What Can We Learn About the Incidence of Foster Care Placement from Birth Records?

Findings from the Cross Jurisdiction Model Replication Project TECHNICAL APPENDIX

September 2022

OPRE Report 2022-255

Authors:
Emily Putnam-Hornstein, University of North Carolina at Chapel Hill, Children's Data Network; Eunhye Ahn, Washington University in St. Louis, Children's Data Network; and Rhema Vaithianathan, Auckland University of Technology, Children's Data Network

Jared Parrish and Robyn Husa, Alaska Division of Public Health

Matthew Walton and James P. Perry, Kentucky Office of Data Analytics

Claire Smither Wulsin, M.C. Bradley, and Beth Varley, Mathematica

Technical Appendix

The technical appendix provides additional contextual information on the jurisdictions (California, Alaska, and Kentucky) that implemented the risk prediction model; the data sources and methods used to develop, validate, and replicate the model; and findings from the CJMR project.

A. State contexts

Exhibit.1. Demographic statistics by state, 2020

	Alaska	California	Kentucky
Racial/ethnic demographics			
White	64.5%	71.1%	87.1%
Black or African American	3.6%	6.5%	8.6%
American Indian/Alaska Native	15.7%	1.7%	0.3%
Asian	6.6%	15.9%	1.7%
Native Hawaiian and Other Pacific Islander	1.6%	0.5%	0.1%
Two or More Races	7.9%	4.2%	2.2%
Hispanic or Latino (includes White and non-White)	7.5%	40.2%	4.2%
Persons in poverty	9.6%	11.5%	14.9%
Median household income	$77,790	$78,672	$52,238

Source: Census Bureau, 2021.

Exhibit 2. Overview of state context and current snapshot, 2020

	California	Alaska	Kentucky	Table title in source
Overview of state context				
Number of child victims with substantiated or indicated maltreatment in 2020	60,317	3,212	16,748	Table 3–3 Child Victims, 2020
Rate of child victims with substantiated or indicated maltreatment per 1,000 children	6.9	18	16.7	Table 3–3 Child Victims, 2020
Snapshot of current child welfare statistics				
Total child population in 2020	8,791,234	178,731	1,001,917	Table C–2 Child Population, 2020
Number of children referred for alleged maltreatment or neglect	359,699	22,687	95,378	Table 2–1 Screened-in and Screened-out Referrals, 2020; “total referrals”
Percentage of the total child population	4.1%	12.7%	9.5%	
Number of referred children who were the subjects of an investigation	306,919	15,460	67,066	Table 3–1 Children Who Received an Investigation or Alternative Response, 2020

	California	Alaska	Kentucky	Table title in source
Percent of referred children who were the subjects of an investigation	85.3%	68.1%	70.3%	
Number of referred children with a substantiated report	60,317	3,212	16,748	Table 3–3 Child Victims, 2020
Percent of referred children with a substantiated report	16.8%	14.2%	17.6%	
Number of children who entered foster care	24,748	1,353	5,204	Entered Foster Care During FY, Number of children' Entering Care & Median Length of Stay
Percent of children referred for alleged maltreatment or neglect who were placed in out-of-home care	6.9%	6.0%	5.5%	

Sources: U.S. Children's Bureau 2020, 2021.

B. Data sources

Exhibit 3. List of variables used in the Alaska and Kentucky models

Variable	Description	Source	Birth years (birth cohort)
Alaska			
BIRTH_DT	child's birthday	Birth Certificate (Vital Records)	2013–2016
sex	child's sex	Birth Certificate (Vital Records)	2013–2016
mage	maternal age at birth	Birth Certificate (Vital Records)	2013–2016
fbthdate	paternal birth date	Birth Certificate (Vital Records)	2013–2016
fage	paternal age at birth	Birth Certificate (Vital Records)	2013–2016
feduc	paternal education	Birth Certificate (Vital Records)	2013–2016
fracecodem	paternal multi-race code	Birth Certificate (Vital Records)	2013–2016
fsporig	paternal ethnicity Hispanic/Latino	Birth Certificate (Vital Records)	2013–2016
meduc	maternal education	Birth Certificate (Vital Records)	2013–2016
mracecodem	maternal multi-race code	Birth Certificate (Vital Records)	2013–2016
msporig	maternal ethnicity Hispanic/Latina	Birth Certificate (Vital Records)	2013–2016
precare	prenatal care	Birth Certificate (Vital Records)	2013–2016
mfood	WIC food	Birth Certificate (Vital Records)	2013–2016
prevlbl	number of previous birth live	Birth Certificate (Vital Records)	2013–2016
prevlbd	number of previous birth dead	Birth Certificate (Vital Records)	2013–2016
paymsopc	birth payment method for delivery	Birth Certificate (Vital Records)	2013–2016
probl_1	pre-pregnancy problems	Birth Certificate (Vital Records)	2013–2016
probl_2	onset of labor problems	Birth Certificate (Vital Records)	2013–2016
mdel	method of delivery - cesarean	Birth Certificate (Vital Records)	2013–2016
fetlbth	fetal presentation	Birth Certificate (Vital Records)	2013–2016
bthweight	child birth weight	Birth Certificate (Vital Records)	2013–2016
obstegest	obstetric estimate of gestation at deliver	Birth Certificate (Vital Records)	2013–2016

Variable	Description	Source	Birth years (birth cohort)
probl_3	code abnormal conditions of child	Birth Certificate (Vital Records)	2013–2016
fst_rpt	indicator of first report to OCS (yes/no)	Office of Children's Services (OCS)	2013–2016
fst_rpt_dt	date of first OCS report	Office of Children's Services (OCS)	2013–2016
fst_sbst	indicator of a substantiated OCS report (yes/no)	Office of Children's Services (OCS)	2013–2016
fst_sbst_dt	date of first substantiated OCS report	Office of Children's Services (OCS)	2013–2016
PE_S	indicator of foster care placement (yes/no)	Office of Children's Services (OCS)	2013–2016
PE_S_DT	date of first foster care placement	Office of Children's Services (OCS)	2013–2016
Kentucky			
BIRTH_DT	child's birthday	Office of Vital Statistics (OVS)	2015–2021
sex	child's sex	Office of Vital Statistics (OVS)	2015–2021
mage	maternal age at birth	Office of Vital Statistics (OVS)	2015–2021
fbthdate	paternal birth date	Office of Vital Statistics (OVS)	2015–2021
fage	paternal age at birth	Office of Vital Statistics (OVS)	2015–2021
feduc	paternal education	Office of Vital Statistics (OVS)	2015–2021
fracecodem	paternal multi-race code	Office of Vital Statistics (OVS)	2015–2021
fsporig	paternal ethnicity Hispanic/Latino	Office of Vital Statistics (OVS)	2015–2021
meduc	maternal education	Office of Vital Statistics (OVS)	2015–2021
mracecodem	maternal multi-race code	Office of Vital Statistics (OVS)	2015–2021
msporig	maternal ethnicity Hispanic/Latina	Office of Vital Statistics (OVS)	2015–2021
precare	prenatal care	Office of Vital Statistics (OVS)	2015–2021
mfood	WIC food	Office of Vital Statistics (OVS)	2015–2021
prevlbl	number of previous birth live	Office of Vital Statistics (OVS)	2015–2021
prevlbd	number of previous birth dead	Office of Vital Statistics (OVS)	2015–2021
paymsopc	birth payment method for delivery	Office of Vital Statistics (OVS)	2015–2021
probl_1	pre-pregnancy problems	Office of Vital Statistics (OVS)	2015–2021
probl_2	onset of labor problems	Office of Vital Statistics (OVS)	2015–2021
mdel	method of delivery - cesarean	Office of Vital Statistics (OVS)	2015–2021
fetlbth	fetal presentation	Office of Vital Statistics (OVS)	2015–2021
bthweight	child birth weight	Office of Vital Statistics (OVS)	2015–2021
obstegest	obstetric estimate of gestation at deliver	Office of Vital Statistics (OVS)	2015–2021
probl_3	code abnormal conditions of child	Office of Vital Statistics (OVS)	2015–2021
PE_S_DT	date of first foster care placement	The Workers Information System (TWIST)	2015–2021

1. Data use agreements

a. Alaska

Prior to this CJMR work, Alaska DPH had an existing, formalized data use agreement to link epidemiological and vital birth records with child welfare records to conduct public health research. The Alaska Longitudinal Child Abuse and Neglect Linkage project (ALCANLink) annually integrates the Alaska Pregnancy Risk Assessment Monitoring System (PRAMS) data with OCS records and other administrative records. This project is easily expandable to integrate all birth records when specific projects (such as this) are warranted. The current data use agreement specifies the data elements that OCS provides to Alaska DPH for integration and provides provision for general use when exploring public health population level research questions. This work falls within the purview of the current data use agreement.

b. Kentucky

CHFS has a data sharing structure in place whereby a master data agreement (MDA) has been executed that issues blanket authority for CHFS agencies to share data among one another, provided a supplemental agreement is completed. This supplemental agreement is effectively a memorandum of understanding between CHFS agencies that outline the terms of a data sharing process (e.g., the purpose for sharing, data to be provided, legal justifications and requirements) and documents each party's agreement to abide by those terms. From beginning to end, this procedure involves communicating with each group to clarify the goals of the project, identifying the necessary data elements, reviewing the privacy and legal implications of the project, and sending the document through the administrative review process before it is executed. For the CJMR project, this process took approximately three months from the initial phase of communication with project partners and data stewards to the completion of the executed agreement.

C. Methods

1. Development

Exhibit 4. List of predictor variables used in the model

Variable	Description	Source	Birth years (birth cohort)
California			
sex	child's sex	Birth Certificate (Vital Records)	2000
mage	maternal age at birth	Birth Certificate (Vital Records)	2000
fbthdate	paternal birth date	Birth Certificate (Vital Records)	2000
fage	paternal age at birth	Birth Certificate (Vital Records)	2000
feduc	paternal education	Birth Certificate (Vital Records)	2000
fracecodem	paternal multi-race code	Birth Certificate (Vital Records)	2000
fsporig	paternal ethnicity Hispanic/Latino	Birth Certificate (Vital Records)	2000
meduc	maternal education	Birth Certificate (Vital Records)	2000
mracecodem	maternal multi-race code	Birth Certificate (Vital Records)	2000
msporig	maternal ethnicity Hispanic/Latina	Birth Certificate (Vital Records)	2000
precare	prenatal care	Birth Certificate (Vital Records)	2000
mfood	WIC food	Birth Certificate (Vital Records)	2000
prevlbl	number of previous birth live	Birth Certificate (Vital Records)	2000
prevlbd	number of previous birth dead	Birth Certificate (Vital Records)	2000
paymsopc	birth payment method for delivery	Birth Certificate (Vital Records)	2000
probl_1	pre-pregnancy problems	Birth Certificate (Vital Records)	2000
probl_2	onset of labor problems	Birth Certificate (Vital Records)	2000

Variable	Description	Source	Birth years (birth cohort)
mdel	method of delivery - cesarean	Birth Certificate (Vital Records)	2000
fetlbth	fetal presentation	Birth Certificate (Vital Records)	2000
bthweight	child birth weight	Birth Certificate (Vital Records)	2000
obstegest	obstetric estimate of gestation at deliver	Birth Certificate (Vital Records)	2000
probl_3	code abnormal conditions of child	Birth Certificate (Vital Records)	2000
PE_S_DT	foster care placement episode start date	Child Protection System Records	2000–2003
BIRTH_DT	child's birth date	Child Protection System Records	2000–2003

2. Machine learning algorithms used to train and test the model

Several candidate classification models were trained and tested using various machine learning algorithms, including LASSO regression, Random Forests, Fast Extreme Gradient Boosting (LightGBM), and Feed-Forward Deep Neural Networks (FNN). LASSO regression performs both shrinkage and coefficient learning for each variable. In other words, some of the predictors are removed from the model to reduce over-fitting. LASSO regression models were trained using 2-times 10-fold cross-validation and the optimal value of the parameters were decided through grid-search over a range of values. After training and testing, the LASSO regression model outperformed other models, and thus, we used the LASSO regression model for all subsequent analysis.

3. Linking data

a. California

Because there was no unique identifier common to birth and child protection services (CPS) records, we probabilistically linked records using a combination of personally identifiable fields captured in both birth and CPS records (e.g., child first and last name, child date of birth, maternal first and last name). The personal identifying information was used only for the linkage purpose, and then removed before the analysis.

b. Alaska

Before using the model, vital birth records for children born between 2013 – 2016 were linked with the Office of Children's Services (OCS) records to obtain OCS IDs (if applicable) and outcomes for each child within the CJMR cohort.

Alaska Department of Family and Community Services (DFCS) implemented both deterministic and probabilistic methods to link the datasets. Prior to all linkages, Alaska DPH conducted systematic record set cleaning, including date, character, and case equalization, standardization of missing data and treatment of special characters, and removal of leading/trailing spaces. Using iterative linkages (deterministic followed by probabilistic), AK DPH reduced the amount of suspected matches requiring manual review. For probabilistic linkages, AK DPH developed comparison patterns based on a Joarowinkler distance metric to account for typos, spelling errors, transpositions, and other edits or deletions between two strings or set of strings and dates. The probabilistic linkage approach automatically accepted matches when the first, last, and alias names, date of birth and sex were identical. Suspected matches that returned a probability match score between 0.85 and 0.99 were manually reviewed, while those below 0.85 were automatically rejected. The RecordLinkage package (Sariyar and Borg 2010) in the R environment (R Core Team 2014) was used for all data linkages (Parrish et al., 2017).

c. Kentucky

Vital birth records and child welfare data are not typically integrated or linked within the Kentucky CHFS system. The two CHFS agencies with primary responsibility and stewardship over these data sources are in distinct operational units within CHFS with separate leadership structures. ODA served as the facilitator to broker the process of obtaining and linking the two datasets.

D. Findings

1. Challenges in the validation and replication phases

a. Validation in California

During the process of validating the model trained on 2012-2015 birth cohorts using 2000 California birth cohort data, several challenges emerged. One issue was that the programs coded for the initial model required revision because they were not coded considering cross-validation. For example, a set of thresholds ('rule_df') had to be recoded to conduct a performance evaluation for the 2000 birth cohort data, and some functions, including 'getPerformanceMetrics,' needed to be revised. We also found that some codes needed to be revisited due to coding errors (e.g., maternal/paternal age and birth payment).

b. Replication in Kentucky

The Kentucky team was unable to obtain birth certificate variables pertaining to maternal smoking during pregnancy but suspects that this would have been a worthwhile variable to include. While it is difficult to gauge whether this variable would have enhanced the predictive power of the model for Kentucky's data, it is worth noting that Kentucky's smoking prevalence is more than twice that of California and among the highest rate in the nation (CDC 2019). Furthermore, there is evidence that this disparity between California and Kentucky is even larger in terms of smoking during pregnancy (Drake et al. 2018).

2. Model performance results

a. California

The area shaded red represents the proportion (that is, density) of children who were not placed in foster care by age 3. The area shaded blue represents the proportion of children who were placed in foster care by age 3. By looking at the x-axis is it possible to examine differences in placement based on the probability score generated by the model. The model works well when there is more overlap between the red and blue areas among the top 10 percent of risk scores. Exhibit 5 shows the predicted risk threshold corresponding to 10 percent of births (the blue dotted line) was 0.68 for the 2012–2015 birth cohorts.

b. Alaska and Kentucky

Focusing on Alaska and Kentucky, the model was better at predicting risk for Alaska (Exhibit 6) than for Kentucky (Exhibit 7). Also note that the predicted risk threshold is higher in Kentucky (0.81) than in Alaska (0.71).

Exhibit 5. Distributions of predicted risk scores stratified by a placement outcome in California

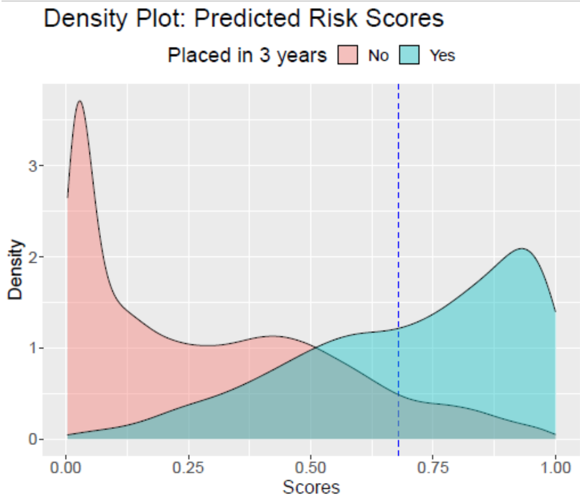


Exhibit 6. Distributions of predicted risk scores stratified by a placement outcome in Alaska

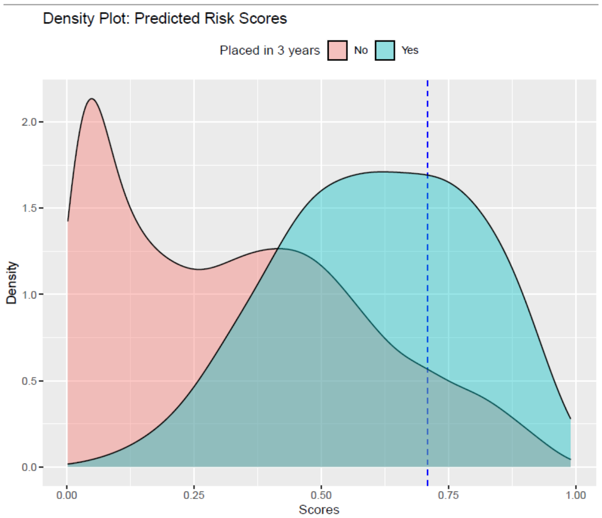
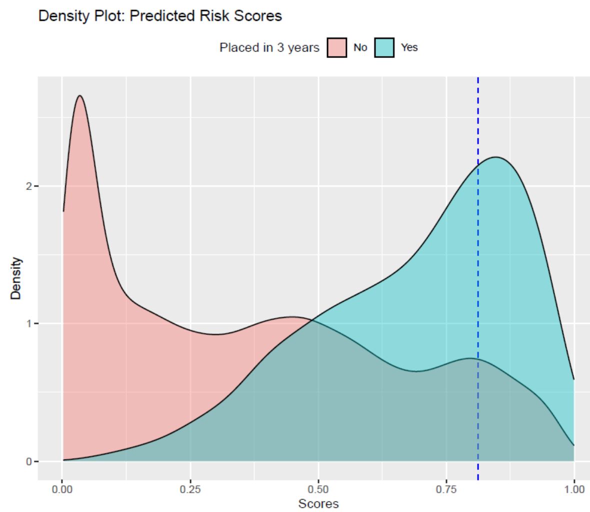


Exhibit 7. Distributions of predicted risk scores stratified by a placement outcome in Kentucky



References

Census Bureau, “QuickFacts.” 2021. https://www.census.gov/quickfacts/fact/table/KY,CA,AK/RHI125220#RHI125220.

Children’s Bureau. “Child Maltreatment 2020.” Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children’s Bureau, 2022. https://www.acf.hhs.gov/sites/default/files/documents/cb/cm2020.pdf.

Children’s Bureau. “Child Welfare Outcomes Report Data: Entered Foster Care During FY.” Adoption and Foster Care Analysis and Reporting System (AFCARS). 2021. https://cwoutcomes.acf.hhs.gov/cwodatasite/enteredCare/index.

Drake, B., M. Jonson-Reid, M.G. Ocampo, M. Morrison, and D. Dvalishvili. “A Practical Framework for Considering the Use of Predictive Risk Modeling in Child Welfare.” *The ANNALS of the American Academy of Political and Social Science,* vol. 692, no. 1, 2020, pp. 162–181. https://doi.org/10.1177/0002716220978200.

Center for Diseases Control and Prevention. “Map of current cigarette use among adults.” Atlanta, GA: Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion. 2019. https://www.cdc.gov/statesystem/cigaretteuseadult.html.

Parrish, J. W., M.E. Shanahan, P. G. Schnitzer, P. Lanier, J. L. Daniels, and S. W. Marshall. “Quantifying Sources of Bias in Longitudinal Data Linkage Studies of Child Abuse and Neglect: Measuring Impact of Outcome Specification, Linkage Error, and Partial Cohort Follow-Up.” Injury Epidemiology, vol. 4, no. 1, 2017, p. 23.

R Development Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Vienna.

Sariyar, M. and Borg, A. “The RecordLinkage Package: Detecting Errors in Data” *The R Journal* (2010) 2:2, pages 61-67. https://journal.r-project.org/archive/2010/RJ-2010-017/index.html

September 2022

OPRE Report #2022-255

Project Officers: Christine Fortunato and Jenessa Malin

Project Specialist: Hanna Katz
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Project Director: Matt Stagner
Mathematica
111 East Wacker Drive, Suite 3000
Chicago, IL 60601-4303

Suggested citation: Emily Putnam-Hornstein, Eunhye Ahn, Rhema Vaithianathan, Jared Parrish, Robyn Husa, Matthew Walton, Claire Smither Wulsin, M. C. Bradley, and Beth Varley (2022). What Can We Learn About the Incidence of Foster Care Placement from Birth Records? Findings from the Cross Jurisdiction Model Replication Project Technical Appendix. OPRE Report #2022-2**55**, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services

This brief was funded by the Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, under contract number HHSP233201500035I/HHSP23337022T

This report and other reports sponsored by the Office of Planning, Research, and Evaluation are available at **www.acf.hhs.gov/opre**.

Connect with OPRE