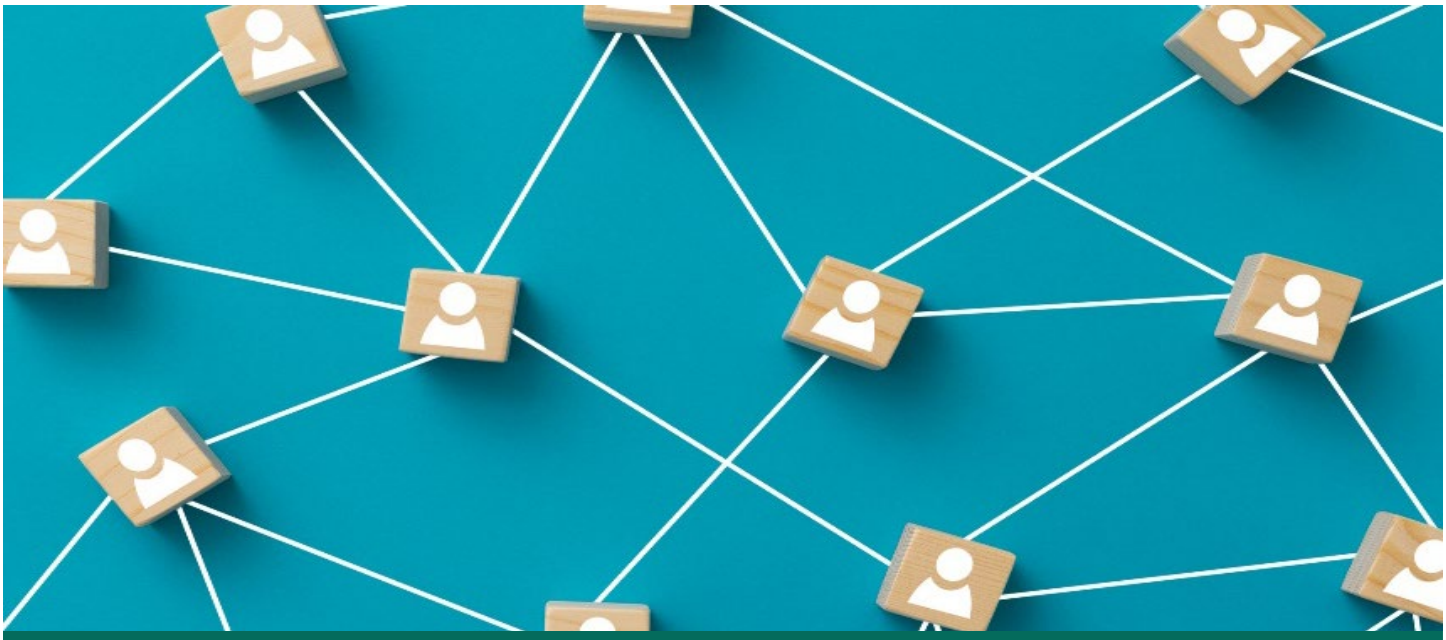
Mathematica
Progress Together



Methods to Estimate the Community Incidence of Child Maltreatment

November 2021

Emily Putnam-Hornstein, John Prindle, Stephanie Cuccaro-Alamin, Wendy Wiegmann, Joe Magruder, Daniel Webster, Eunhye Ahn, and Regan Foust

Submitted to:

Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
200 Independence Avenue, S.W.
Washington, D.C. 20201
Attention: Jenessa Malin and Christine Fortunato
Project Officers
Contract number: HHSP233201500035I
OPRE report number: 2021-253

Submitted by:

Mathematica
111 East Wacker Drive, Suite 3000
Chicago, IL 60601-4303
Phone: (312) 994-1002
Fax: (312) 994-1003
Matt Stagner, Project Director

This page has been left blank for double-sided copying.

Introduction

This appendix is an overview of the CDN/CCWIP project, which was conducted as part of the CMI Data Linkages work. The site team authored the appendix, although the Mathematica team worked with the site to ensure consistency in information, level of detail, and presentation across sites.

Overview

The site team used California administrative data to understand the extent to which methods and methodological decisions affected estimates of the incidence of child maltreatment. The goal was to develop a methodology that could be generalized to other states and then used in conjunction with or in place of the National Incidence Study (NIS).

NIS estimates include estimates of the number of abused and neglected children who do and do not come to the attention of child protective services (CPS). The NIS is, however, expensive to conduct due to the complexities of a nationally representative sampling approach and the nature of primary data collection; consequently, its estimates are outdated. It has been shown to suffer from problems in terms of lacking the precision and statistical power needed to assess critical group differences in maltreatment (for example, by race or ethnicity). (Drake and Jonson-Reid 2011)

Improving the collection, management, accessibility, and integration of administrative records positions these data as an increasingly important source of information for research, evaluation, and policy analysis (Putnam-Hornstein, Needell, & Rhodes 2013). In the context of a population-level understanding of maltreatment exposure, linked administrative data have the potential to be prospectively leveraged as an alternative or complement to the NIS, generating more cost-effective, timely, local, and (potentially) more accurate estimates of the victims of child abuse and neglect.

Partnership history

The site team included staff affiliated with the Children's Data Network (CDN) and California Child Welfare Indicators Project (CCWIP). See Table B.1. Data were already in hand based on the long-standing relationships between CDN, CCWIP, and California state agencies.

CDN is a data and research collaborative focused on the linkage and analysis of administrative records. In partnership with public agencies, philanthropic funders such as First 5 LA, affiliated researchers, and community stakeholders, CDN seeks to generate knowledge and advance evidence-rich policies that will improve the health, safety, and well-being of the children of California. CDN maintains data use agreements with numerous agencies that give permission to link cross-sector data and configure them longitudinally. The population-based, cross-sector data can be leveraged to develop applied and actionable research, support cost-effective program evaluations, and address policy-relevant questions.

CCWIP is a long-standing university/agency data partnership between CDSS and the University of California at Berkeley (UCB), supported through funding from CDSS and the Conrad N. Hilton Foundation. The project, housed in the UCB's School of Social Welfare, gives policymakers, child welfare workers, researchers, and the public direct access to customizable information on California's entire child welfare system.

CCWIP and the CDN have been data and research partners since the CDN's inception in 2013 and are well positioned to collaborate on this project. CDN maintains a formal data and research collaboration

with CCWIP, working closely with CCWIP researchers to provide technical support to state and county child welfare agencies. CCWIP is included in CDN's memorandum of understanding/contract with the California Department of Social Services (CDSS); likewise, CDN is named in CCWIP's agreement with CDSS.

Table B.1. Staff involved in the project

Name	Title	Affiliation	Role
Emily Putnam-Hornstein, Ph.D.	CDN principal investigator	USC	Oversee all aspects of the project, including design, analysis, and reporting. Serve as primary point of contact with state agency and CMI Data Linkages project.
Regan Foust, Ph.D.	CDN research scientist, co-investigator	USC	Project management, results translation/dissemination/communication, and manuscript preparation.
John Prindle, Ph.D.	CDN co-investigator	USC	Lead cross-sectional and longitudinal analytic strategies.
Daniel Webster, Ph.D.	CCWIP principal investigator	UC-Berkeley	Facilitate extract. Provide technical assistance for CWS/CMS records. Help with deliverables and manuscript preparation (co-authorship).
Stephanie Cuccaro-Alamin, Ph.D.	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).
Wendy Weigmann, Ph.D	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).
Joe Magruder, Ph.D.	CCWIP analyst	UC-Berkeley	Facilitate extract. Provide technical assistance and analytic support. Help with deliverables and manuscript preparation (co-authorship).

Source: Project plan.

Note: CDN = Children's Data Network; CCWIP = California Child Welfare Indicators Project; USC = University of Southern California; UC-Berkeley = University of California at Berkeley.

Background

The project developed a framework for an administrative-record–based methodology to estimate the number of children who are victims of abuse or neglect. The project, designed and tested using data from California, had the goal of developing a methodology that could be generalized to other states, and could produce estimates inclusive of both children who are identified as victims of abuse or neglect in a given year and those who are victims, but do not come to the attention of CPS.

It is important to note that the objective of this project was not to produce a single set of estimates held out as the "true" rate of abuse and neglect in a given community. Instead, the goal was to develop a number of upper- and lower-bound maltreatment estimates based on a range of assumptions and different methodological approaches. To document the potential to use administrative records to develop such estimates, we drew on information available through California vital birth, child protection, and death records. Thanks to the existing research infrastructure available at CDN and CCWIP, each of these data sources had already been cleaned, standardized, geocoded, and probabilistically linked using an algorithm customized to California's data.

Using records from 1998 through 2018, along with two different analytic approaches to structuring records (that is, cross-sectional and longitudinal), we exploited variability in the likelihood of CPS involvement and substantiation across:

- Counties with different policy and practice environments
- Child and family demographics, as measured for the population overall using vital birth records
- Time (that is, annual and lifetime involvement with CPS)
- Maltreatment type (that is, emotional, physical, and sexual abuse; neglect)
- Referral patterns before and after substantiation events

The observed variation allows the development of reasonable population-based approximations of child maltreatment victimization, which were compared with official substantiation rates.

Research Questions

This project has two main strategies, described below.

Cross-sectional strategy

How do estimates of victimization of child maltreatment vary based on (a) number of years in an estimate window (for example, only in focal year [2015] or up to three years on either side of focal year) or (b) county-level variation in how likely victims are to be identified?

We used California Child Protective Service records from 2012 through 2018 to identify a cross-section of children, across counties and demographics, identified as substantiated victims of maltreatment in 2015. We identified all other contacts (for example, no referrals, referrals to the CPS hotline that were screened out without investigation, referrals that were investigated but unfounded, and referrals that were substantiated) during the three years before the focal substantiation for a given child victim and the three years following substantiation. We estimated county-level variation in the annual/cross-sectional substantiation rates (conditioned on child-level characteristics gleaned from birth and CPS data) and use that county variability to extrapolate a range of estimated maltreatment rates.

Longitudinal strategy

How do annual incidence and cumulative prevalence rates differ by county and demographic characteristics at birth for children born in California in 1999? To what extent does earlier involvement with the child welfare agency predict substantiated child maltreatment?

We organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative childhood risk of abuse and neglect in California. With these data, we examined differences in annual incidence and cumulative prevalence rates by county and by demographic characteristics at birth.

Data

Sources

Three data sets were linked for this project (Table B.2).

Table B.2. Data sources

Name	Years	Geography covered	Source	Measures
Child protection (CWS/CMS)	1999–2018	California (statewide)	California Department of Social Services	Child protection encounters (referrals, substantiations) across time and counties
Vital birth records	1999	California (statewide)	California Department of Public Health	Birth characteristics
Vital death records	1999–2017	California (statewide)	California Department of Public Health	Deaths

Source: Project plan, interim, and final memo.

Note: CDN and CCWIP had access to all three data sets before the project began.

Linking process

CDN links and analyzes birth, child welfare, and death records (among other data sources) under approved state and university IRB protocols. The CDN uses unique identifiers created from linked birth, child welfare, and death records using previously published machine learning methods (see E. Putnam-Hornstein et al., 2020 for a detailed description). Briefly, linkages are developed using probabilistic matching methodologies that incorporate identifying information including names and dates of birth of both children and parents. ChoiceMaker, the probabilistic record linkage software that CDN uses, is based on a machine learning technique called Maximum Entropy. ChoiceMaker is based on Clues, which are Boolean tests of similarly between fields in a record pair. Each clue is assigned a weight, which is learned through machine learning on manually reviewed training record pairs. The probability of a match is output by ChoiceMaker, and all pairs above a certain threshold are "matches" and included in the de-identified analytic data set. Those pairs with a value less than a second threshold are "differs" and not include in the de-identified analytic data set. Pairs with values between the two thresholds, called "holds," are held, reviewed, and may be added to the de-identified analytic data set. The de-identified data set with unique identifiers was used for analysis by only identified CDN and CCWIP researchers.

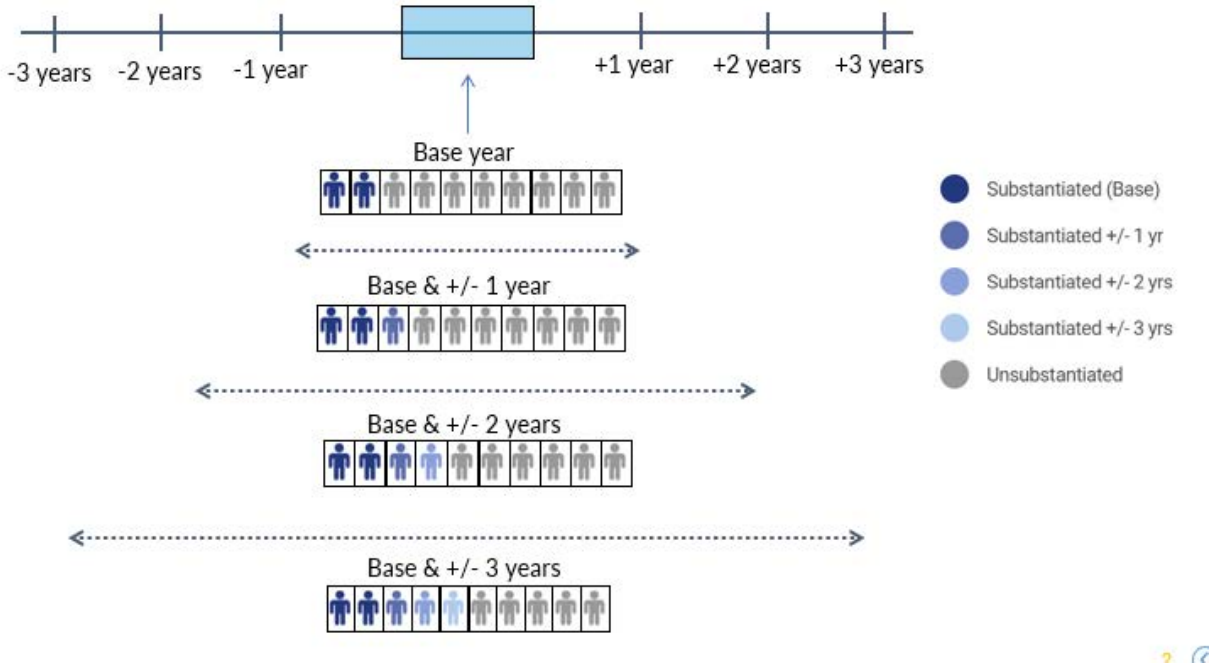
The team subjected 1,000 record pairs that represented complex clue patterns (that is, they were more difficult to categorize as a match or differ) to clerical review. We found that our model achieved a match recall rate of 92.5 percent, correctly classifying 92.5 percent of all "true" matches in the sample as matches. Meanwhile, false positive rates (pairs incorrectly identified as matches of out of all pairs) and false negative rates (pairs that were true matches but were incorrectly identified as holds or differs) were very low at 1.84 percent and 1.20 percent, respectively. This evaluation, training, and refinement process is ongoing; we are continually integrating new records, assessing match quality, and feeding that information back into our model to optimize accuracy.

Methods

Cross-sectional strategy

For our initial exploratory analysis, we used California CPS records from 2010 through 2018 to identify, for each year, a cross-section of children across counties and demographics who were referred because of alleged maltreatment. For each annual referral cohort, we then identified all other contacts (that is, no referrals; referrals to the CPS hotline that were screened out without investigation; referrals that were investigated but unfounded; referrals that were substantiated) during the three years before the focal substantiation for a given child victim and the three years after substantiation (see Figure C.1).

Figure B.1. Methodology for cross-sectional estimation of cumulative substantiation rates



We calculated annual base estimates of maltreatment victimization by examining the proportion of referred children with a substantiation during the year. We then documented how these base estimates changed if we *included children who had been referred but were not substantiated as a victim* in the specific year, but had been a substantiated victim in the year immediately prior, two years prior, and so forth. Similarly, we examined this for the years going forward with substantiations at one, two and three years following the base year. Using these data, we calculated a series of revised cumulative victimization rates that takes into account both children who were substantiated victims in a base year, as well as those who were referred to CPS but were substantiated as a victim within three years on either side of the base year. In addition, we explored maltreatment victimization rate estimates by allegation type (emotional, physical, and sexual abuse, neglect) and by geography, as well as interactions between allegation type and geography.

Longitudinal strategy

We also organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative risk of childhood abuse and neglect in California, providing a longitudinal record of system

interactions from birth (1999) to age 18 (through 2017). With these data, we examined differences in annual incidence and cumulative prevalence rates by county and demographic characteristics at birth. Further work to model probability of exposure of abuse or neglect symptoms to mandated reporters was an opportunity to assess a range of incidence rate estimates. In addition, the longitudinal approach allowed us to examine "persistence," which would identify children whose risk for long-term negative outcomes is heightened because of their repeated involvement with the child welfare system. We regard this longitudinal strategy as a method for exploring victimization estimates based on assumptions that will serve as an extension of the cross-sectional strategy, but with additional opportunities to exploit annual vs. lifetime data.

Findings

Cross-sectional findings

In 2014, 82,388 children were substantiated as victims in California. If we add to that number the children who were referred to CPS but were substantiated as a victim within three years on either side of 2014, there was a 72.5 percent increase in the number of children classified as victims over the base estimate (142,157; Figure B.2). Applying the percent changes from the three-year cumulative analysis to the 2015 National Child Abuse and Neglect Data System (NCANDS) child victim total (n = 683,487), we found that an additional 524,614 children would be classified as victims, for a total of 1.2 million child victims in 2015 (Figure B.3).

Figure B.2. 2013 - 2015 Children with referrals: cumulative substantiation rate at -/+ 1, 2, and 3 years

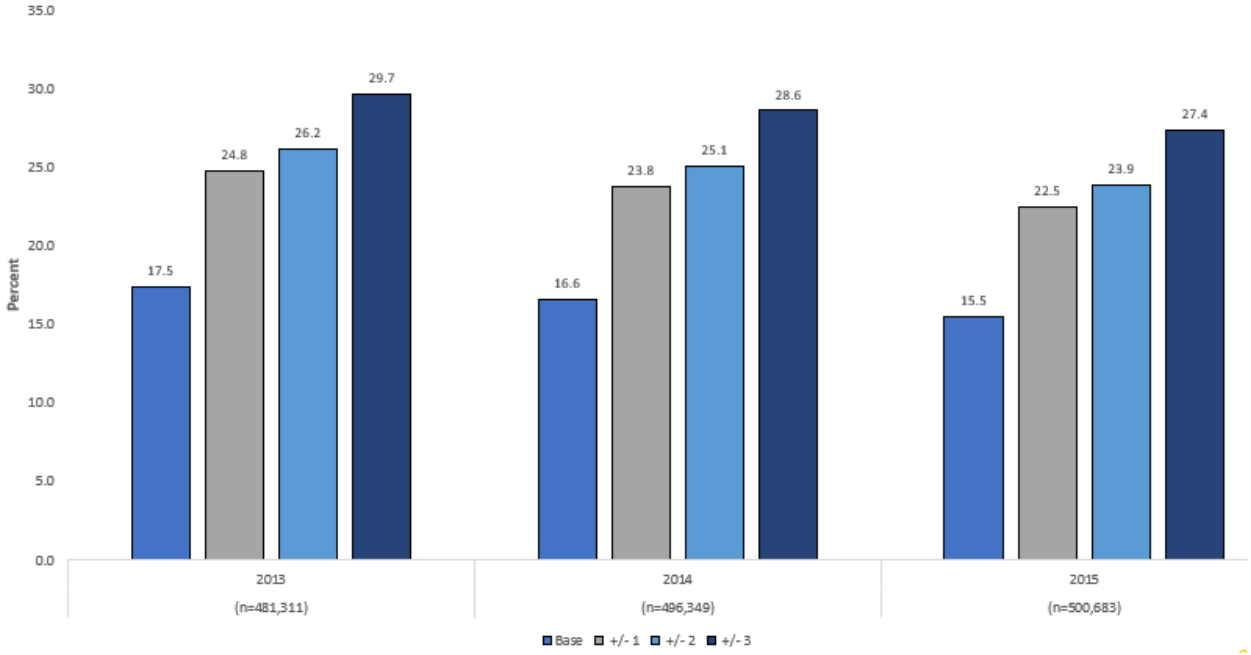
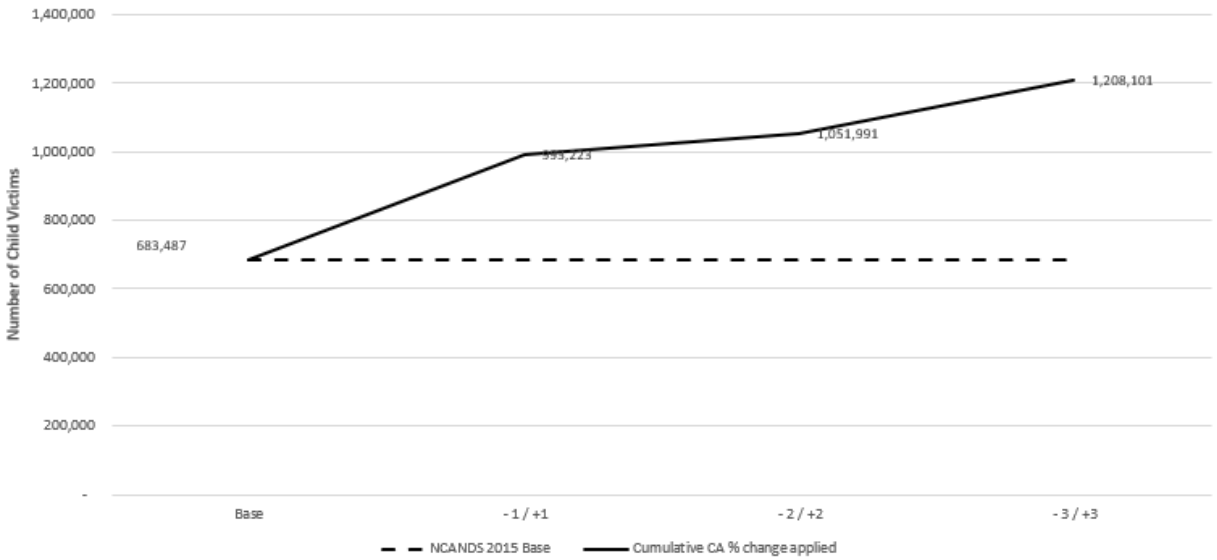


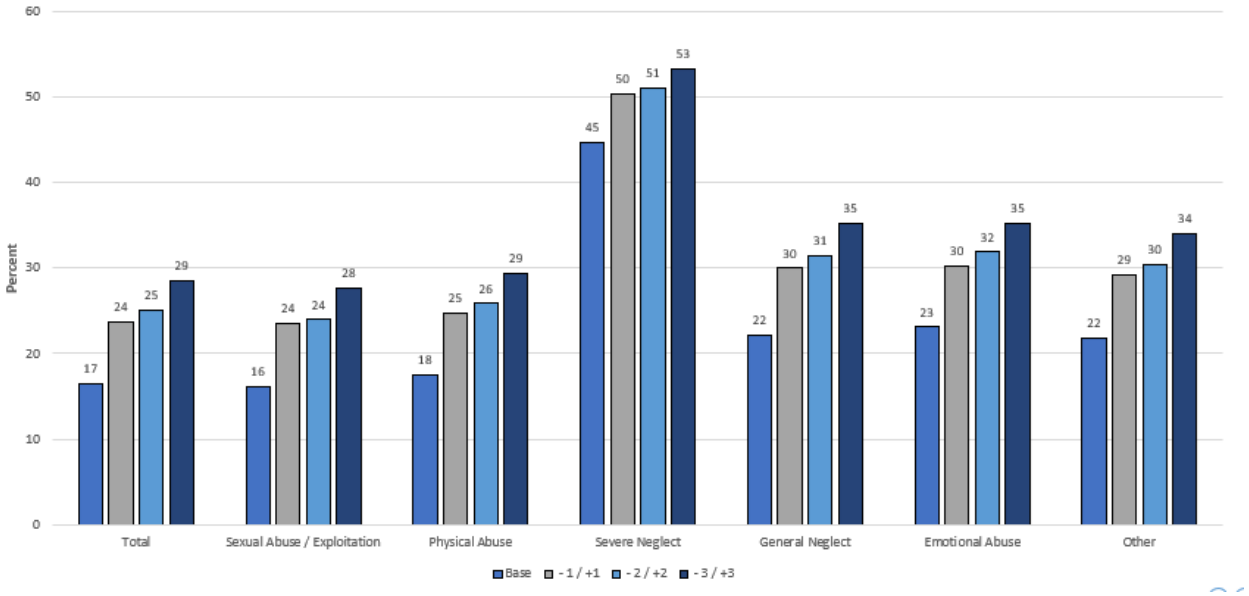
Figure B.3. 2015 U.S. child victim estimates derived using California's cumulative % change at -/+ 1, 2, and 3 years



In addition, if the 3-year cumulative window were applied on top of the 683,487 child victims in 2015 (i.e., if we included both children who were substantiated victims in a base year and those referred to CPS but substantiated as a victim within three years on either side of that base year) an additional 524,614 children would have also been counted as victims, totaling 1.2 million child victims.

We also found variation in base and cumulative substantiation rates (+/ - 1, 2, 3 years) by county and by allegation type (Figure B.4).

Figure B.4. 2014 Children with referrals cumulative substantiation rate at -/+ 1, 2, & 3 years by allegation type



Longitudinal findings

We organized CPS records longitudinally for a cohort of children born in 1999 to estimate the cumulative childhood risk of abuse and neglect in California through age 18 (through 2017). Cumulative prevalence rates by county and demographic characteristics at birth are in Table B.3 (see Putnam-Hornstein et al. 2021). In California, 519,248 children were born in 1999. Mirroring the secondary sex ratio nationally, the cohort was defined by a slight male majority. Consistent with California demographics, a plurality was born to Hispanic mothers. Slightly more than 1 in 10 children were born to teen mothers and approximately 30 percent of children were born to mothers with less than a high school degree. Overall, 92.8 percent of children had paternity established at birth.

Cumulatively, 29.9 percent of children in the cohort were investigated for alleged maltreatment and 12.1 percent were substantiated as victims of abuse or neglect. Between birth and age 18, 4.4 percent of children in the cohort experienced a removal and placement in foster care at least once. 1.1 percent experienced a legal termination of parental rights. Although the magnitude of the relative group differences varied across levels of CPS involvement, demographic patterns were directionally consistent. The cumulative proportions of Black and Native American children who had child protection encounters were significantly higher than those of other children. In the cohort overall, approximately half (51.3%) of Black and Native American (55.4%) children were investigated for alleged maltreatment before age 18. Both groups experienced termination of parental rights at twice the rate of White children in the cohort.

The likelihood of child protection involvement exhibited an inverse relationship to both maternal age at birth and maternal education levels. The rate of termination of parental rights was twice as great for children born to teen mothers versus children born to mothers 25 and older (IRR: 2.52, 95% CI: 2.36, 2.69). Likewise, children born to mothers with less than a high school degree experienced a termination of parental right at 2.6 times the rate of those who had completed high school (IRR: 2.60; 95% CI: 2.47, 2.74). Receipt of public health insurance and missing paternity were also strongly related to all levels of CPS involvement. Among children whose births were covered by public insurance, the rate of terminations of parental rights as 6 times that of children in the cohort covered by private insurance (IRR: 6.14; 95% CI 5.74, 6.57). Although only 1 in 14 children in California was born without paternity established (n = 37,513), parental rights were terminated for nearly 6 percent (n = 2,000) of those children. Children with missing paternity experienced termination of parental rights at seven times the rate of those with established paternity (IRR: 7.77; 95% CI: 7.37, 8.19).

Table B.3. Demographic characteristics of 1999 birth cohort in California (incidence rate per 1,000 children)

Characteristic	1999 Birth Cohort		Referred for Alleged Maltreatment		Substantiated as Victim of Abuse or Neglect		Removed and Placed in Foster Care		Termination of Parental Rights	
	N	%	Cumul. %	RR (99%CI)	Cumul. %	RR (95% CI)	Cumul.%	RR (95% CI)	Cumul. %	RR (95% CI)
Total	519,248	100.0	29.9	-	12.1	- -	4.4	- -	1.1	- -
Child sex										
Female	253,734	48.9	30.7	1.05 (1.05, 1.06)	12.5	1.07 (1.05, 1.09)	4.4	1.01 (0.99, 1.04)	1.1	1.06 (1.01, 1.12)
Male (ref.)	265,511	51.1	29.1	1.00	11.8	1.00	4.3	1.00	1.1	1.00
Maternal race/ethnicity										
Black	34,156	6.6	51.3	1.97 (1.95, 2.00)	24.0	2.22 (2.13, 2.23)	12.3	2.97 (2.86, 3.08)	3.2	2.47 (2.30, 2.65)
Native American	2,532	0.5	55.4	2.13 (2.06, 2.21)	30.0	2.83 (2.65, 3.02)	14.5	3.49 (3.17, 3.85)	3.9	2.97 (2.44, 3.63)
Latinx / Hispanic	252,691	48.7	32.7	1.26 (1.25, 1.27)	12.7	1.13 (1.14, 1.18)	4.0	0.96 (0.93, 0.99)	0.8	0.68 (0.61, 0.69)
Asian / PI	55,422	10.7	15.1	0.58 (0.57, 0.59)	4.9	0.46 (0.42, 0.46)	1.2	0.29 (0.27, 0.31)	0.3	0.20 (0.16, 0.23)
White (ref.)	172,188	33.2	26.0	1.00	11.0	1.0	4.2	1.00	1.3	1.00
Maternal age at birth										
< 20 years	57,693	11.1	50.4	2.14 (2.12, 2.16)	22.8	2.66 (2.60, 2.71)	9.2	2.99 (2.98, 3.08)	2.1	2.52 (2.36, 2.69)
20 – 24 years	120,519	23.2	38.0	1.61 (1.60, 1.63)	15.8	1.81 (1.78, 1.84)	5.7	1.85 (1.80, 1.91)	1.3	1.57 (1.48, 1.67)
25+ years (ref.)	340,974	65.7	23.6	1.00	9.0	1.00	3.1	1.00	0.8	1.00
Maternal education										
Less than high school	155,364	29.9	40.0	1.57 (1.56, 1.98)	17.9	1.93 (1.90, 1.96)	7.1	2.30 (2.25, 2.36)	1.9	2.60 (2.47, 2.74)
High school degree (ref.)	356,358	68.6	25.4	1.00	0.00	1.00	3.1	1.00	0.7	1.00
Birth payment method										
Public	218,643	42.1	41.7	1.96 (1.95, 1.98)	18.9	2.79 (2.75, 2.84)	7.7	4.11 (3.99, 4.24)	2.1	6.14 (5.74, 6.57)
Private (ref.)	298,178	57.4	21.2	1.00	7.1	1.00	1.9	1.00	0.3	1.00
Paternity established										
Missing	37,513	7.2	52.8	1.88 (1.86, 1.90)	28.4	2.79 (2.74, 2.85)	15.7	4.53 (4.40, 4.65)	5.8	7.77(7.37, 8.19)
Established (ref.)	481,735	92.8	28.1	1.00	10.9	1.00	3.5	1.00	0.7	1.00

Source: Project documents.

Note: Cumul. = cumulative. CI = confidence interval. RR = risk ratio. Rates of missing sociodemographic variables were low for all groups: child sex (0.03%), maternal race/ethnicity (0.05%), maternal age (0.01%), birth payment method (0.47%), and maternal education (1.45%).

Further details on the longitudinal analyses and findings can be found in:

1. Prindle, J., R. Foust, and E. Putnam-Hornstein. "Maltreatment Type Classifications and Transitions During Childhood for a California Birth Cohort." *Child Maltreatment,* 2021. Available at https://doi.org/10.1177/10775595211006784.
2. Putnam-Hornstein, E., E. Ahn, J.J. Prindle, J. Magruder, D. Webster, and C. Wildeman. "Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999 – 2017." *American Journal of Public Health*, 2021, pp. e1-e7. Available at https://doi.org/10.2105/AJPH.2021.306214.

Next steps

The methodology developed in the cross-sectional work holds the promise of helping other jurisdictions develop alternative estimates of child maltreatment incidence through simple extensions of the window in which substantiation is counted. Findings from California suggest that many children who are reported in any given year may be experiencing conditions that at another proximate point in time the system classifies as official maltreatment. Future work should explore the extent to which these same adjustments yield similar increases in maltreatment incident estimates in other jurisdictions.

Lessons learned about administrative data linkage practices related to examining the incidence and risk of child maltreatment

We learned much from the process, and from the people we were able to collaborate with in framing these findings. Thanks in large part to the grant support, we had the time and space to apply new methodologies to and explore new conceptualizations of maltreatment. The protected time to collaborate, discuss, and present analyses that were in progress to an expanded group of researchers and stakeholders broadened our horizons and informed the development of our methodology to produce a range of community estimates of the incidence of abuse or neglect. In addition, our approach to using a base cohort of children with reports of maltreatment in any given year and then making adjustments to the substantiated victimization rate for that year by looking at how many of those children were classified as victims in the recent past, or will be in the near future, increases the number of child maltreatment victims in the United States by approximately 75 percent. Although academic researchers and others are correct to question the value of substantiation given the uncertainties about what contributes to that classification, we still believe this approach is a useful way to think about an alternative approach to estimating maltreatment incidence. It is also important to note that state comparisons and other demographic statistics reported by the Children's Bureau (i.e., annual Child Maltreatment report) continue to be strongly oriented around substantiation rather than children who are reported, meaning that this designation continues to drive policy decisions. In terms of recommendations or suggestions for other jurisdictions, gaining access to vital birth records would be useful for replication and characterization of maltreated children.

REFERENCES

Drake, B. and M. Jonson-Reid. "NIS interpretations: Race and the National Incidence Studies of Child Abuse and Neglect." *Children and Youth Services Review*, vol. 33, no. 1, 2011. pp. 16 – 20.

Putnam-Hornstein, E., E. Ahn, J. Prindle, J. Magruder, D. Webster, and C. Wildeman. "Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999 – 2017." *American Journal of Public Health*, 2021, pp. e1 – e7. Available at https://doi.org/10.2105./AJPH.2021.306214.

Putnam-Hornstein, E., M. Ghaly, and M. Wilkening. "Integrating Data to Advance Research: Operations, and Client-Centered Services in California." *Health Affairs*, vol. 39, 2020, pp. 655 – 661. Available at https://doi.org/10.1377/hlthaff.2019.01752.

Putnam-Hornstein, E., B. Needell, and A. E. Rhodes. "Understanding risk and protective factors for child maltreatment: the value of integrated, population-based data." *Child Abuse & Neglect*, vol. 37, no. 2-3, 2013. pp. 116 – 119.