 7 ○ 8 ○ 9 ○</p> <p><input type="radio"/> Jan—Mar <input checked="" type="checkbox"/> <input type="radio"/> 5 ○ 6 ○ 7 ○ 8 ○ 9 ○</p> <p><input type="radio"/> Apr—June</p> <p><input type="radio"/> July—Sept</p> <p><input type="radio"/> Oct—Dec</p>	
<p>6 Marital status</p> <p>Fill one circle.</p>	<p><input type="radio"/> Now married <input type="radio"/> Separated</p> <p><input type="radio"/> Widowed <input type="radio"/> Never married</p> <p><input type="radio"/> Divorced</p>	<p><input type="radio"/> Now married <input type="radio"/> Separated</p> <p><input type="radio"/> Widowed <input type="radio"/> Never married</p> <p><input type="radio"/> Divorced</p>	
<p>7 Is this person of Spanish/Hispanic origin or descent?</p> <p>Fill one circle</p>	<p><input type="radio"/> No (not Spanish/Hispanic)</p> <p><input type="radio"/> Yes, Mexican, Mexican-Amer, Chicano</p> <p><input type="radio"/> Yes, Puerto Rican</p> <p><input type="radio"/> Yes, Cuban <input checked="" type="checkbox"/> <input type="radio"/> Yes, other Spanish/Hispanic</p>	<p><input type="radio"/> No (not Spanish/Hispanic)</p> <p><input type="radio"/> Yes, Mexican, Mexican Amer, Chicano</p> <p><input type="radio"/> Yes, Puerto Rican</p> <p><input type="radio"/> Yes, Cuban <input checked="" type="checkbox"/> <input type="radio"/> Yes, other Spanish/Hispanic</p>	
	<p>CENSUS USE ONLY A ○ I ○ N ○ ○</p>	<p>CENSUS USE ONLY A ○ I ○ N ○ ○</p>	

1980 Census Sample for Linkage

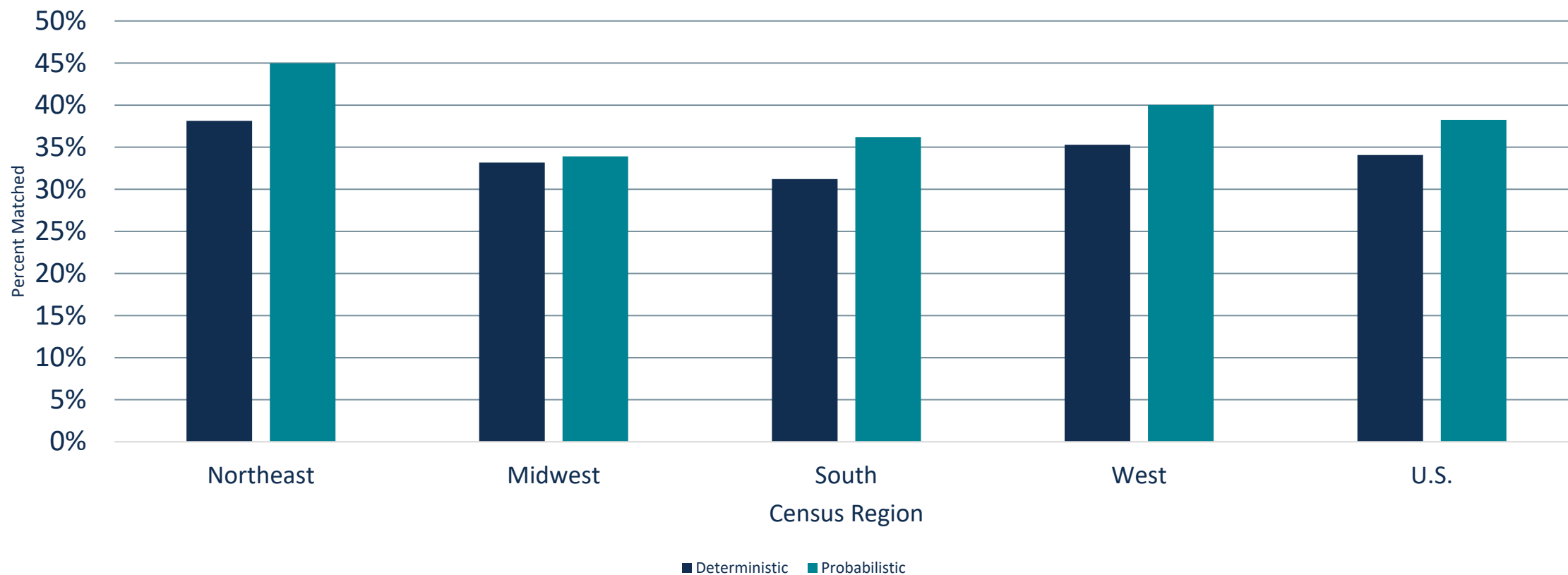
- Link the E-Sample to 1980 Census data
- Attach Names from E-Sample to 1980 Census record
- Use the 1980 Census records with names for development of PIK assignment strategies

Record Linkage

- Linking based on household and personal characteristics
 - Variables: geography, sex, race, age, marital status, relationship to household head
- Deterministic record linkage: Exact matching
- Probabilistic record linkage: Window blocking on age (+/- 1)

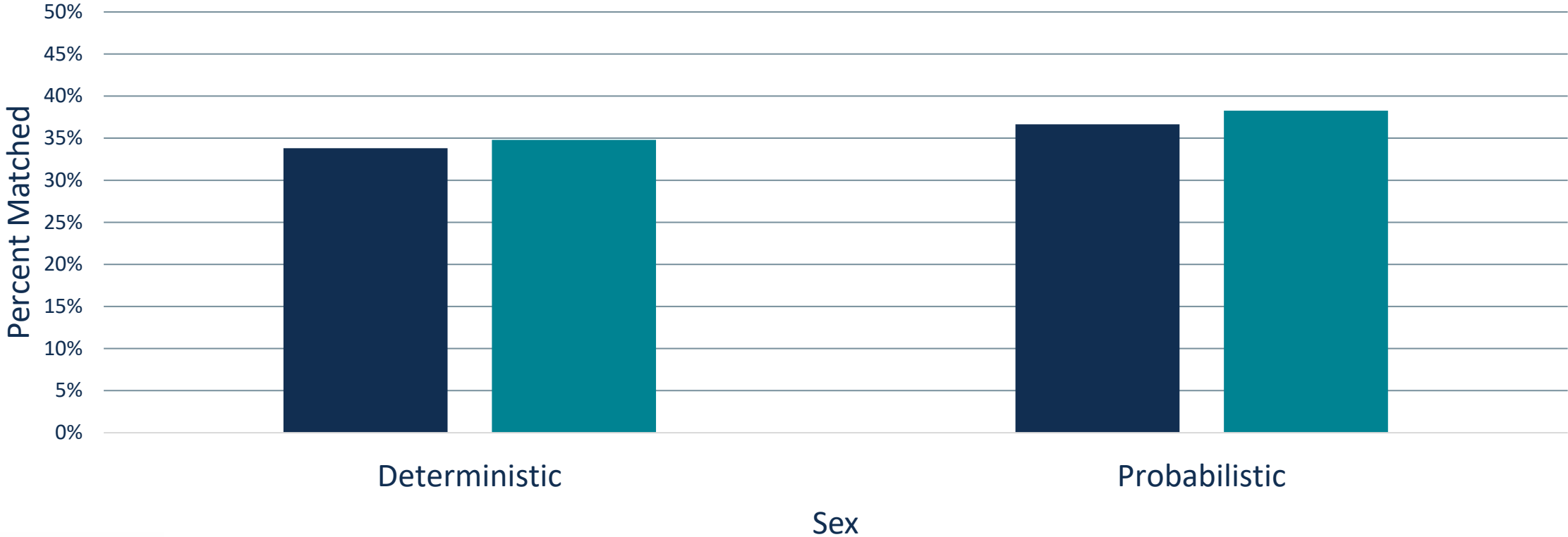
Results - Region

Percent of E-Sample Records Matched to the 1980 Census by Region



Results - Sex

Percent of E-Sample Records Matched to the 1980 Census by Sex



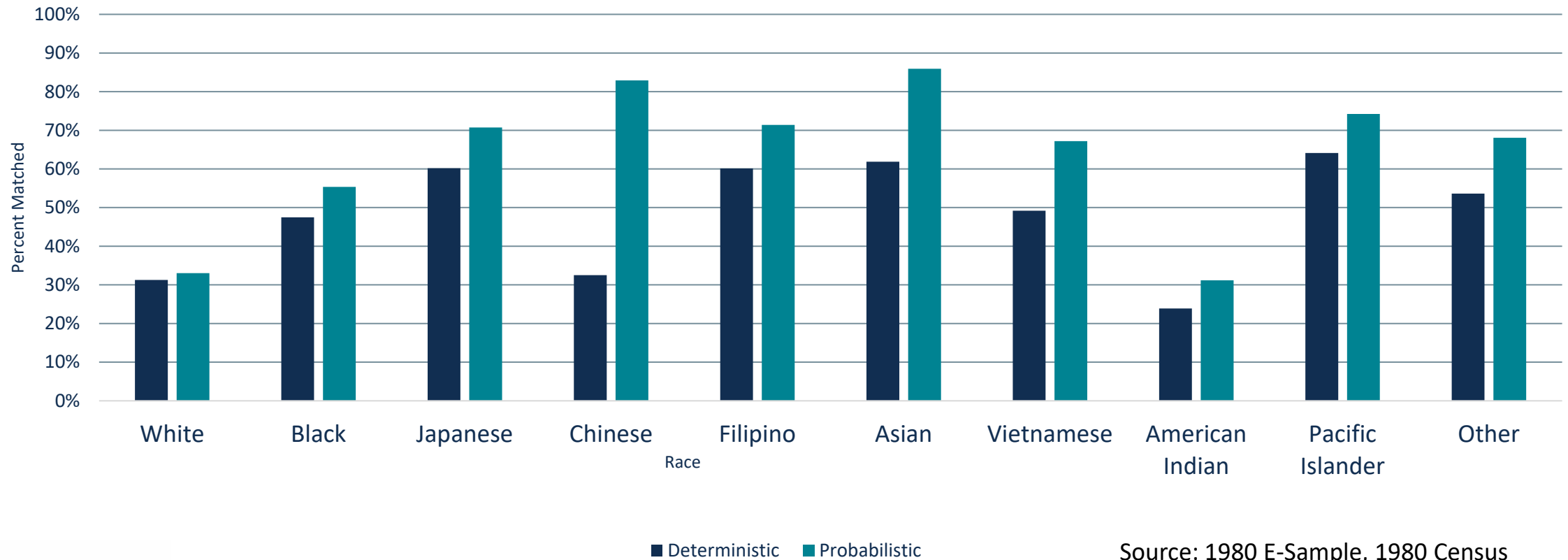
■ Male ■ Female

Source: 1980 E-Sample, 1980 Census

All results were approved for release by the Disclose Review Board of the U.S. Census Bureau (Data Management System number: P- 7522572, Disclosure Review Board (DRB) approval number: CBDRB-FY23-ADEP001-006.

Results - Race

Percent of E-Sample Records Matched to the 1980 Census by Race



Source: 1980 E-Sample, 1980 Census

Conclusions

- Linked 38% of E-Sample to 1980 Census data
- Record linkage using historical data is challenging due to the lack of unique identifying variables
- Probabilistic record linkage performs slightly better than deterministic record linkage, but there are still a lot of records we are unable to link
- 1980 Census sample with names ready for PIK assignment

DCDL Team

Aaron Glover
Abigail Hiller
Adam Galemore
Alessandro Ferrucci
Allison D Fox
Amanda Pait
Ana Cherry
Andrew Hennesy
Andrew Larson
Angeline K Brock
Anna Simpson
Aretha Wilkerson
Arrington Smith
Babatunde Adewuyi
Barbara J Maupin
Becky J Francis
Benjamin Peterson
Benny Alvey
Betty J Jones
Bijou Kisula
Bill Peil
Bittany Dixon
Brenda Figueroa
Cam Dang
Carla Medalia
Carlos Bacerra
Carlos Mena
Carrie A Ridge
Catherine L Beauchamp
Cathy Sharp
Cecile Murray
Chad Adkins
Charles Andrew Stahl
Cheryl Grim
Cheryl Smeltekop
Christian Moscardi
Christine Mosely
Christine Parkhill
Christophe Saliboko
Christopher Baker
Christopher Ison
Christy E Stewart
Christy S Scoggins

Cindy Johnson
Claids Shacklette
Claudia Shacklette
Connie Hong
Consolee Mulindahabi
Corey Silvia
Cory Reed |
Damon M Lewis
Dana Rowland
Darcie Oubezza
Darlene Poisson
David Bleckley
David Grusky
David Middleton
Dean Michalak
Debbie Johnson
Deborah Kay Dietz
Debra A Pound
Denise Henderson
Denise L Nein
Diana Magnuson
Don Badrack
Erick Wingfield
Erin Meyer
Felix Hayes
Felix Ramirez
Francesca Marini
Fred Krenke
Freddie Branham
Gina Erickson
Girsie D Lee
Halla Joumaa-Jouney
Helen Keeton
Holly Neilson
Huaigu Cao
Ian Dodd
Jackie Alexander
James Christy
James Richardson
Jan Dickenson
Jarred M Roberts
Jason Hall
Jeff Behler

Jeffrey A Evans
Jeffrey Monroe
Jennifer L Violette
Jennifer Seitz
Jerry Collins
Jim Csoka
Joey Morales
John Abowd
John Elting
John Friedman
John Haskill
John Sullivan
John T Smith
John Trambaugh
Jon Morgan
Jonathan Cropper
Jonathan Fisher
Judy M Brading
Jumanne Young
Kanese Gardner
Karen McGuffin
Karen Walker
Keith Finlay
Kelsey Drotning
Kent Patterson
Kevin Deardorff
Krista Chan
Lara Karacasu
Laveda Hall
Lesley Wallis
Linda G Robinson
Linda Vernooy
Lisa Carroll
Lisa Kelly
London Johnson
Lucia Foster
Lukas Althoff
Marcus Whitesides
Margaret Levenstein
Marguerite Molner
Mark Weng
Martia L Riley
Marvin Credle

Mary A O Bryan
Mary Buchanan
Mary Holbrook
Mary O'Rourke
Mary Reynolds
Matthew Lenfert
Melissa Banzhaf
Michael Baker
Michael K Hunter
Michael Winggerd
Michele K Middleton
Michelle Karlsson
Michelle Snow
Mindi Hanna
Morgan Miller
Nathaniel Kly
Nevada Basdeao
Nhut K Dao
Nick Orsini
Nicole Lane
Norm Johnson
Norma B Ettlting
Patricia De la Cruz
Paul Friday
Paula Moore
Peggy Robinson
Phyllis M Smitson
Prem Natarajan
Raj Chetty
Rana W Alwazani
Randy Becker
Rhonda L Tobey
Richard Flagg
Robert Karlen
Robin Lawrence
Robin Roberts
Rodney Gaddis
Rolanda R Scherer
Ron Seymour
Roy Ashley
Ryann Brown
Sabdi Taylor
Sallie Keller

Sam Jones
Samantha Gognat
Sandra E Rice
Sandra L Post
Sara Roseman
Sarah Herman
Sarah Oppenhiemer
Shanna M Benson
Shannon Felton Spence
Shannon G Koenig
Sharon Moitoso
Sherry L Nichols
Silvia N Nava
Sohial Kamdar
Stephanie Studds
Sumit Khaneja
Susan Heitz-Carpenter
Tahir Zeb
Tammy Ann Shelton
Tammy Samuels
Tara Jones
Taylor Hanson
Teresa Holmes Scott
Terrah A Bruner
Terri Kennerly
Terri Sheckles
Terrie L Henderson
Theresa Toole
Thomas Sherman
Timothy Morris
Todd Gardner
Tonya L Roberts
Toshica Harvey
Tracy Coulter
Trent Alexander
Troy Erickson
Tynell Johnson
Vincent Biller
Wendy R Montgomery
Winona Miniard
Xujun Peng
Yukari Nakata