Appendix Materials for "Children's Indirect Exposure to the U.S. Justice System: Evidence from Longitudinal Links between Survey and Administrative Data"

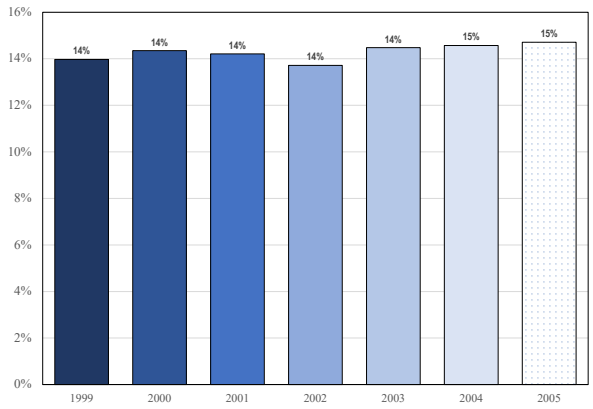
Keith Finlay, Michael Mueller-Smith, and Brittany Street

Table of Contents

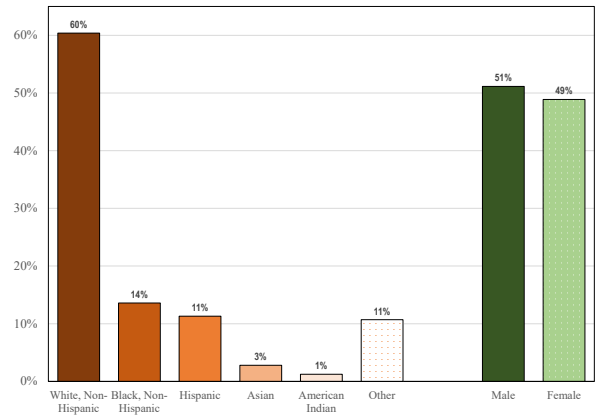
A Supplementary results

Figure A1: Composition of sample

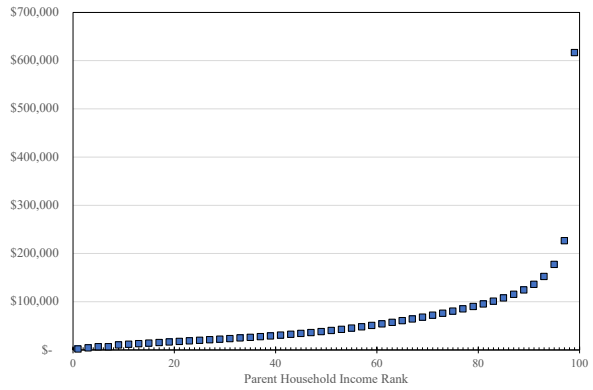
A: Birth year



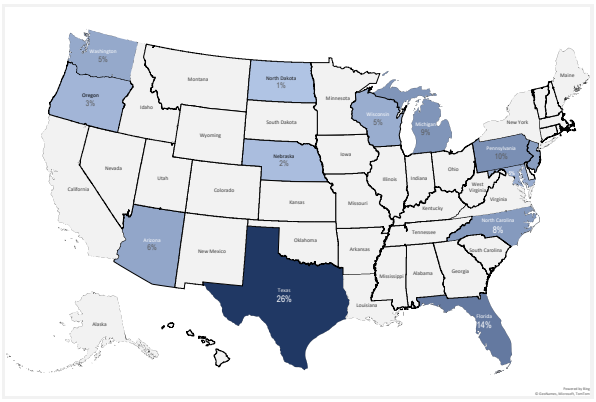
B: Child's race, ethnicity, and sex



C: Average household AGI by percentile rank



D: Place of birth

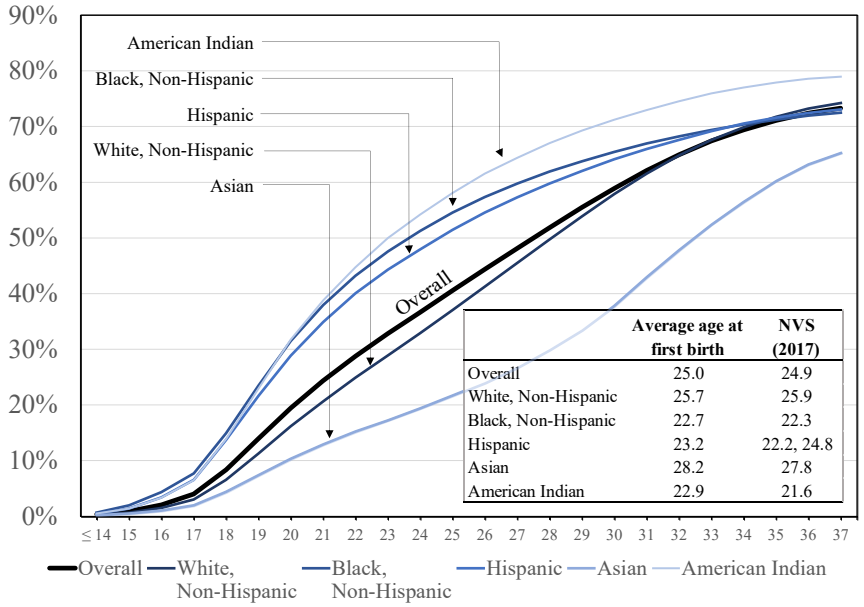


Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident born between 1999 and 2005 in CJARS-covered states. Child's year of birth, place of birth, and sex are measured using the Census Numident. Child's race and ethnicity is measured using the Census BestRace files. Average Adjusted Gross Income (AGI) is measured using the IRS Form 1040s on which the child is claimed in their year of birth and the subsequent four years. AGI is reported as zero if the child is not claimed. Children that are not claimed in their first five years of life or are ever claimed on a form reporting negative AGI are dropped from the sample in Panels C and D. Panels C and D depict average AGI, number of caregiver links observed, and the number of tax filings within the first five years for children within percentile bins as rank-ordered within birth cohorts. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-009.

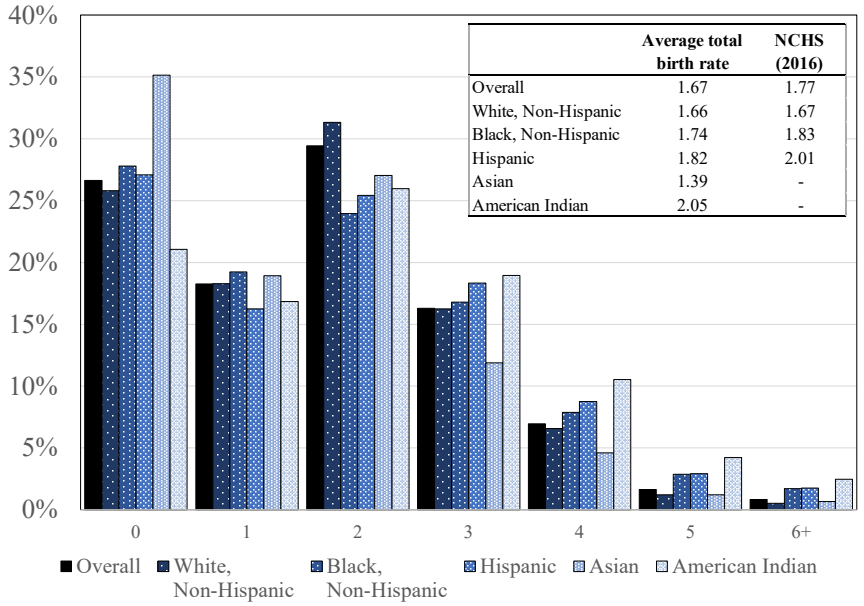
Figure A2: Life cycle fertility for females born in 1981, by race and ethnicity

A: Share with observed birth by age



	Average age at first birth	NVS (2017)
Overall	25.0	24.9
White, Non-Hispanic	25.7	25.9
Black, Non-Hispanic	22.7	22.3
Hispanic	23.2	22.2, 24.8
Asian	28.2	27.8
American Indian	22.9	21.6

B: Number of births



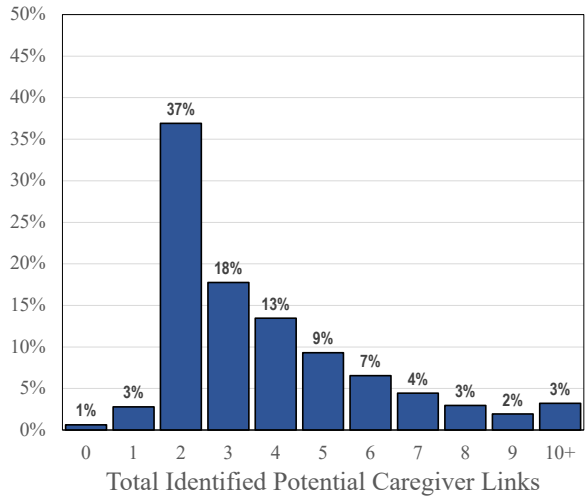
	Average total birth rate	NCHS (2016)
Overall	1.67	1.77
White, Non-Hispanic	1.66	1.67
Black, Non-Hispanic	1.74	1.83
Hispanic	1.82	2.01
Asian	1.39	-
American Indian	2.05	-

Source: Calculations are based on the Census Numident, the Census BestRace files, and the CJARS relationship crosswalk.

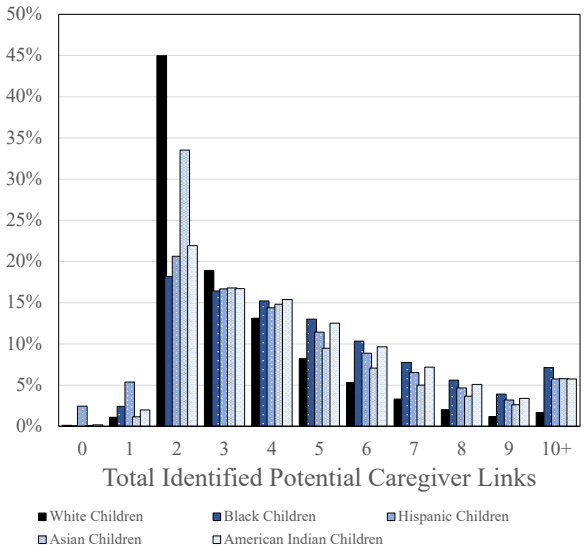
Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of females in the Census Numident born in 1981 in any state. Race and ethnicity are measured using the Census BestRace files. Individuals are linked to the CJARS family crosswalk to measure fertility and age of first birth, defined as identifying a biological child and determining age at birth based on the year of birth of the child. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-009. These estimates can be compared to publicly available national statistics on fertility. For Panel A, the National Center for Health Statistics reports the following ages at first birth in 2000: overall 24.9, White 25.9, Black 22.3, Mexican 22.2, Central/South American 24.8, Asian 27.8, and American Indian/Alaska Native 21.6 (Mathews and Hamilton, 2016). For Panel B, the 2017 National Vital Statistics reports total birth rates: overall 1.77, White 1.67, Black 1.83, and Hispanic 2.01 (Mathews and Hamilton, 2019).

Figure A3: Distribution of total potential caregiver links identified per child

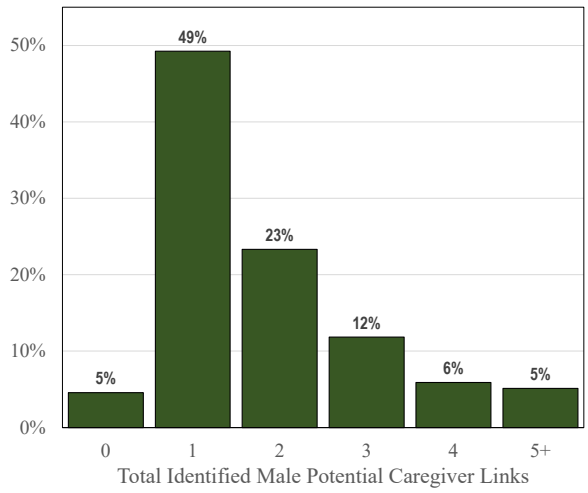
A: All children, all potential caregivers



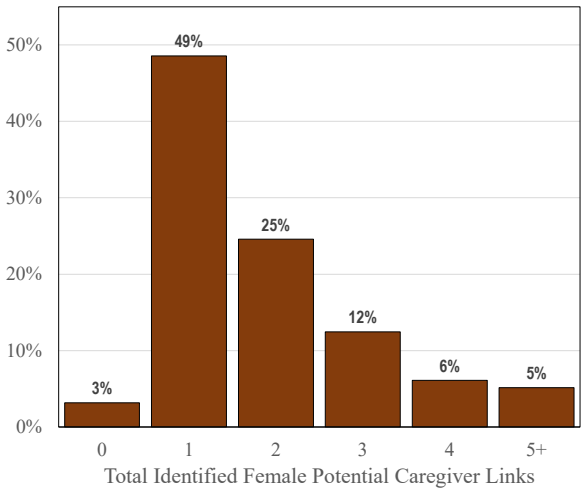
B: By child's race/ethnicity, all potential caregivers



C: All children, male potential caregivers



D: All children, female potential caregivers

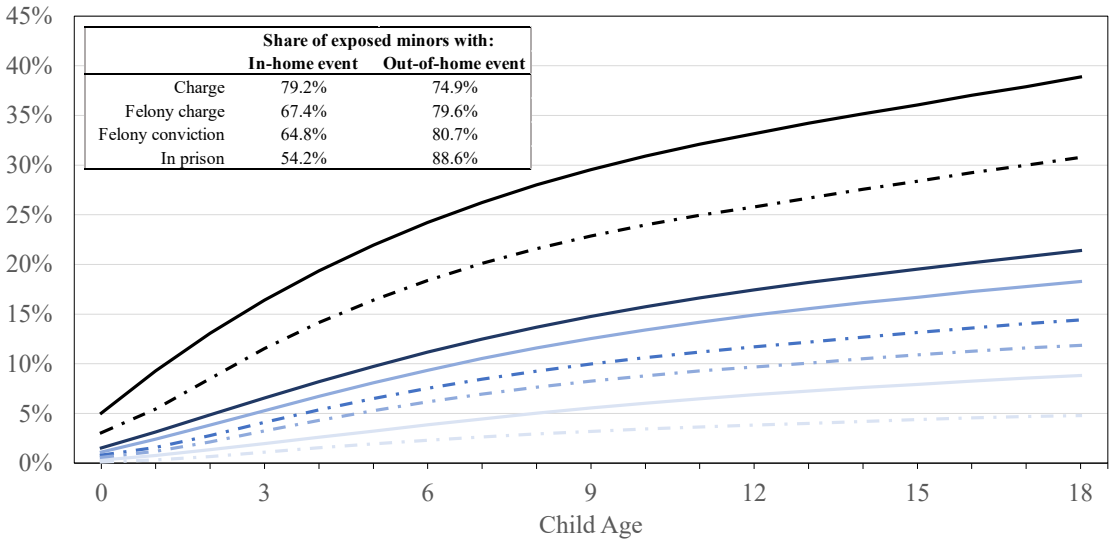


Source: Calculations are based on the Census Numident, the Census BestRace files, and the CJARS relationship crosswalk.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident born between 1999 and 2005 in all states. Child's race and ethnicity are measured using the Census BestRace files. Potential caregivers are defined as biological parents, stepparents, adopted parents, foster parents, unclassified caregivers, grandparents, aunts/uncles, non-familial adults (cohabiting 2+ years), and unclassified adults (cohabiting 2+ years). All results were approved for release by the U.S. Census Bureau, authorization numbers CBDRB-FY22-ERD002-001 and CBDRB-FY22-ERD002-003.

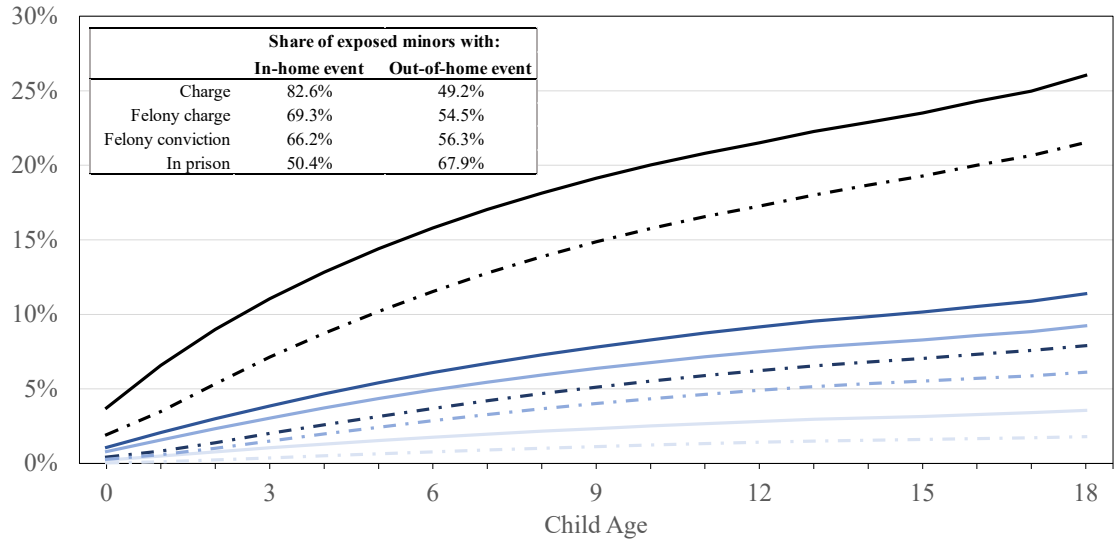
Figure A4: Cumulative exposure to the criminal justice system, by current or recent coresidence of adult

A: All potential caregivers

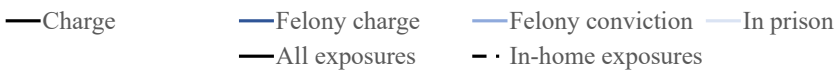


	Share of exposed minors with:	
	In-home event	Out-of-home event
Charge	79.2%	74.9%
Felony charge	67.4%	79.6%
Felony conviction	64.8%	80.7%
In prison	54.2%	88.6%

B: All biological parents



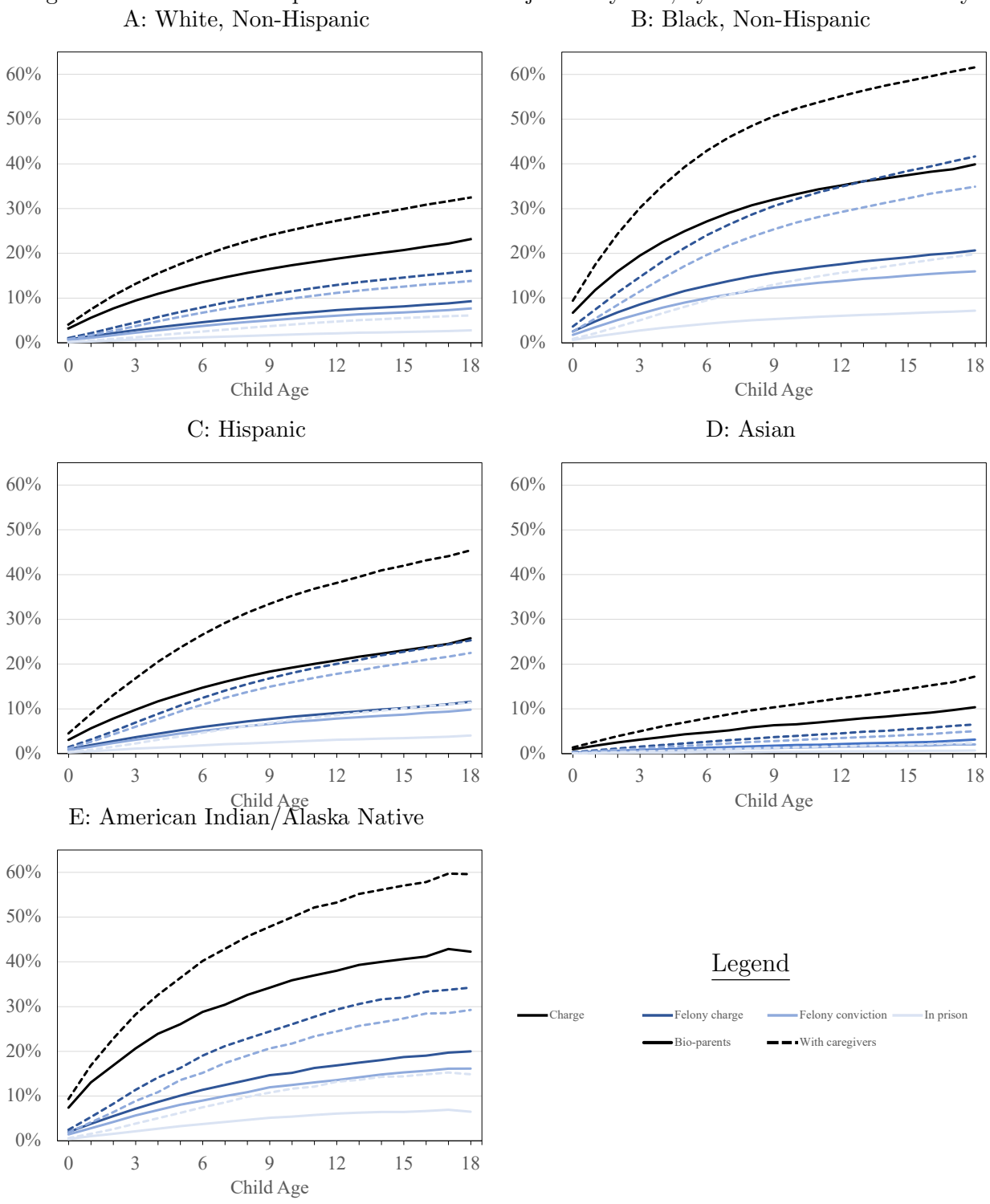
	Share of exposed minors with:	
	In-home event	Out-of-home event
Charge	82.6%	49.2%
Felony charge	69.3%	54.5%
Felony conviction	66.2%	56.3%
In prison	50.4%	67.9%



Source: Calculations are based on the Census Numident, CJARS, and CJARS residence and relations crosswalks.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2005 birth cohorts in CJARS-covered geographies from birth until X, where X represents years since birth (0–18) with the place of birth still covered or year 2018. In-home exposure is defined as exposure by an individual who was coresiding with the child in the year of the event or in the preceding two years. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval numbers CBDRB-FY22-ERD002-001 and CBDRB-FY22-ERD002-003.

Figure A5: Cumulative exposure to the criminal justice system, by child's race and ethnicity

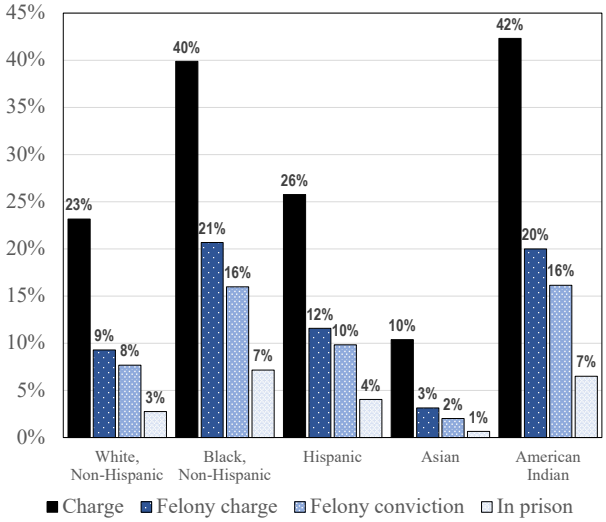


Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

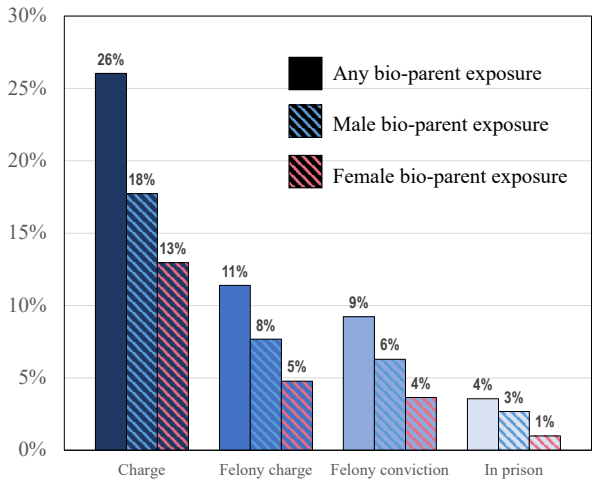
Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2005 birth cohorts in CJARS-covered geographies from birth until X, where X represents years since birth (0–18) with the place of birth still covered or year 2018. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-001.

Figure A6: Heterogeneous cumulative exposure by biological parents

A: By child's race and ethnicity



B: By biological parents' sex

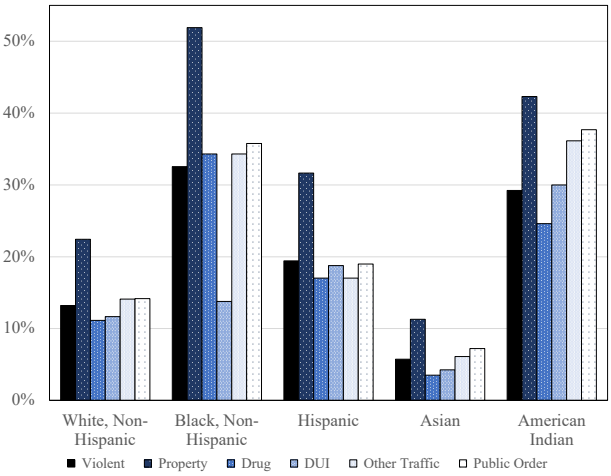


Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

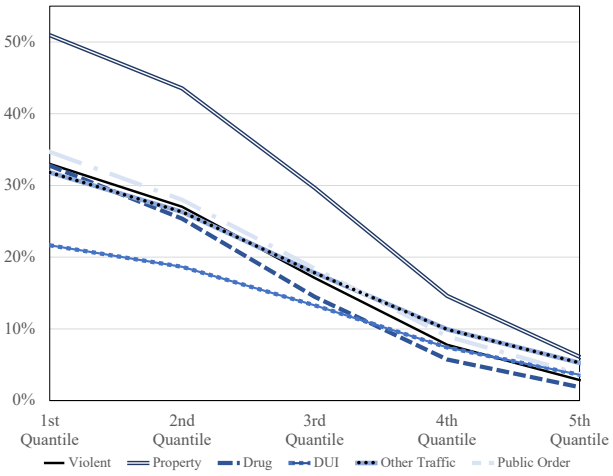
Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2005 birth cohorts in CJARS-covered geographies from birth until X, where X represents years since birth (0–18) with the place of birth still covered or year 2018. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-001.

Figure A7: Heterogeneous cumulative exposure to criminal charges by offense type

A: By child's race and ethnicity



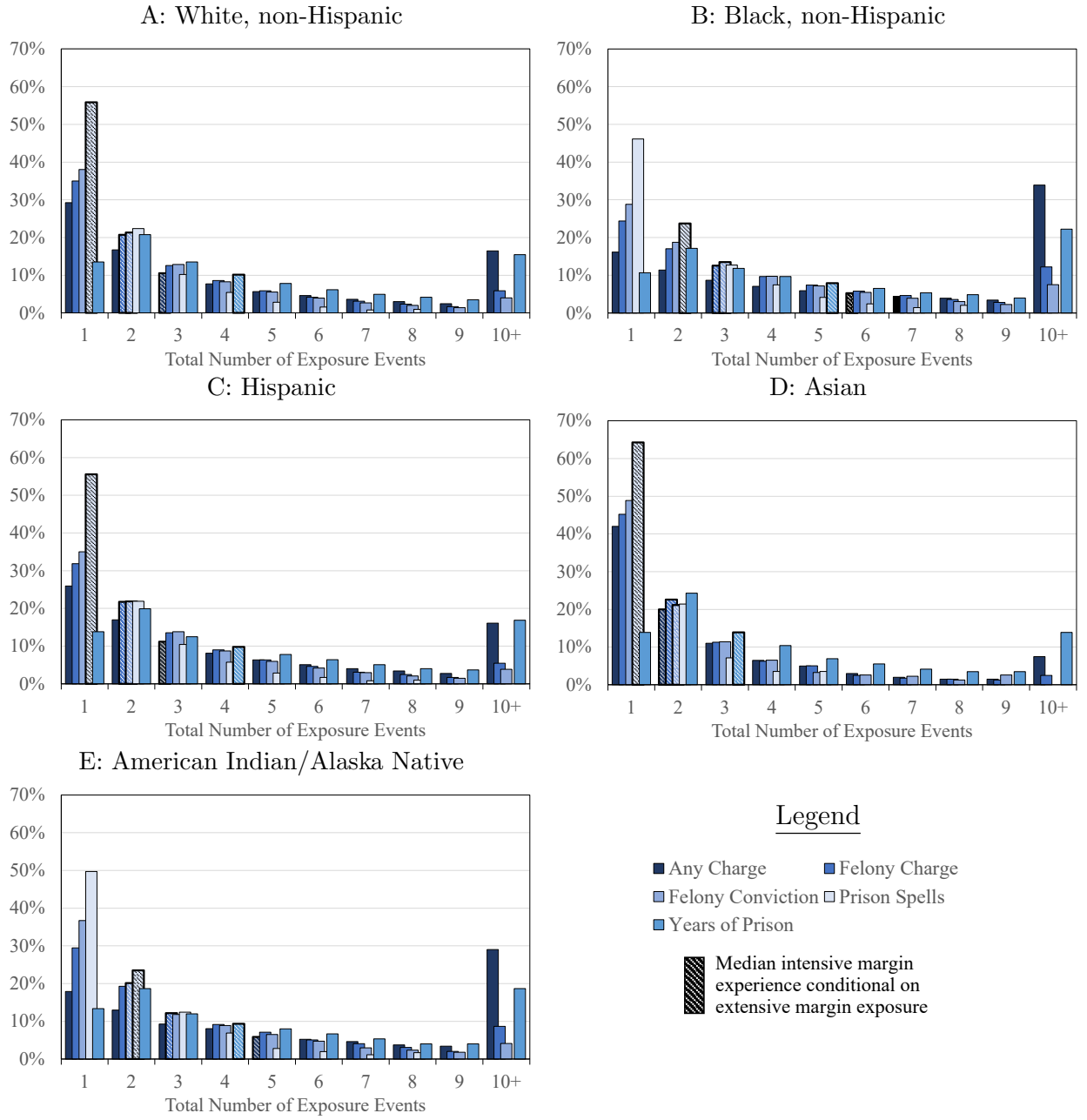
B: By parents' income rank



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, the CJARS relations and residency crosswalks, and IRS Form 1040s.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2000 birth cohorts in CJARS-covered geographies from birth until age 18. Average exposure by age 18 to specific charge types is depicted by child race/ethnicity (Panel A) and by household income quantiles (Panel B). Income quantiles are determined using the average adjusted gross income reported on IRS Form 1040s, in which the child is claimed for the first five years. Children claimed on a form with negative AGI or never claimed in the first five years are not included in the sample. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-009.

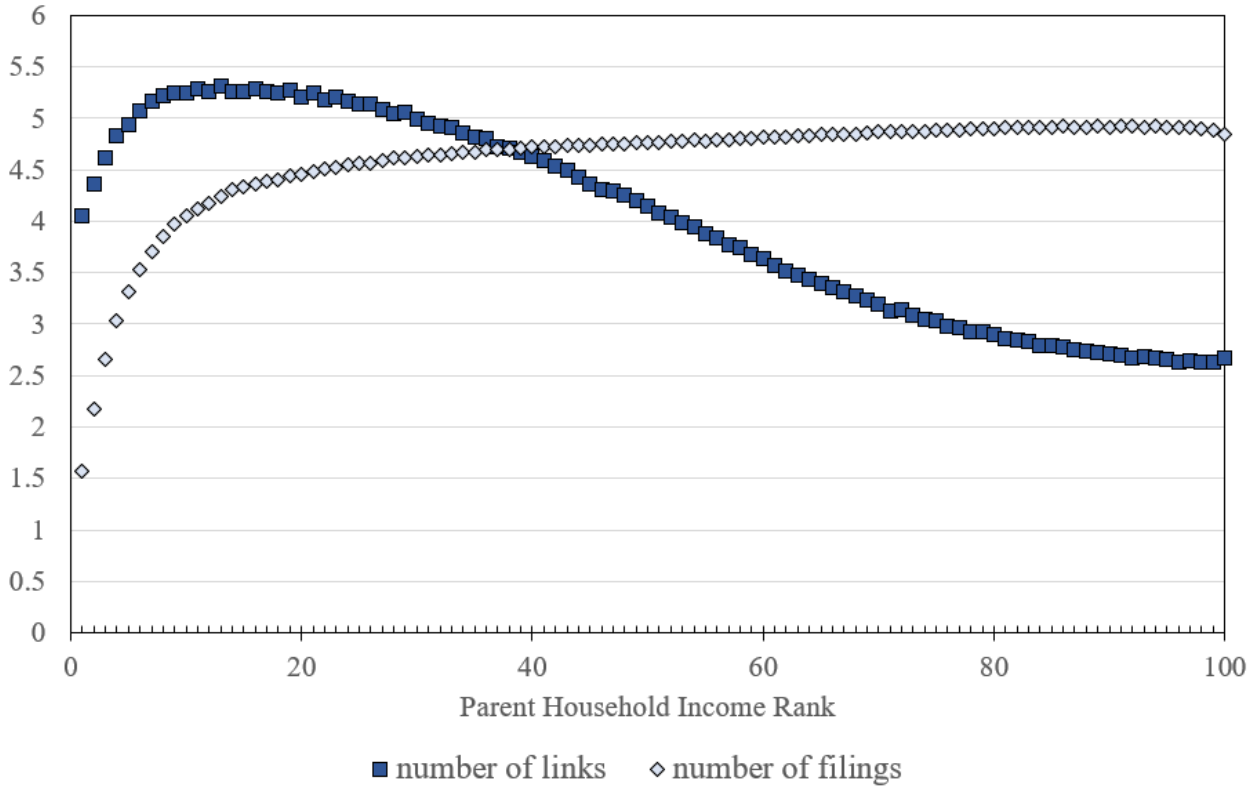
Figure A8: Intensive margin of criminal justice exposure by parents and other potential caregivers by age 18, by child's race and ethnicity



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2000 birth cohorts in CJARS-covered geographies from birth until age 18. Distinct events are counted among children with any exposure. Thus, multiple charges filed on the same date are considered one event, and similarly for the other types of criminal justice events. The number of events is truncated at 10+ events for all events *except* prison spells, which are top coded at 8+. Asian felony convictions and prison spells are top coded at 9+ and 5+ to preserve confidentiality, respectively. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval numbers CBDRB-FY22-ERD002-003 and CBDRB-FY22-ERD002-009.

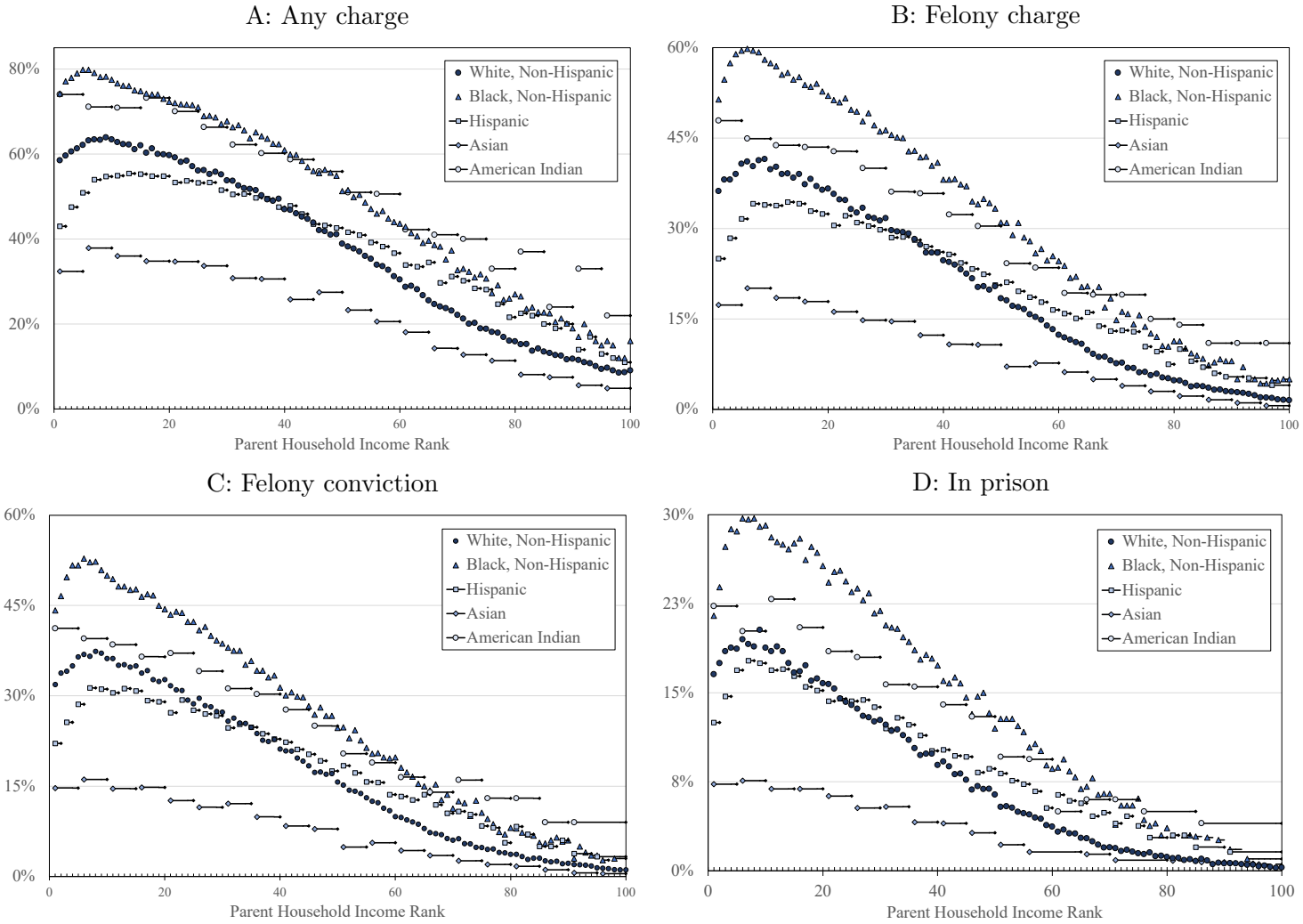
Figure A9: Average number of caregiver links and tax filings, by parent income rank



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, the CJARS relationship crosswalk, and IRS Form 1040s.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident born in 1999 and 2000 in CJARS-covered geographies from birth until age 18. Income percentile bins are determined using the average adjusted gross income reported on IRS Form 1040s, in which the child is claimed for the first five years. Children claimed on a form with negative AGI or never claimed in the first five years are not included in the sample. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY23-013.

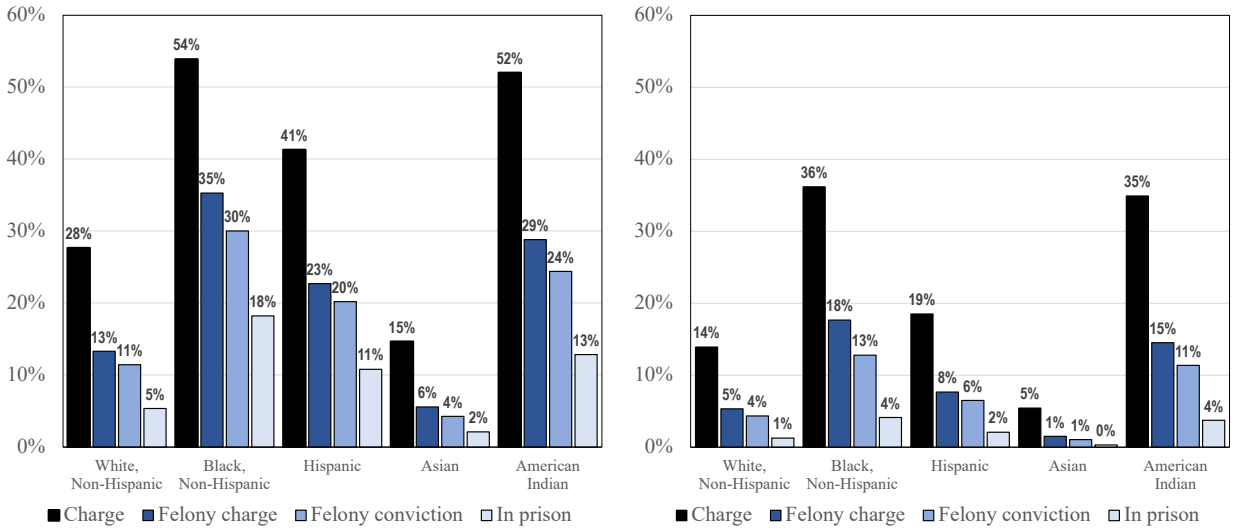
Figure A10: Exposure heterogeneity, by exposure event, parent income rank, and child's race and ethnicity



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, the CJARS relationship crosswalk, and IRS Form 1040s.
Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident born in 1999 and 2000 in CJARS-covered geographies from birth until age 18. Average exposure by age 18 is depicted for children across income percentile bins. Some bins, marked with horizontal black lines, are wider to satisfy disclosure requirements. Income percentile bins are determined using the average adjusted gross income reported on IRS Form 1040s, in which the child is claimed for the first five years. Children claimed on a form with negative AGI or never claimed in the first five years are not included in the sample. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-003.

Figure A11: Heterogeneous cumulative exposure, by child race and ethnicity and adult sex

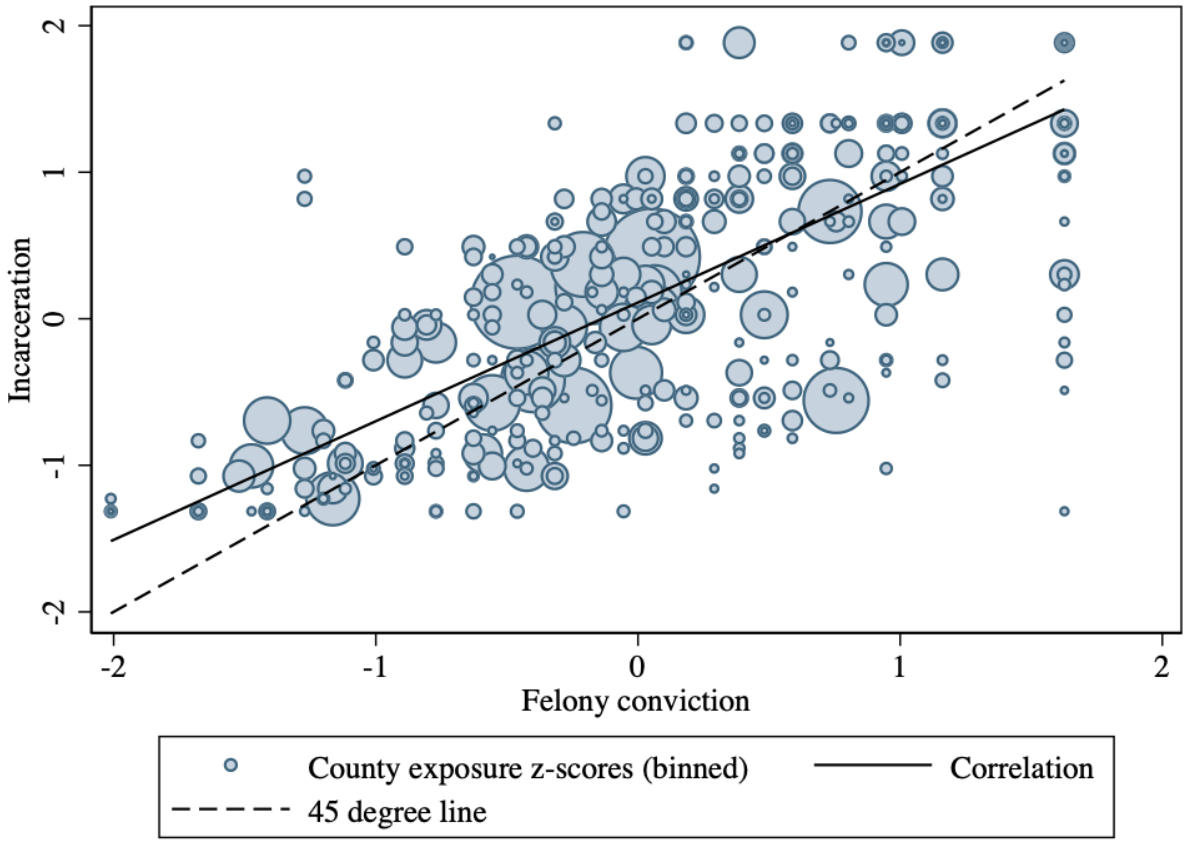
A: Male potential caregivers B: Female potential caregivers



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2005 birth cohorts in CJARS-covered geographies from birth until X, where X represents years since birth (0–18) with the place of birth still covered or year 2018. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY22-ERD002-001.

Figure A12: Relationship between county-level felony conviction and incarceration exposure z-scores

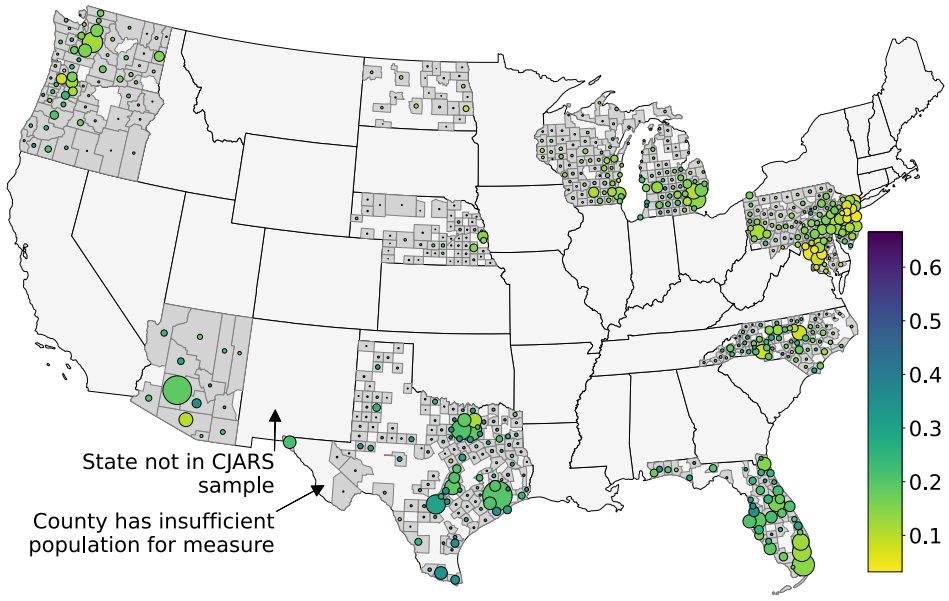


Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relationship crosswalk.

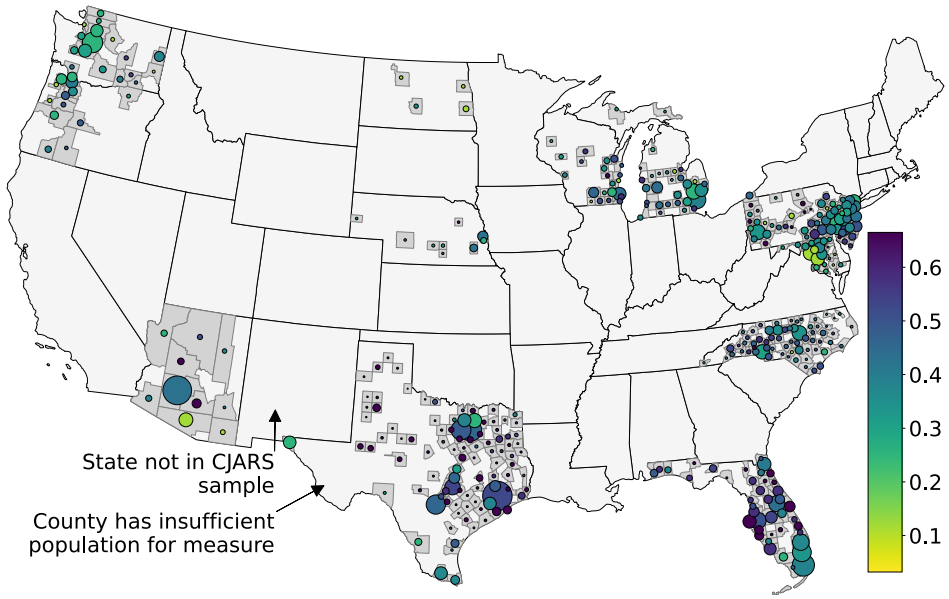
Notes: Estimates have been binned and rounded to preserve confidentiality. Plot points are sized according to 2021 Census Bureau county total population estimates. The sample consists of individuals in the Census Numident 1999–2005 birth cohorts in CJARS dual-covered geographies (courts and prison coverage) from birth until X, where X represents years since birth (0–18) with the place of birth still covered or year 2018. A regression of incarceration z-scores from disclosed exposure rates by county on felony conviction z-scores shows a correlation of 0.81 and R^2 of 0.58 for the subset of counties which include both court and incarceration coverage. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY23-013.

Figure A13: County variation in degree of child indirect exposure rates, by child race

A: Map of county-level index variation, White children



B: Map of county-level index variation, Black children



Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, the CJARS relations and residence crosswalks, and IRS Form 1040s.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. The sample consists of individuals in the Census Numident 1999–2000 birth cohorts in CJARS-covered geographies. Map markers are sized according to 2021 Census Bureau county total population estimates. All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY23-0138.

Table A1: Correlation between estimated race-specific county-level exposure index and county and commuting zone characteristics

	Estimated coefficient	p-value
School expenditure per student	-0.429***	0.000
Opportunity Insights' place effect for children with parents in the 25th percentile of income (county-level)	-0.260***	0.000
Teacher student ratio	-0.226***	0.000
Manufacturing employment share	-0.147***	0.000
Growth in Chinese imports 1990–2000	-0.093**	0.017
Racial segregation	-0.067*	0.085
Fraction with commute less than 15 minutes	-0.065*	0.094
Indicator for urban areas	-0.032	0.407
Income segregation	0.026	0.510
Fraction foreign born	0.090**	0.021
Local tax rate	0.091**	0.019
Migration inflow rate	0.222***	0.000
Migration outflow rate	0.232***	0.000
Fraction black in the population	0.245***	0.000
Number of CJARS-covered county by race cells	2,900	

Source: Calculations are based on the Census Numident, the Census BestRace files, CJARS, and the CJARS relations and residence crosswalks.

Notes: Estimates and sample sizes have been rounded to preserve confidentiality. Reported are correlation coefficients (after standardizing variables to mean zero and standard deviation of 1) and corresponding p-values between county and commuting zone measures in CJARS-covered geographies. The exposure index is created using CJARS and all other measures are from publicly available Opportunity Insight (OI) data. The OI county effects are the percentage gain (or loss) in income at age 26 from spending one more year of childhood in a given county relative to the national mean. All other OI measures are at the commuting zone level and come from Chetty et al. (2014). All results were approved for release by the U.S. Census Bureau, Data Management System number P-7500378 and approval number CBDRB-FY23-0235. * p<0.1, ** p<0.05, *** p<0.01.

B Data appendix

Constructing the residency and relations crosswalks

The crosswalks are currently created for all individuals in the Census Numident with a valid birth year and born between 1960 and 2018 and not deceased by 1969; currently the first year we observe residence information.[34] The Census Numident is the "backbone" of the residence crosswalks, setting the population and identifying date and place of birth.[35] Address-level information is harmonized for all subsequent years based on the 2000 and 2010 Decennial Censuses, American Community Survey (2001–2018), IRS Form 1040 tax filings (1969, 1974, 1979, 1984, 1989, 1994, 1995, 1998–2018 tax years), IRS Form 1040 electronic tax filings (2005, 2008–2012), Department of Housing and Urban Development (HUD) program data (Longitudinal PIC/TRACS: 1995–2016, 2018; PIC: 2000–2014; TRACS: 2000–2014) and county-level information from Medicare (2000–2017 EBD) and Medicaid (2000–2014 MSIS) enrollment databases, Indian Health Service (IHS) from 1999–2017, and the MAF-ARF (2000–2018).[36] Data are linked at the person level using a Protected Identification Key (PIK) created by the Census Bureau's Person Identification Validation System (PVS).[37] Similarly, addresses are assigned MAFIDs, a numeric key, to protect PII. If more than one MAFID (i.e., address) is provided for an individual in a given year, the following priority ranking is applied: decennial census, IRS Form 1040, IRS Form 1040 ELF, American Community Survey, CMS EDB, HUD Longitudinal PIC/TRACS, HUD PIC, HUD TRACS, IHS, MAF-ARF, CMS MSIS.

The residence crosswalks are the basis of the familial crosswalks. First, for each year, all coresidence pairs are created at a given address. Group quarters and addresses with more than 20 individuals identified at the locations in a given year are suspect of not having familiar relations and thus not used to create pairwise relations among all cohabitants.

[34]The Census Numident is sourced from the Social Security Administration (SSA) Numident file, which tracks all events related to Social Security Numbers (SSN) and Individual Taxpayer Identification Numbers (ITINs) including applications, changes, and deaths. The Census Numident is a research file that de-identifies the information by assigning a unique Protected Identification Key (PIK) for all SSNs and ITINs. For further explanation of these files, see Finlay and Genadek (2021) and Genadek et al. (2022).

[35]We use a place-of-birth crosswalk that links unique place names and states to county and state FIPS codes. This crosswalk is an adaptation of the ones developed by Bailey et al. (2020) and others.

[36]We note that the MAF-ARF is used as a last resort when determining the best address for each individual in each year; this is because multiple address are reported per individual in years prior to 2012 with only one address reported after 2012, without knowing the source of the multiple addresses or the single address chosen post 2012.

[37]PIKs are used to link data at the person level within the Census Bureau's Data Linkage Infrastructure. PIKs can be assigned deterministically using only SSN or probabilistically using names, dates of birth, addresses, and other information as inputs into the Person Identification Validation System (PVS). For further explanation of this process, see Wagner and Lane (2014).

This should not impact individuals living in apartment buildings, since individual units are assigned unique address identification numbers. Instead, examples of group quarters include dormitory facilities, assisted living facilities, homeless shelters, nursing homes, and prisons. For children who are living temporarily in group quarters (like a homeless or emergency shelter), we rely on the years of their childhood when they are not living in group quarters environments to build their relationship information. So if they coreside with their parents and other potential caregivers before or after the period of time in group quarters, there will be no loss of linkages based on our processing algorithm.

Relationships are enhanced above cohabitation based on information in the 2000 and 2010 decennial censuses, American Community Survey (2001–2018), IRS Form 1040 tax filings (1969, 1974, 1979, 1984, 1989, 1994, 1995, 1998–2018 tax years), IRS Form 1040 electronic tax filings (2005, 2008–2012), and HUD program data (1995–2018, 2018). The Decennial Censuses, American Community Surveys, and HUD program data each provide relationships between the household head and other household members, which is used to directly establish relationship types and infer the relationship between household members. Additionally, tax filing and claiming behavior establish spousal relationships and dependents. Finally, we include the Census Household Composition Key (CHCK) which creates links between children and parents based on information on birth certificates for children born between 1999–2018 (Luque and Wagner, 2015).[38] Children who do not have parental links established by the CHCK file could be due to the father's or mother's information being left off of the birth certificate, inaccurate parent information, the parent not being assigned a PIK (SSN or ITIN), or an inability to match the child-parent pair to the same address to confirm the link (Luque and Wagner, 2015; Genadek et al., 2022; Bond et al., 2014). We provided new statistics validating the relationship pairs identified by the CHCK file. First, We confirm that 93% of the biological relationships identified by our crosswalk are also observed in the CHCK file. Second, we document that about 70% of all CHCK relations are confirmed as biological parents by survey microdata or HUD program data, which increases to 97% once unclassified caregivers (those observed claiming a child on a 1040 tax form accompanied by no other observable information) are reclassified as biological parents.

Relationship types are established by combining the multitude of observations between pairs across data sources and years into the sets show in Table B1. For several reasons, many relations can only be classified into one of the main categories without the additional detailed information required to classify the relationship pair into a subcategory. First, relationships in the decennial censuses, American Community Survey, and HUD program data are all

[38]The Census Bureau uses parents' names from birth certificates to probabilistically assign PIKs through the PVS and the child's SSN which uniquely determines the PIK to match to children in the Census Numident ages 0–18 as of 2018 and 2019. Since parents' SSNs are not available, the CHCK file requires that the child-parent link be confirmed at the same address in the PVS reference file.

Table B1: Relationship types in CJARS crosswalks

Relation Code	Relation Description
10	Cohabiting adults (=<13 years apart)
11	Spouse
12	Domestic partner
13	Romantic unmarried (e.g., boyfriend/girlfriend)
14	Unclassified romantic
15	Adult, non-romantic (e.g., roommate/boarder)
20	Cohabiting adult-minor (>13 years apart)
21	Bio parent - child
22	Adopted parent - child
23	Stepparent - child
24	Foster parent - child
25	Unclassified parent - child
26	Parent - child-in-law
27	Grandparent-grandchild
28	Aunt/uncle-niece/nephew
29	Non-familial adult - child
40	Cohabiting minors (=< 13 years apart)
41	Bio-siblings
42	Adopted-siblings
43	Step-siblings
44	Foster-siblings
45	Unclassified siblings
45	Cousins
45	Second cousins
45	Siblings-in-law

expressed in relation to the household head or adult respondent. Several assumptions are imposed to infer relations between other household members, but often there is not enough information to classify a link beyond an unclassified parent-child link.[39] Second, the American Community Survey between 2001 and 2007 and the 2010 Decennial Census use broader relationship definitions.[40] Finally, parental relations that are established only in the tax

[39]Some of the assumptions imposed to define parent-child relations (and vice versa) are: (1) if a household head has a biological child, then the spouse to the household head is also linked as a biological parent; (2) if a household head has a stepchild, then the spouse to the household head is assumed to be the biological parent; (3) if a household head has an adopted child, then the spouse to the household head is assumed to be an adopted parent; and (4) if a household head has a foster child, then the spouse to the household head is assumed to be the foster parent as well. Examples of inferred parent-child links where subclassification can not be ascertained are: (1) a household head linked to their parent is not further classified as biological; step, adopted, or foster; (2) a spouse of a household head linked to the household head's mother/father-in-law is not further classified as biological, step, adopted or foster; (3) a sibling of the household head linked to the niece/nephew of a household head, subject to age restrictions and other information if available; and (4) a child of a household head linked to a grandchild of the household head, subject to age restrictions and other information if available.

[40]The following relationship types to the household head were removed from the 2010 Decennial Census: sibling-in-law, nephew/niece, uncle/aunt, cousin, grandparent, and foster child. The ACS did not offer more

records and not observed in the Census surveys or HUD program data (or are observed with ambiguous relationships) can only be classified as a parent with no further information to sub-classify into biological, step, adopted, foster, aunt/uncle, etc.[41]

A relation pair may be observed multiple times across source files and years. Relationship types only need to be defined once in order to assign it to the cohabiting pair. This approach is beneficial since individuals may be observed multiple years in the tax records, but without detailed relational information and may be observed only once by the Decennial Census or ACS.

Relationship types between pairs are sequentially established based on the strength of the source information. For example, relations established in surveys with the head of household directly define a relationship, while relations between other household members are inferred. Figure B1 demonstrates the iterative process and assumptions used when defining relationship types.

Performance of the residential and relationship crosswalks

First, we benchmark fertility statistics calculated from the CJARS family crosswalk against published statistics in the 2016 National Center for Health Statistics and 2017 National Vital Statistics System reports (Mathews and Hamilton, 2016, 2019). In Figure A2 Panel A, we show the cumulative distribution of age at first birth by race for females born in the U.S. in 1981, as measured by the Census Numident.[42] The average age of first birth overall is 25.01 based on our crosswalks, which is inline with an overall age of first birth in 2000 and 2014 of 24.9 and 26.3, respectively (Mathews and Hamilton, 2016). Asian women have the highest average age at first birth (our estimate 28.15; NCHS in 2000: 27.8), followed by White (25.67; 25.9), Hispanic (23.23; Mexican 22.2, Central and South American 24.8), American Indian/Alaska Native (22.85; 21.6), and Black (22.67; 22.3) women. In Panel B, we show the number of births observed until the age of 37, the latest possible age for this cohort. Again, our observed birth rates for women born in 1981 overall (1.673) are inline with the reported birth rate by NVSS in (1.766). American Indian/Alaska Native women have the highest

detailed classifications for parent-child links (namely, biological, adopted, or step) or in-law links (parent or child) until 2008.

[41] A parent-child relationship is assumed if the age difference between the individuals is less than 45 years and a grandparent-child relation if the age gap is 45 years or greater. However, it is possible for an aunt or uncle to claim a child on tax records, although treating them as an unclassified parent in this circumstance is likely permissible for our application.

[42] Due to data availability, we do not observe fertility beyond age 37 for this cohort and have more limited ability to observe birth prior to age 18 since the CHCK file covers children starting in 1999. Observed births are measured in our crosswalks as a reported biological child, and age at birth is defined using the year of birth for the child.

number of observed births (our estimate: 2.047, NVSS not available), followed by Hispanic (1.823; 2.01), Black (1.744; 1.825), White (1.655; 1.666), and Asian (1.39; not available) women.

Next, we turn to our ability to measure caregiver links for children in our core sample. Using the previously described crosswalks, we identify female (male) biological parents for 90% (76%) of children born between 1999 and 2005 (Figure IA). When we relax the requirement for an explicitly defined relationship status and expand to include other types of adult household members, we observe 97% and 95% of children are linked with one or more female and male adult potential caregivers, respectively.[43]

There are important differences by race, due to both systematic differences in household structure and in our ability to observe potential caregivers in administrative and survey records. First, we see that only 55% of Black, non-Hispanic children are connected to a biological male parent in the data. This could be due to either fathers being excluded from birth records, not coresiding with children during household surveys, or not claiming their children as dependents in tax filings. However, White, non-Hispanic and Black, non-Hispanic children have very similar rates of being linked to a female biological parent (94%). Second, Hispanic children are much less likely to be observed with a female biological parent (75%) as well as male biological parents (60%), and just less than 90% are observed with a male and female potential caregiver; this is likely the result of these individuals being less likely to have a Social Security Number (SSN) or Individual Tax Identification Number (ITIN), which is needed for individuals to receive a PIK and be linked across data sets within the Census Bureau's Data Linkage Infrastructure (Bond et al., 2014).

Figure IB documents the share of children born between 1999 and 2005 observed with other types of intergenerational relationships in the household. We find that 4.7%, 22%, 29%, and 46.2% of children are observed with a step/adopted/foster parent, extended family (grandparent/aunt/uncle), unclassified caregivers, and unclassified cohabiting adults, respectively.[44] Unclassified caregivers, which occur when we observe an adult-child link resulting of dependency claims in IRS tax records without any other information to pin down the nature

[43]Figure A3 documents the distribution of number of links identified overall (Panel A) and by racial and ethnic subgroup (Panel B).

[44]Fewer children are observed with step/adopted/foster parents than what is observed in the SIPP (Sweeney, 2010; Kreider and Ellis, 2011; Raley and Sweeney, 2020); this is likely due to relationship misclassification into biological parents and unclassified caregivers due to relations being reported to the household head in the Census surveys and HUD program data and a lack of relational information beyond claiming behavior on tax forms. For example, if a household head has a biological child, then it is assumed they share the biological child with their spouse. However, the relation could be a step or adopted parent. Similarly, if a child is claimed by someone with an age gap less than 45 years and the relation is not observed in the CHCK file, then the relation is considered an unclassified caregiver. However, the relation could be a step/adopted/foster parent, an aunt or uncle, or a younger grandparent.

of their relationship, are more likely in multigenerational households, and among biological parents that are left off of birth records (as proxied by the CHCK file).[45] Importantly, all these relations *seem* like important parental figures given the tax filing relationship even if they are not included in the CHCK, which is our closest approximation to a birth record database. Unclassified adults, in contrast, are other cohabiting relations that are either explicitly classified as non-familial in Census Bureau surveys or just adults that we observe coresiding at the same address as the child (e.g., live-in boyfriends, roommates, etc); notably, we require that these relations must coreside for 2 years or more in order to focus on those with greater potential familial attachment.[46]

Again, there are important differences in the share of children that are observed with various caregiver relations by race. Minority children are much more likely to have an extended family caregiver (Black, non-Hispanic 38%, Hispanic 28%, Asian 32%, and American Indian/Alaska Native 31%) than their White, non-Hispanic counterparts (17%); this is consistent with previously documented differences in household structures by race and ethnicity in the U.S. (e.g., Lofquist, 2012; Cohen and Passel, 2018).

Approximately 48.6% (49.3%) of children are linked with just one female (male) potential caregiver, a grouping that combines biological parents and other potential adult caregivers in their household. 24.6% (23.3%) of children are observed with two female (male) potential caregivers, while 23.7% (22.8%) are linked with 3 or more female (male) potential caregivers. A variety of living circumstances could give rise to more than one potential caregiver of the same sex being linked with a child, for example: (1) parents with multiple romantic partners (due to divorce or separation) while raising their children, (2) households with same-sex romantic partners, (3) multigenerational households, or (4) doubled up households where multiple families share the same accommodations. The high rate of children linked with 2 or more potential caregivers of the same sex reflects the experience that many children today in the U.S. have of growing up with multiple adult influences in their households beyond those of the traditional nuclear family.

[45] Household surveys enumerate relationships of all individuals in the household with respect to the head of household. Relationships between other individuals must be inferred, which is increasingly complicated in households that extend beyond nuclear families.

[46] Utilizing a 2-year coresidency requirement also minimizes the influence of errors in the probabilistic record linking process for address information that might lead children to be labeled as coresiding with adults who in fact do not live at the same address.

Figure B1: Sequential process to establish relations beyond cohabitation

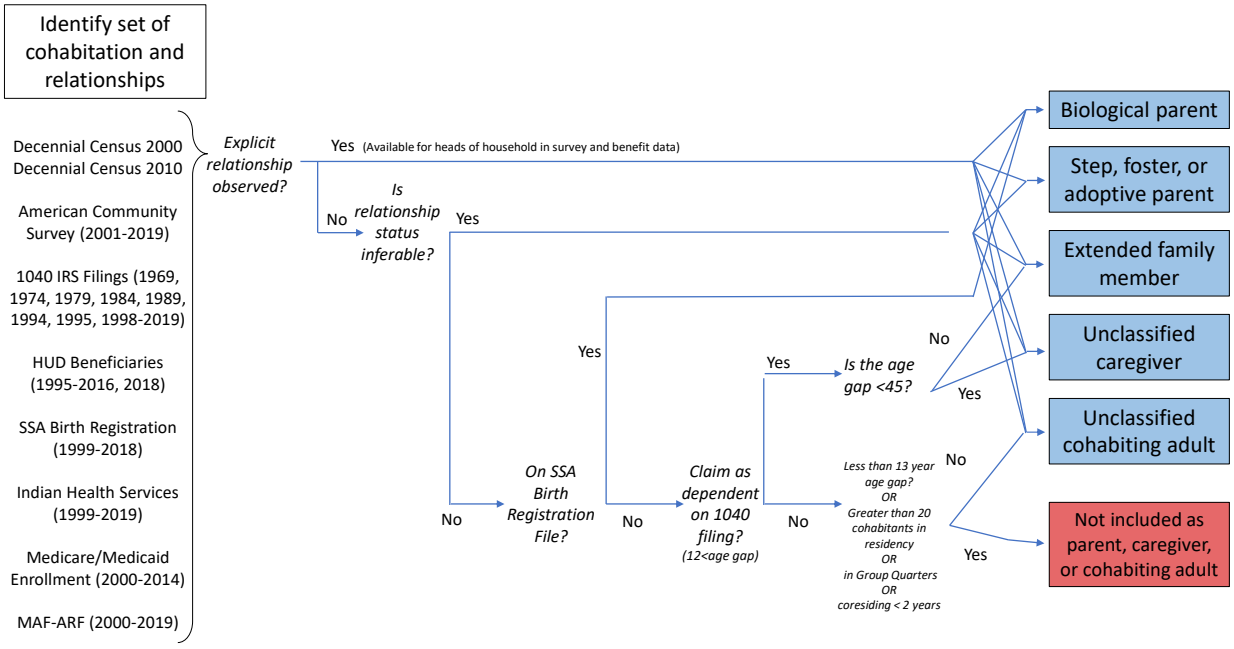


Table B2: Source files contributing to the residence crosswalk

Source	Years	Variables
MAF-X	2017	MAFID, state, county, group quarters flag
IRS Form 1040	1969	primary filer, MAFID, state
—	1974, 1979, 1984,1989	primary and secondary filer, MAFID, state
—	1994, 1995, 1998–2019	primary and secondary filer, four dependents, MAFID, state
IRS Form 1040 ELF	2005, 2008–2012	primary and secondary filer, 20 dependents
Decennial Census	2000, 2010	household members, MAFID, state, county, group quarters flag
ACS	2001–2004†	household members, state, county
—	2005–2018	household members, state, county, MAFID, group quarters flag
HUD Longitudinal PIC/TRACS	1995–2016, 2018	household members of enrollees, state, county, MAFID
HUD PIC	2000–2014	household members of enrollees, state, county, MAFID
HUD TRACS	2000–2014	household members of enrollees, state, county, MAFID
CMS EDB	2000–2019	enrollees, state, county, MAFID
CMS MSIS	2000–2014	enrollees, state, county
Indian Health Service	1999–2019	enrollees, state, county, MAFID
MAF-ARF	2000–2018	individuals with SSN or ITIN, state, county, MAFID
Census Numident	2021Q1	individuals with SSN or ITIN, place of birth (linked to state, county, and commuting zone), date of birth, date of death, sex, race

† There are small samples during the ACS trial years between 2001 and 2004, so while statistical information is not used for these years, respondents' address-level information and household relationships are used in the creation of these crosswalks.

Figure B2: Construction of residential and relations crosswalks

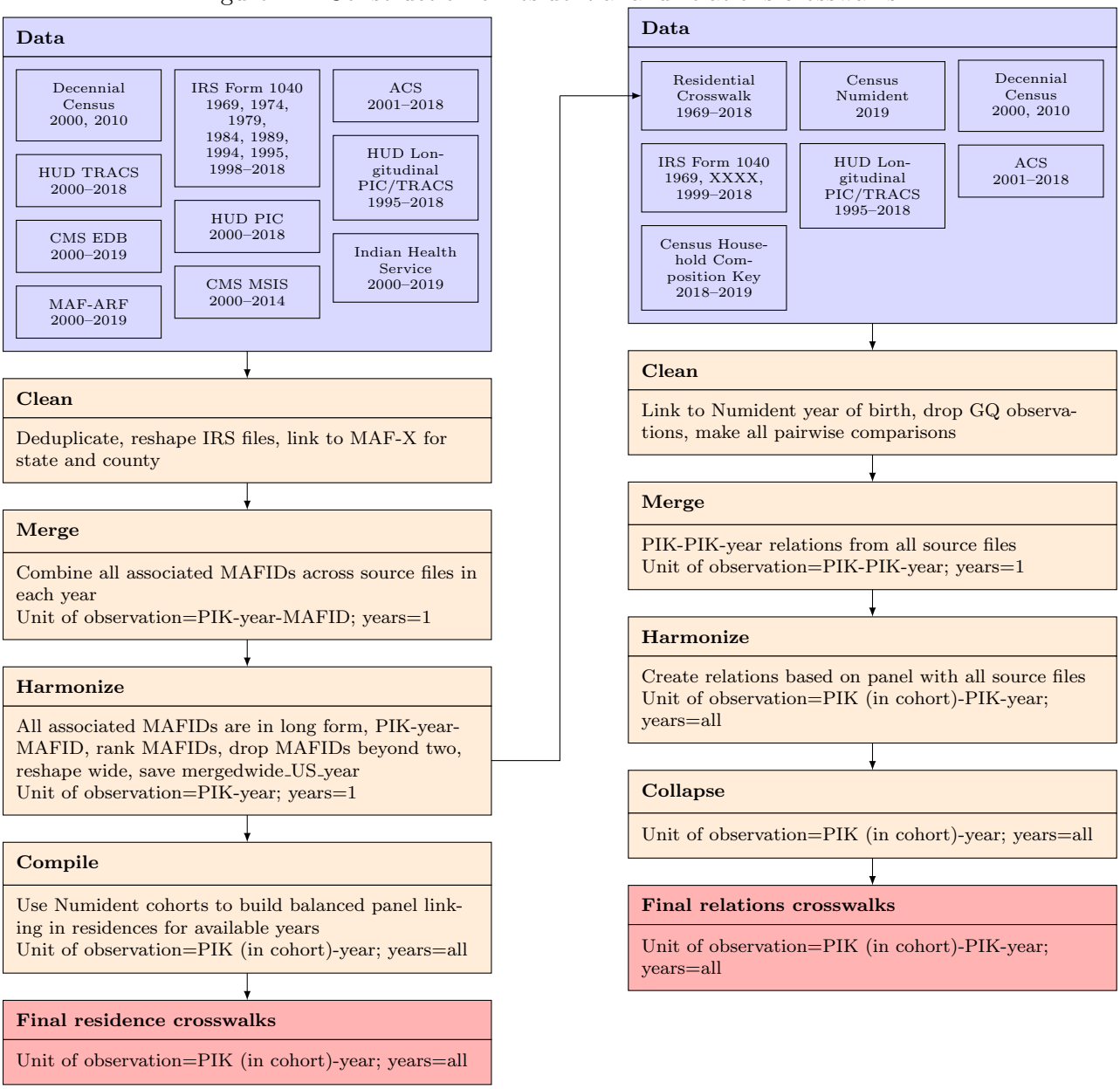
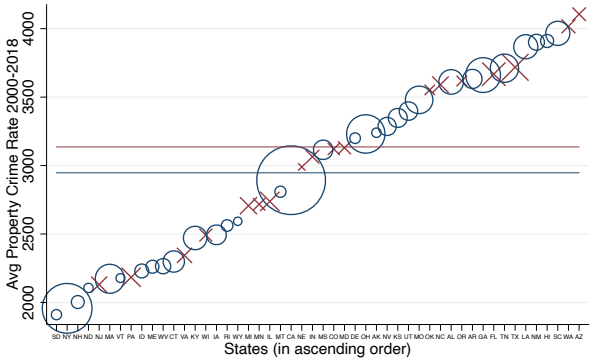
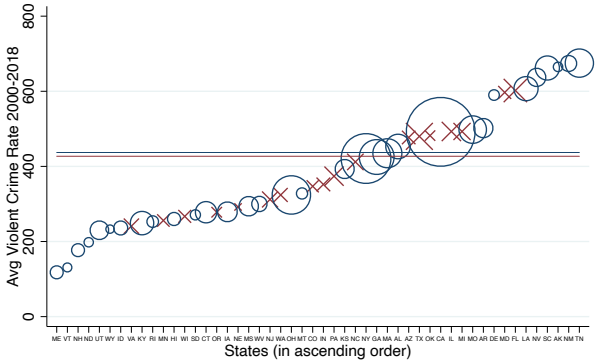


Figure B3: State comparisons of crime and criminal justice outcomes

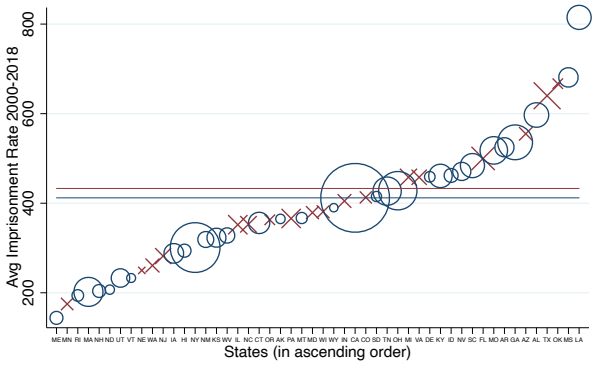
A: Property Crime Rates (UCR)

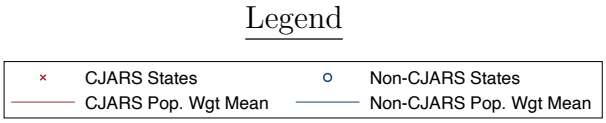


B: Violent Crime Rates (UCR)



C: Incarceration Rates (BJS)



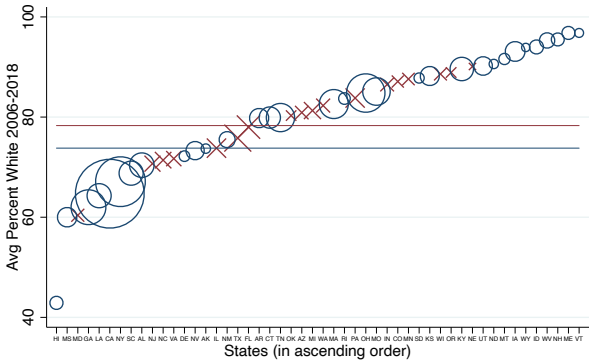


Source: Calculations from publicly available data from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program and the Bureau of Justice Statistics National Prisoner Statistics program.

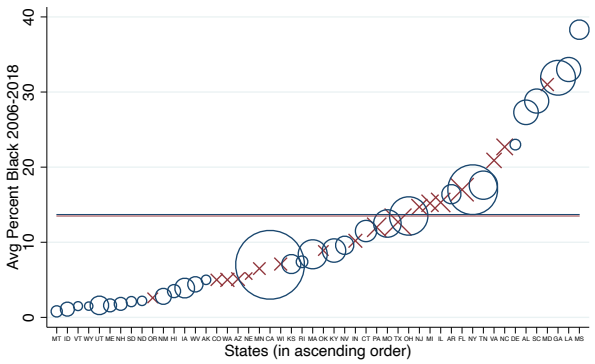
Notes: The rates per 100,000 residents have been averaged for each state over the period 2000–2018. Marker sizes are proportional to each state's population, averaged over the years 2000–2018.

Figure B4: State comparisons, socioeconomic outcomes

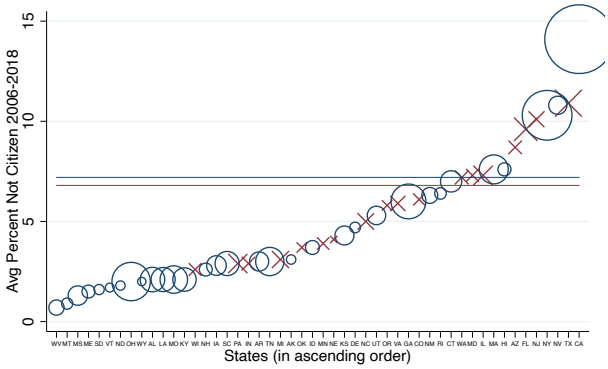
A: Percent Population White



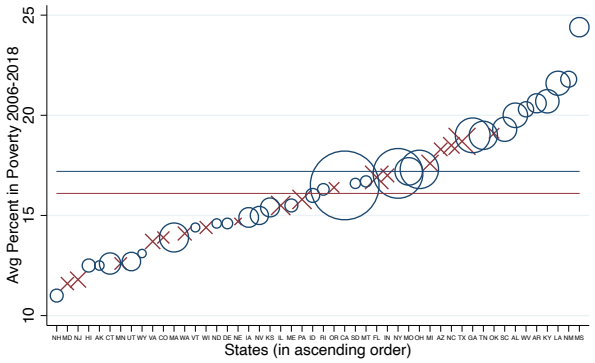
B: Percent Population Black



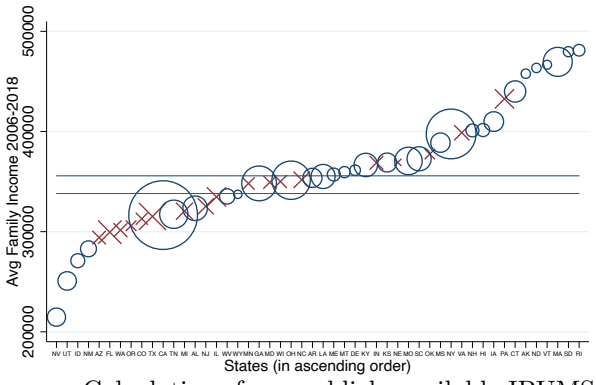
C: Percent Not A Citizen

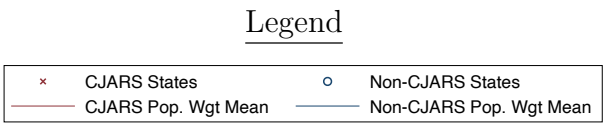


D: Percent In Poverty



E: Family Total Income





Source: Calculations from publicly available IPUMS USA 2006–2018 ACS data.

Notes: State averages reported and weighted by the average population during the same time period (Ruggles et al., 2021). Marker sizes are proportional to each state's population averaged over the years 2000–2018.

Table B3: Source files contributing to the relations crosswalk

Source	Years	Variables
Residential Crosswalk	1969–2019	cohabiting pairs, except those in group quarters or at an address with more than 20 individuals in a single year
IRS Form 1040	1974, 1979, 1984, 1989	primary and secondary filer
—	1994, 1995, 1998–2019	primary and secondary filer, four dependents
IRS Form 1040 ELF	2005, 2008–2012	primary and secondary filer, 20 dependents
Decennial Census	2000, 2010	household members and relation to household head
ACS	2001–2018[†]	household members and relation to household head
HUD Longitudinal PIC/TRACS	1995–2016, 2018	household members of enrollees and relation to household head
Census Household Comp. Key	2018, 2019	individuals with SSN between 0 and 18 years of age, child, mother, father links
Census Numident	2021Q2	individuals with SSN or ITIN, place of birth (linked to state, county, and commuting zone), date of birth, date of death, sex, race

[†] There are small samples during the ACS trial years between 2001 and 2004, so while statistical information is not used for these years, respondents' address-level information and household relationships are used in the creation of these crosswalks.

Table B4: Ten most common criminal charges within each offense category

Violent	Property	Drug
1200 Aggravated assault	2040 Forgery/fraud	3150 Possession/use of marijuana
1240 Extortion threat	2070 Theft	3160 Possession/use of unspecified drug
1180 Armed robbery	2010 Burglary	3250 Drug paraphernalia
1230 Simple assault	2140 Criminal trespass	3080 Distribution, drug unspecified
1990 Other violent offense	2050 Grand theft	3110 Possession/use of cocaine or crack
1090 Child molestation	2110 Destruction of property	3990 Other drug offense
1220 Child abuse	2060 Petty theft	3140 Possession/use of controlled substance
1070 Rape	2100 Receiving stolen property	3070 Distribution of marijuana
1010 Murder	2120 Hit and run driving, property damage	3030 Distribution of cocaine or crack
1060 Kidnapping	2130 Unauthorized use of a vehicle	3100 Possession of amphetamines
Driving under the influence	**Public order**	**Criminal traffic**
4020 Driving under the influence of alcohol	5130 Obstruction/resisting arrest	6010 Traffic offense, minor
4010 Driving while intoxicated	5170 Disorderly conduct offense	
4030 Driving under the influence of drugs	5990 Public order offense, other	
	5180 Liquor law violation	
	5040 Weapons offense	
	5090 Other court offense	
	5080 Contempt of court/court order violation	
	5070 Probation violation	
	5150 Commercialized vice	
	5110 Offense against morals/decency	

Source: Estimates calculated from CJARS court records held by the University of Michigan and not protected by 13 USC §9a.
Notes: Offense codes follow the classification schema outlined in Choi et al. (2023).