CALIFORNIA STRONG START INDEX DOCUMENTATION

Children's Data Network

REGAN FOUST, PHD
JOHN PRINDLE, PHD
ANDREA LANE EASTMAN, PHD
WILLIAM C. DAWSON, MSW
MICHAEL MITCHELL, PHD
HUY TRAN NGHIEM, MS
EMILY PUTNAM-HORNSTEIN, PHD

CALIFORNIA STRONG START INDEX DOCUMENTATION

ACKNOWLEDGEMENTS

This project was supported through a grant from the Heising-Simons Foundation. We wish to acknowledge our collaborating colleagues at the USC Children's Data Network, the California Child Welfare Indicators Project, First 5 Association, and First 5 County Commissions, as well as our state and county data partners. Additionally, none of this work would be made possible without our other core infrastructure supporters: The Heising-Simons Foundation, First 5 LA, and the Conrad N. Hilton Foundation. Their steadfast partnership, keen insights, and ongoing commitment to the children and families of California are critical to the development and continued iteration of the California Strong Start Index, and to all of our work here at the Children's Data Network.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Ideally, every child would be healthy, growing, and thriving in a strong family, and supported by a safe and nurturing community. The reality is, however, that the human, social, and material assets present at birth vary widely across California's nearly 500,000 infants born each year. And this variation is not inconsequential. A large, and growing, body of literature affirms the importance of early childhood experiences in influencing adolescent and adult behavior. The human, social, and material assets present at birth lay the foundation for the emergence of protective factors during childhood that we know are tied to good outcomes and resilience throughout the life course.[1]

Information universally registered at birth can be used to document assets available to each California newborn. Specifically, information regarding infant health and circumstances surrounding the birth (e.g., birthweight, presence of birth abnormalities), family socioeconomic status (e.g., type of insurance), maternal health behaviors and access to services (e.g., timing of initiation of prenatal care), and the age, education, and nativity of both parents (if paternity is established), all provide insight into the conditions into which individual children are born. Of course, assets and conditions at birth are not destiny. But thoughtful supports and services may be required to ensure that children with fewer assets find themselves on equal footing with their peers in California. Monitoring the distribution of assets among newborns in different communities can help ensure our investments are intentional and equitable.

The California Strong Start Index uses data that already exist for children and families to summarize, in a standardized way, the conditions into which children are born. It comprises a total of 12 variables that fall into four domains. A birth asset score is calculated by simply counting the number of assets present (0-12).

[1]The Five Protective Factors include: Parental Resilience, Social Connections, Concrete Support in Times of Need, Knowledge of Parenting and Child Development, and Social and Emotional Competence of Children (CSSP, n.d.)

TABLE 1. CALIFORNIA STRONG START INDEX INDICATORS

FAMILY

- Legal parentage established at birth
- Born to non-teen parents
- Born to parents with at least a HS degree

HEALTH

- Healthy birthweight
- Absence of congenital anomalies, abnormalities, or complications at birth
- Born to parents with at least a High School degree

SERVICE

- Access to and receipt of timely prenatal care
- Receipt of nutritional services (WIC) if eligible
- Hospital with high percentage of births with timely prenatal care

FINANCIAL

- Ability to afford and access healthcare
- Born to a parent with a college degree
- Born to parents with employment history

These asset indicators are universally measured at birth with strong validity, and set the stage for the emergence of protective factors and healthy development throughout the life course. A review of literature and external validity checks confirm that the Strong Start Index adds unique insight into the conditions into which children are born in California and its scores are related to at least two important indicators of child health and well-being (i.e., child protection involvement and death).

The Strong Start Index allows us to characterize the number of assets children have at birth, including how California communities vary in the distribution of children at different asset levels.

Specifically, the Strong Start Index:

- Facilitates the identification of communities in which children have fewer assets at birth and where additional services and supports may be important to promote equity
- Characterizes how asset levels of children in different communities have changed over time, highlighting where disparities persist

And it has the potential to:

- Act as a standardized and cost-effective anchor for community needs assessments
- Guide a more strategic stewardship of public dollars, with increased accountability
- Promote the adoption of a common language across communities, commissions, and other stakeholder groups for conceptualizing and discussing early childhood investments.

The goal of this document is to describe the development of the Strong Start Index, and specifically the landscape of indices that influenced our thinking and the theoretical and methodological rationale that underpins the index, as it is currently constructed. Please visit **www.strongstartindex.org** to explore the data, and to learn more about how communities are using it to facilitate equitable investment.

INTRODUCTION

Could information from birth records, thoughtfully assembled and simply scored, be used to help communities more efficiently and equitably allocate resources?

This question emerged during a conversation among colleagues from the Children's Data Network, First 5 Association, and Heising-Simons Foundation. We had gathered to discuss the challenges County Commissions, service providers, funders, legislators, and advocates face when assessing the service needs of families with young children.

Due, in large part, to a dearth of recent, publically available data on young children and families, we lamented that they have to conduct costly community needs assessments, use survey-based county estimates for neighborhood-level program planning, extrapolate historical data about adults into actionable information about current families with newborns, and base many decisions on a community's poverty level—there had to be a better way!

The idea for the California Strong Start Index was born.

Successful proof-of-concept analyses funded by First 5 LA and the Orange County Commission on Children and Families confirmed that administrative birth records could be used to not only **better characterize the changing demographics and birth outcomes of babies born in LA County,** but also help **Los Angeles County chart a course toward universal and targeted Home Visiting.** And First 5 colleagues immediately recognized the potential for an index comprised of information from administrative birth records to provide recent, specific, holistic, and asset-focused information about the children they were charged with supporting. Geocoded records could be aggregated and viewed at a state level, but also provide granular, local information to support equitable resource allocation. These child / family-level indicators could be flexibly overlaid with other community-level indicators or indices. These data could develop a much more complete picture of our state's children. And they could be easily updated for each new cohort of children born.

Given the potential to better characterize young children and families, streamline processes for stakeholders, and, ultimately, change the conversation around investments for children and families, we proceeded to explore the idea of the Strong Start Index. The following described our process and learnings.

SECTION 1: MODEL INDICES

First, we analyzed the landscape of existing indices, public datasets, and published analyses in order to confirm that we weren't reinventing the wheel. Finding no other indices that rely solely on information presented in vital birth records, we proceeded to review the landscape of indices with the potential to assist in the development of the Strong Start Index. The following section presents indices developed using similar approaches and goals. This compilation of indices inspired the presentation and visualization of Strong Start Index scores and two were used as benchmarks in our assessment of the Strong Start Index's external validity (See Section 3: Strong Start Index).

INDICES USING POPULATION DATA SETS

Of all indices identified, the following population-based indices are most similar in method and intended presentation to the Strong Start Index.[2] They acted as useful models for not only describing our methods, but also presenting / visualizing data (please see "Effective Visualizations and Summaries" below). Use of population data and presentation at a high level of geographic detail are elements that the Strong Start Index shares with these U.K. indices. The level of precision possible with these approaches supports informed analysis and decision-making.

INDICES USING POPULATION DATA SETS

INDEX	FOCUS	FRAME	GEOGRAPHY	PRESENTATION *(smallest geo unit)*
Welsh Child Index for Multiple Deprivation (WCIMD)	Deprivation	Deficit-based	Wales	Lower Super Output Areas (LSOA), average 1,600 people
The Scottish Index of Multiple Deprivation (SIMD)	Deprivation	Deficit-based	Scotland	Data Zones, average 760 people
English Index of Multiple Deprivation (IMD)	Deprivation	Deficit-based	England	Lower-layer Super Output Areas, average 1,500 people
Northern Ireland Multiple Deprivation Measure (NIMDM)	Deprivation	Deficit-based	Northern Ireland	Super Output Areas (SOAs), average 2,100 people

[2] This list will be updated as new indices become publically available.

These indices use relative ranking of geographic units to report their findings. In this regard, they differ from the approach of the Strong Start Index, which provides an absolute (0 through 12) score, rather than ranking counties, census tract, or other geographies. Absolute scores allow a community or county to examine their progress over time in a straightforward way, not affected by changes in other areas. An additional difference is that they are deficit-based.

WELSH CHILD INDEX FOR MULTIPLE DEPRIVATION (WCIMD)

https://gov.wales/statistics-and-research/welsh-index-multiple-deprivation/?lang=en
From: **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0113592**

The Welsh Child Index of Multiple Deprivation (WCIMD) measures concentration of deprivation for small areas in Wales, with a specific focus on children. It is based on the Welsh Index of Multiple Deprivation (WIMD). The smallest unit of analysis is Lower Super Output Areas (LSOAs). Based on the average population (1,600), these correspond roughly to U.S. census block groups.[3] The index ranks areas based on seven distinct domains of deprivation: Income, Health, Education, Access to Services, Community Safety, Physical Environment, and Housing. Each domain is a composite variable and ranked 0-100 (highest level of deprivation). Within the online interface, users can rank order each LSOA from least to most deprived, compare the seven distinct deprivation domains in each area, and compare the ratio of LSOAs in local governments that are exceedingly deprived. The Index, however, cannot be used to compare ranks over time, measure how much more deprived one area is than another, or compare deprivation with other UK countries that use different indices. Also, given its deficit-focus, it cannot measure affluence.

THE SCOTTISH INDEX OF MULTIPLE DEPRIVATION (SIMD)

http://www.gov.scot/Topics/Statistics/SIMD
Please see: **http://simd.scot/2016/#/simd2016/BTTTFTT/9/-4.0000/55.9000/**

The Scottish Index of Multiple Deprivation (SIMD) measures concentrations of deprivation for small areas (called data zones) in Scotland. In each zone there are between 500 and 1,000 residents. There are 6,505 data zones total. SIMD is based on 38 indicators from seven domains: Income, employment, health, education, access, housing and crime. SIMD can be used to determine areas where residents may experience multiple deprivation, determine areas of greater need, compare small areas across Scotland, and provide statistical profiles of individual data zones. It also offers a dynamic tool consisting of a map with graphical tables

[3] https://current360.com/research-101-census-tracts-vs-census-block-groups/

of data zone descriptive statistics and SIMD domain rankings, visible when you mouse over the zone.

ENGLISH INDICES OF DEPRIVATION

https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
From: *Statistical release - main findings*, p.2

The English Indices of Deprivation 2015 measures deprivation for small areas, called Lower-layer Super Output Areas (LSOAs) in England. There is a total of 32,844 areas. It is based on 37 indicators from seven domains of deprivation (i.e., income; employment; education, skills, and training; health and disability; crime; barriers to housing and services; living environment). In addition to the seven domains, there are two supplementary indices: Income deprivation affecting children and income deprivation affecting older people. The index ranks areas from most deprived to least deprived. The areas are divided into 10 deciles based on their deprivation rank. The index is used to compare small areas across England, determine the most deprived area, examine how areas rank on domains, and compare the small areas to local authorities. Mapping tools are available to examine the relative deprivation for small areas.

NORTHERN IRELAND MULTIPLE DEPRIVATION MEASURE 2017 (NIMDM2017)

https://www.nisra.gov.uk/statistics/deprivation/northern-ireland-multiple-deprivation-measure-2017-nimdm2017

[1] See indicator description document at: https://www.nisra.gov.uk/publications/nimdm17-results

[2] See page 6 of the NIMDM 2017 Blueprint document: https://www.nisra.gov.uk/publications/nimdm17-consultation-results

From: *NIMDM2017 Report*, p.2

Northern Ireland Multiple Deprivation Measure 2017 (NIMDM 2017) identifies deprivation concentrations of small areas, called Super Output Areas (SOAs), across Northern Ireland. There is a total of 890 areas, averaging 2,100 people. The index uses 38 indicators, organized into seven domains, to identify different types of deprivation (i.e., income; employment; health and disability; education, skills, and training; access to services; living environment; crime and disorder). NIMDM 2017 is used to compare deprivation ranks between two or more areas, identify which area is most or least deprived, and examine how deprivation ranks changed over time.

INDICES USING SAMPLE DATA SETS

We identified three geographically stratified indices that were constructed using survey data gathered through sampling methods, rather than through census or population-based administrative data. The California Healthy Places Index, the Social Vulnerability Index, and Wisconsin's Risk and Protective Factors Related to Child Abuse and Neglect emerged as the most instructive for building the Strong Start Index.[4]

INDICES USING SAMPLE DATA SETS

INDEX	FOCUS	FRAME	GEOGRAPHY	PRESENTATION *(smallest geo unit)*
Healthy Places Index (HPI)	Health	Asset-based	California	Census tract
American Human Development Index (HDI)	Human Development	Asset-based	California	County
Social Vulnerability Index (SVI)	Social Vulnerability	Deficit-based	National, includes California	Census tract
Wisconsin's Risk and Protective Factors Related to Child Abuse and Neglect	Child Welfare	Deficit-based *(Uses risk and protective factors to determine risk)*	Wisconsin	County

HEALTHY PLACES INDEX (HPI)

[California]
https://healthyplacesindex.org/

The California Healthy Places Index (HPI) is an asset-based index that examines the social determinants of health associated with life expectancy. Each census tract is scored. The index can be used to compare each census tract across California, and group them into a single score to compare zip codes, project areas, and other geography areas. The index uses 25 indicators based on eight domains (i.e., economic; education; housing; health care access; neighborhood; clean environment; transportation; social factors). We used the HPI as a reference for an external validation of the Strong Start Index. Please see *External Validation in Section 3: Strong Start Index* for our examination of the correspondence of the Strong Start Index and HPI scores.

[4] This list will be updated as new indices and underlying datasets (e.g., Child Opportunity Index) become publically available.

AMERICAN HUMAN DEVELOPMENT INDEX (HDI)

[California]
http://www.measureofamerica.org/california2014-15/

The American Human Development Index (HDI) is an asset-based index that assesses, using a 10-point scale, how people are faring in terms of a long and healthy life, access to knowledge, and a decent standard of living at the California and county level. All three dimensions are given equal weight. The index is used to question national policy choices based on human development. The index does not show inequalities, poverty, human security, empowerment, etc. We used the HDI as a reference for an external validation of the Strong Start Index. Please see *External Validation in Section 3: Strong Start Index* for our examination of the correspondence of the Strong Start Index and HDI scores.

SOCIAL VULNERABILITY INDEX

[National, Includes California]
https://svi.cdc.gov/Index.html

The Social Vulnerability Index (SVI), a project of the CDC's Agency for Toxic Substances and Disease Registry, uses U.S. Census data to evaluate the social vulnerability of every Census tract. This information is used to determine the capacity of neighborhoods to respond to disaster. The SVI uses 14 American Community Survey variables to rank each tract and groups them into four domains: Socioeconomic, household composition and disability, minority status and language, and housing and transportation. Each tract receives a separate ranking for each of the four themes, as well as an overall ranking. The SVI is used to calculate the number of basic supplies and emergency personnel needed, plan evacuation strategies, and identify areas that need shelters and continued support after the disaster.

WISCONSIN'S RISK AND PROTECTIVE FACTORS RELATED TO CHILD ABUSE AND NEGLECT

https://preventionboard.wi.gov/Documents/WIWTBrief4Full.pdf
Maguire-Jack, K., Kibble, N., Cranley, M., & O'Connor, C. (2010). Risk and protective factors related to child abuse and neglect. *What it will take: Investing in Wisconsin's future by keeping kids safe today.* Madison, WI: Wisconsin Children's Trust Fund and Wisconsin Council on Children and Families.

Wisconsin's Risk and Protective Factors Related to Child Abuse and Neglect brief used county-level data to calculate risk levels by county and to identify prominent risk and protective factors. The risk factors were organized into four domains: parent characteristics, family situations, child characteristics, and economic circumstances. The risk factors include: parental substance abuse and mental health problems, domestic violence, single parenthood,

teen parenthood, low maternal education, low birth weight, child disability, child emotional or behavioral problems, poverty, and unemployment. The risk factors can be used to identify families who may need support to prevent child maltreatment.

EFFECTIVE VISUALIZATIONS, SUMMARIES, AND DATA TOOLS

The above-referenced indices inspired the clear and engaging presentation of the Strong Start Index on www.strongstartindex.org.

Other local, state, and national, data sources and tools identified as part of the Landscape Analysis are included in *Appendix A. Other Local, State, and National Data Sources and Tools*.

FINDINGS

Our survey of existing indices reinforced that simplicity and validity are key; indices need to be easy to understand and the scores need to relate to actual child health and well-being outcomes. We also appreciated the availability of recent data in the indices we reviewed and this further encouraged a focus on births. A focus on conditions observable at birth that are correlated with childhood outcomes would provide communities actionable data, with no follow-up required.

Our review of existing indices and their public presentation also reinforced the value of using population-based (i.e., universal), rather than survey (i.e., sample-based), data. When the use of population data is possible, it is easier to explain to users of the data and to connect to their community. Population data also tends to allow reporting to a more granular level and, as such, enhances a more localized understanding. The most effective visual presentations of indices we found enabled users to explore data at various levels of analysis—zooming in and out on a map, applying various boundary overlays—and such presentations included data aggregated at a detailed level and indicators reliably available across the entire area under consideration.

These guiding principles informed what indicators were ultimately included in the Index, as well as how they are 'counted' to produce an Index score.

SECTION 2: METHODS

Incorporating information gleaned from our analysis of the landscape of existing indices and public datasets, we constructed the Strong Start Index.

INDICATORS

Of the many fields recorded at birth, which could be considered key descriptors of the conditions into which children are born? Which are ultimately related to child outcomes? We consulted the literature to understand which indicators should be included.

Our literature review confirmed that certain fields recorded on vital birth records could be treated as 'assets' due to their relationship to child health and well-being outcomes. As such, an index comprising these indicators could help communities document the range of contexts into which children are born, supporting more strategic and equitable investment in services and programs. (Please refer to *Appendix B. Evidence for Inclusion in the California Strong Start Index, by Indicator*, for more information)

The results of our literature review resulted in a list of 12 indicators[5] constructed from fields available on vital birth records that would comprise the Strong Start Index (Table 2).

TABLE 2: STRONG START INDEX INDICATORS AND SCORING

FAMILY

- Legal parentage established at birth (+1)
  1=yes
  0=missing paternity
- Born to non-teen parents (+1)
  1=yes
  0=missing parent age or mom teen or dad teen or both parents teens
- Born to parents with at least a High School degree (+1)
  1=HS degree or more reported for both mom and dad
  0=<HS degree for at least one parent or missing education for at least one parent

[5] An initial version included the indicators: 'No Prenatal Cigarette Exposure' and 'Non-Negative APGAR Score,' but the results of an exploratory and confirmatory factor analysis (results available upon request) suggested replacement with other, more reliable, internally consistent, and intuitive health-related indicators: 'Absence of Congenital Anomalies, Abnormalities, or Complications at Birth' and 'Absence of Transmissible (Mother-to-Child) Infections.'

HEALTH

- Healthy birthweight (+1)
  1=birthweight >2500g
  0=missing or <2500g
- Absence of congenital anomalies, abnormalities, or complications at birth (+1)[6]
  1=No congenital anomalies, abnormalities, or complications
  0=Congenital anomalies, abnormalities, or complications
- Absence of transmissible (mother-to-child) infections (+1)
  1=No maternal infections present or treated during this pregnancy
  0=Maternal infections present or treated during this pregnancy

SERVICE

- Access to and receipt of timely prenatal care (+1)
  1=prenatal care began during first trimester
  0=prenatal care began after first trimester or not at all
- Receipt of nutritional services (WIC) if eligible (+1)
  1=private insurance and WIC or no WIC OR public insurance and WIC
  0=public insurance and no WIC
- Hospital with high percentage of births that received timely prenatal care (+1)
  1=birth in a hospital where the mean percentage of births with prenatal care beginning in the first trimester exceeded the state mean
  0=birth in other hospital

FINANCIAL

- Ability to afford and access healthcare (+1)
  1=US-born mother w/private health insurance OR non-US-born mother w/private or public insurance
  0=US-born mother w/public health insurance
- Born to a parent with a college degree (+1)
  1=yes
  0=both parents less than college degree or both parents missing education
- Born to parents with employment history (+1)
  1= work reported for both parents
  0=no work or missing work for at least one parent

[6] A child is scored as having a birth complication if he/she had any of the data recorded in the raw field of the birth record, which includes having epidural or induction. We plan to revisit our coding of this variable in future iterations.

DATASET

DATA SOURCE

Annual birth record files were obtained from California's Department of Public Health (CDPH) for the years spanning 1999-2016. These birth files reflect all live births in California during each calendar year and reflect the universe of children included in the development and production of the Strong Start Index.

Considerations
During our initial exploratory data phase, we considered supplementing the information collected through vital birth records with data collected at a community-level. We were concerned that it could complicate the calculation and interpretation of our straightforward index, and that adding other scores could make it more difficult to overlay our Strong Start scores with other community indices – an important next step – but decided to at least explore the idea. We assessed the potential value of including Healthy Places Index (HPI) total and subscale scores in predicting future child welfare involvement (our proxy outcome for child health and well-being) for a cohort of children born in California in 2007. Results indicated that supplementing the Strong Start Index with HPI total or subscale scores did not result in a meaningful improvement in predictive quality. See *Appendix C. Assessment of Inclusion of Healthy Places Index (HPI)* for more information and model fit statistics.

We also considered linking birth records to other administrative data that might provide a more direct measure of the dynamics in which we are interested (e.g., CalWORKs). One of the main goals of the project was to see if we could use existing, population-based data that the state makes available on a set schedule to facilitate program planning and policy. Birth-records, on their own, fit the bill, but we are excited that the California Health and Human Services Agency (CHHS) is moving steadily in the direction of integrated data and that there seems to be support among departments and offices for this next step. As such, these decisions may be revisited in future iterations, but we decided to restrict the Strong Start Index variables to person-level, birth record data.

SCORING

As presented in Table 2, a simple 0/1 scoring for index items was established. Variables were constructed to reflect the positive condition (i.e., the asset) scored as 1, and the absence of the asset as 0. Where various asset thresholds were possible (e.g., defining the level of completed parental education as a family asset), literature informed our decisions, with a

tendency toward inclusivity.

Considerations
We considered providing domain-specific (i.e., family, health, service, financial) scores alongside total scores, but decided against it. First, we wanted stakeholders to come away with a more holistic (rather than domain-based) understanding of the resources available to children and families. Second, we relied on the literature, rather than on statistical evidence of internal consistency, to guide our grouping strategy. Finally, further stratifying scores by domain could introduce confidentiality issues not otherwise present when presenting aggregate totals. For this reason, we chose to present only aggregate totals by geography.

DE-IDENTIFICATION

Guided by the intended use case and inspired by effective visualizations, we planned to display Strong Start descriptive statistics within an interactive mapping interface so that stakeholders could easily find, view, and digest Strong Start information for their region of interest. Recognizing that both the mean level of assets and the distribution of total assets for each birth within each census tract would provide users with the most actionable information, we planned to include both in the mapping interface.

We consulted the CHHS Data De-identification Guidelines (DDG)[7] in order to assess data files for risk of exposure of personal characteristics for this use case. Following step 2 of the statistical de-identification assessment (Figure 5), we identified a number of census tracts with fewer than 11 births. Out of concern for re-identification of individual birth events, we amended our original plan to show the distribution of asset scores at the census tract level to instead display asset distributions for geographies with more births (i.e., county and state assembly / senate districts).

We again assessed risk of re-identification with the revised dataset. Following advice of our consulting statistician, we applied statistical masking for the resulting dataset (See *Appendix D. Strong Start Score De-identification Recommendations*):

1. We suppressed census tracts with fewer than 11 births. A total of 244 census tracts met this specific criterion.

[7] California Health and Human Services Agency. (2016). *Data De-identification Guidelines (DDG)*. Retrieved from: https://chhsdata.github.io/dataplaybook/documents/CHHS-DDG-V1.0-092316.pdf

2. We further suppressed distributions, means and numbers of births as the census tract level when there were fewer than 11 births.
3. We suppressed these statistics for counties with fewer than 11 births. Alpine county was the only county to meet this criterion for the birth cohort year in this study.
4. There were no state legislative districts (assembly and senate) which required suppression under these criteria.
5. LA county sub-geographies (Supervisorial Districts and Service Planning Areas) included in the mapping tool did not require suppression.

A further consideration for de-identifying data involved examination of the margin of error for each census tract with respect to the average Strong Start score (see *Appendix E. Precision of Strong Start Scores*). The margin of error serves as a confidence interval for each census tract, identifying the expected range of scores. Given the distribution of scores is not presented for all census tracts, the margin of error provides an indication of variability within census tracts, not specifically how many births deviate from the average.

Our statistician confirmed that he was unable to identify individual Strong Start scores once the statistical masking was applied.

SECTION 3: STRONG START INDEX

De-identified birth records were used to construct a Strong Start Index dataset, reflecting all California births registered for calendar year 2016. This dataset included geocoded residential addresses assigned to census tracts, thus allowing each record, which included birth assets (detailed above in Table 2: Strong Start Index Indicators and Scoring) and resulting Strong Start Index score (0-12), to be assigned to a California census tract.[8] We calculated preliminary descriptive statistics using this dataset, as well as performed external validity checks in order to confirm our efforts weren't duplicative of other indices, but also related to actual child outcomes (i.e., child protection involvement and child death).

DESCRIPTIVE STATISTICS

The following tables (Tables 3 through 11) provide a summary of the number and percentage of births with each Strong Start asset present.

TABLE 3. 2016 CALIFORNIA BIRTHS BY STRONG START ASSET PRESENT

	Number of Births with Asset Present	Percentage of Births with Asset Present
All Births	485,573	--
Legal parentage established at birth	454,269	93.6%
Born to non-teen parents	435,173	89.6%
Born to parents with at least a HS degree	338,088	69.6%
Healthy birthweight	452,439	93.2%
Absence of congenital anomalies, abnormalities, or complications at birth[9]	101,510	20.9%
Absence of transmissible (mother-to-child) infections	484,422	99.8%
Access to and receipt of timely prenatal care	400,130	82.4%
Receipt of nutritional services (WIC) if eligible	435,204	89.6%
Hospital with high percentage of births with timely prenatal care	296,341	61.0%
Ability to afford and access healthcare	353,765	72.9%
Born to a parent with a college degree	208,241	42.9%
Born to parents with employment history	335,680	69.1%

[8]5,600 2016 birth records were excluded because they could not be geocoded at the census tract level (1.1% of all 2016 birth records).

[9]See earlier mention that this asset is likely to be recoded in future iterations due to its inclusion of inductions and epidurals as complications of labor and delivery.

TABLE 4. 2016 CALIFORNIA BIRTHS BY NUMBER OF STRONG START ASSETS PRESENT

	Number of Births with Asset Present	Percentage of Births with Asset Present
All Births	485,573	--
1 Asset	129	0.0%
2 Assets	1,206	0.2%
3 Assets	4,622	1.0%
4 Assets	10,306	2.1%
5 Assets	16,807	3.5%
6 Assets	28,655	5.9%
7 Assets	50,905	10.5%
8 Assets	77,833	16.0%
9 Assets	85,074	17.5%
10 Assets	90,991	18.7%
11 Assets	101,447	20.9%
12 Assets	17,598	3.6%

TABLE 5. 2016 BIRTHS AND STRONG START SCORES FOR CALIFORNIA AND LOS ANGELES

	Number of Births	Mean Strong Start Index Score
State of California	485,573	8.9
Los Angeles County	122,139	8.8

TABLE 6. 2016 CALIFORNIA BIRTHS AND STRONG START SCORES BY TERRITORIAL UNITS

	Number of Units (within California)	Mean Number of Births	Median Number of Births	Min. Number of Births	Max. Number of Births	Mean Mean Strong Start Index Score	Median Mean Strong Start Index Score	Min. Mean Strong Start Index Score	Max. Mean Strong Start Index Score
County*	57	8,518.8	2,326.0	25	122,139	8.4	8.5	6.8	9.9
State Assembly District	80	6,069.7	6,056.5	4,264	8,265	8.9	8.9	7.2	10.3
State Senate District	40	12,139.3	12,268.5	8,628	16,228	8.9	8.9	7.6	10.2
LA County SPA	8	15,267.3	15,978.0	5,546	24,159	8.7	8.8	7.1	9.8
LA County Supervisorial District	5	24,427.6	24,675.0	21,751	27,640	8.8	8.9	8.1	9.1
Census Tract*	7,742	60.9	55.0	11	858	9.0	9.0	5.5	11.6

* 1 County and 227 census tracts not included because of low cell sizes.

TABLE 7. 2016 CALIFORNIA BIRTHS AND STRONG START SCORES BY COUNTY

		Number of Births	Mean Strong Start Index Score
State of California		485,573	8.9
	Alameda County	19,476	9.4
	Alpine County	NA	NA
	Amador County	304	9.1
	Butte County	2,476	8.0
	Calaveras County	376	8.5
	Colusa County	318	7.7
	Contra Costa County	12,313	9.6
	Del Norte County	282	7.4
	El Dorado County	1,578	8.9
	Fresno County	15,077	8.6
	Glenn County	372	7.7
	Humboldt County	1,469	8.0
	Imperial County	2,953	7.4
	Inyo County	170	8.0
	Kern County	13,633	7.7
	Kings County	2,206	7.9
	Lake County	736	7.8
	Lassen County	265	7.7

Los Angeles County	122,139	8.8
Madera County	2,326	7.9
Marin County	2,248	9.3
Mariposa County	144	8.1
Mendocino County	1,025	7.7
Merced County	4,092	7.8
Modoc County	52	6.8
Mono County	122	8.7
Monterey County	6,230	8.1
Napa County	1,394	9.4
Nevada County	750	8.5
Orange County	37,763	9.7
Placer County	3,707	9.3
Plumas County	145	7.8
Riverside County	30,457	8.9
Sacramento County	19,486	8.8
San Benito County	784	9.6
San Bernardino County	30,945	8.5
San Diego County	42,518	8.8
San Francisco County	9,020	9.9
San Joaquin County	10,184	8.5
San Luis Obispo County	2,567	8.4
San Mateo County	8,933	9.9
Santa Barbara County	5,475	8.5
Santa Clara County	22,910	9.6
Santa Cruz County	2,741	9.3
Shasta County	2,002	7.6
Sierra County	25	8.3
Siskiyou County	358	7.7
Solano County	5,246	8.7
Sonoma County	4,940	9.0
Stanislaus County	7,794	8.5
Sutter County	1,358	7.9
Tehama County	826	7.7
Trinity County	116	7.5
Tulare County	7,128	7.3
Tuolumne County	451	8.2
Ventura County	9,535	8.9
Yolo County	2,391	8.7
Yuba County	1,239	7.8

TABLE 8. 2016 CALIFORNIA BIRTHS AND STRONG START SCORES BY STATE ASSEMBLY DISTRICT

		Number of Births	Mean Strong Start Index Score
State of California		485,573	8.9
	Assembly District 1	4,264	7.9
	Assembly District 2	4,750	8.4
	Assembly District 3	6,237	7.9
	Assembly District 4	5,004	8.7
	Assembly District 5	5,019	8.2
	Assembly District 6	4,756	9.6
	Assembly District 7	7,029	8.7
	Assembly District 8	6,635	8.8
	Assembly District 9	6,268	8.7
	Assembly District 10	4,600	9.1
	Assembly District 11	6,249	9.0
	Assembly District 12	6,437	8.9
	Assembly District 13	7,145	8.4
	Assembly District 14	5,809	9.4
	Assembly District 15	5,292	9.7
	Assembly District 16	4,758	9.9
	Assembly District 17	5,465	9.7
	Assembly District 18	6,321	9.0
	Assembly District 19	4,931	10.1
	Assembly District 20	6,474	9.4
	Assembly District 21	7,497	7.9
	Assembly District 22	5,881	10.1
	Assembly District 23	6,886	9.1
	Assembly District 24	5,900	10.0
	Assembly District 25	7,162	9.9
	Assembly District 26	7,341	7.4
	Assembly District 27	5,959	8.7
	Assembly District 28	4,978	10.0
	Assembly District 29	4,720	9.5
	Assembly District 30	7,466	8.2
	Assembly District 31	8,265	8.1
	Assembly District 32	8,041	7.5
	Assembly District 33	6,958	8.1
	Assembly District 34	6,952	8.0
	Assembly District 35	6,062	8.1
	Assembly District 36	6,662	7.2

Assembly District 37	4,723	9.2
Assembly District 38	4,701	9.4
Assembly District 39	5,933	8.2
Assembly District 40	7,414	8.6
Assembly District 41	5,011	9.7
Assembly District 42	5,508	8.5
Assembly District 43	4,332	9.4
Assembly District 44	5,600	8.8
Assembly District 45	5,140	9.4
Assembly District 46	5,949	8.7
Assembly District 47	7,738	8.3
Assembly District 48	5,783	9.2
Assembly District 49	6,023	9.6
Assembly District 50	4,317	9.6
Assembly District 51	5,319	8.7
Assembly District 52	7,060	8.8
Assembly District 53	5,568	8.1
Assembly District 54	5,264	8.9
Assembly District 55	7,012	10.0
Assembly District 56	7,183	8.0
Assembly District 57	6,475	9.4
Assembly District 58	5,566	8.9
Assembly District 59	7,541	7.6
Assembly District 60	7,175	9.1
Assembly District 61	7,072	8.8
Assembly District 62	6,051	8.7
Assembly District 63	6,422	8.5
Assembly District 64	7,294	7.9
Assembly District 65	5,527	9.3
Assembly District 66	4,674	9.2
Assembly District 67	6,379	9.2
Assembly District 68	6,871	10.0
Assembly District 69	6,964	8.9
Assembly District 70	5,660	8.8
Assembly District 71	6,063	8.3
Assembly District 72	4,880	9.3
Assembly District 73	4,793	10.1
Assembly District 74	6,102	10.3
Assembly District 75	6,178	8.8
Assembly District 76	6,531	9.0
Assembly District 77	5,671	9.8
Assembly District 78	4,952	9.4
Assembly District 79	7,183	8.8
Assembly District 80	7,797	8.1

TABLE 9. 2016 CALIFORNIA BIRTHS AND STRONG START SCORES BY STATE SENATE DISTRICT

	Number of Births	Mean Strong Start Index Score
State of California	485,573	8.9
State Senate District 1	8,628	8.6
State Senate District 2	9,474	8.6
State Senate District 3	10,554	8.9
State Senate District 4	12,450	8.3
State Senate District 5	13,702	8.5
State Senate District 6	13,627	8.7
State Senate District 7	10,852	9.6
State Senate District 8	12,604	9.0
State Senate District 9	11,629	9.3
State Senate District 10	13,687	9.6
State Senate District 11	10,347	9.9
State Senate District 12	15,562	7.9
State Senate District 13	11,811	10.1
State Senate District 14	16,228	7.6
State Senate District 15	10,633	9.4
State Senate District 16	13,925	7.8
State Senate District 17	10,304	9.0
State Senate District 18	11,690	8.4
State Senate District 19	12,087	8.5
State Senate District 20	14,798	8.5
State Senate District 21	13,192	7.9
State Senate District 22	12,860	9.4
State Senate District 23	13,713	8.6
State Senate District 24	10,642	8.5
State Senate District 25	9,543	9.6
State Senate District 26	8,692	9.6
State Senate District 27	9,329	9.5
State Senate District 28	11,531	9.0
State Senate District 29	12,735	9.6
State Senate District 30	12,828	8.0
State Senate District 31	14,247	8.9
State Senate District 32	11,109	9.2
State Senate District 33	13,066	8.4
State Senate District 34	11,403	9.2
State Senate District 35	12,811	8.3
State Senate District 36	11,413	9.4

State Senate District 37	13,015	10.2
State Senate District 38	12,507	8.7
State Senate District 39	11,122	9.4
State Senate District 40	15,222	8.1

TABLE 10. 2016 LOS ANGELES COUNTY BIRTHS AND STRONG START SCORES BY SUPERVISORIAL DISTRICT

		Number of Births	Mean Strong Start Index Score
Los Angeles County		122,139	8.8
	SD 1	25,789	8.8
	SD 2	27,640	8.1
	SD 3	21,751	8.9
	SD 4	24,675	9.1
	SD 5	22,283	9.0

TABLE 11. 2016 LOS ANGELES COUNTY BIRTHS AND STRONG START SCORES BY SERVICE PLANNING AREA

		Number of Births	Mean Strong Start Index Score
Los Angeles County		122,139	8.8
	SPA 1	5,546	7.1
	SPA 2	24,159	9.0
	SPA 3	23,976	9.5
	SPA 4	11,827	8.6
	SPA 5	6,545	9.8
	SPA 6	15,878	7.7
	SPA 7	16,078	8.9
	SPA 8	18,129	8.7

EXTERNAL VALIDATION

We wanted to confirm that the Strong Start Index, in its current form, added unique insight into the conditions into which children were born in California, and yet correlated with outcomes where expected. To that end, we explored the Strong Start scores relationship to both community-level information (i.e., other published indices) and client-level outcomes (i.e., child protection involvement and death).

COMMUNITY-LEVEL INFORMATION

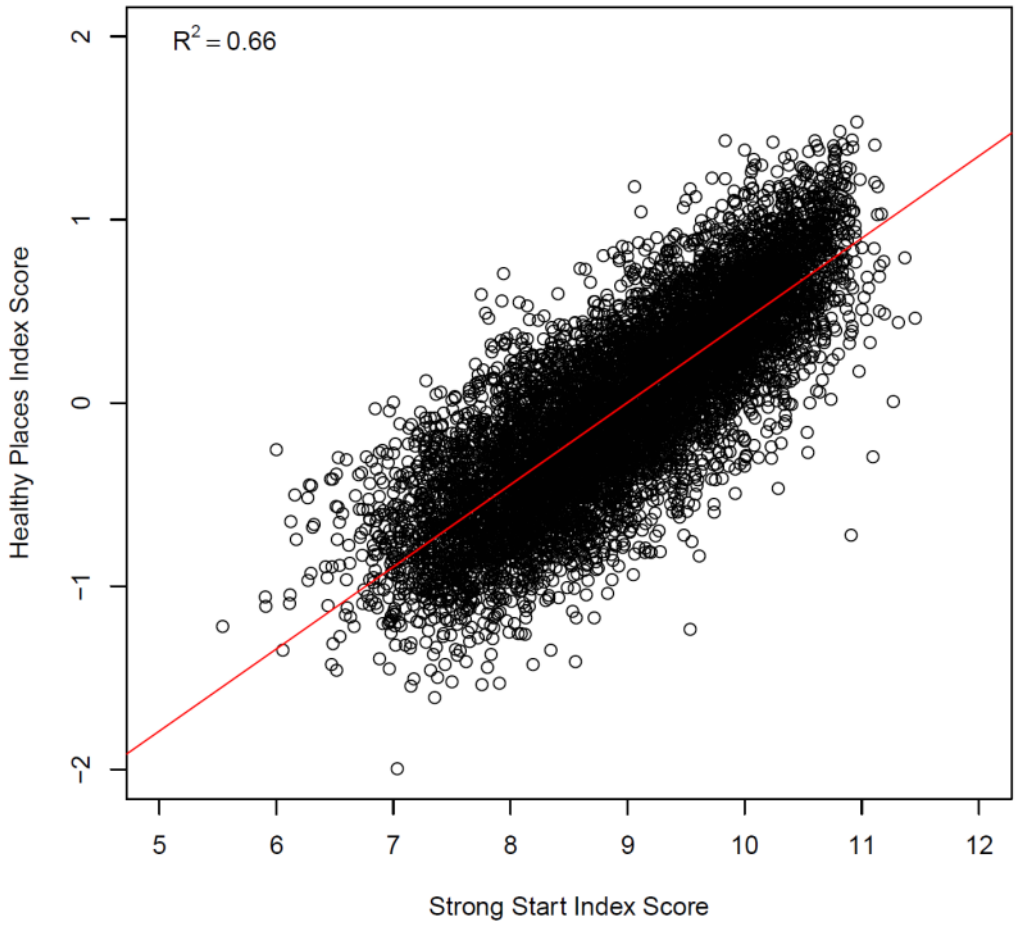
First, we explored the correlation between Strong Start Index scores and two community-level indices in order to make sure that the Index was neither exceedingly divergent, nor duplicative of other published measures. The Healthy Places Index (HPI) and American Human Development Index (HDI) emerged from the landscape analysis (*Section 1: Model Indices*) as potential benchmarks for external validation.

Healthy Places Index (HPI)
The Healthy Places Index (HPI) is an asset-based index that examines the social determinants of health associated with life expectancy. The index uses 25 indicators based on eight domains (i.e., economic; education; housing; health care access; neighborhood; clean environment; transportation; social factors). It is scored at a census tract level. Because the HPI is a compilation of sample-based (i.e., survey based) information, individual indicators that comprise the total HDI score are collected at various time points.

For comparative purposes, we obtained census tract-level HPI scores and plotted them against Strong Start Index scores. As reflected in the scatterplot below (Figure 1), the correlation coefficient (R^2) between these two indices was 0.66, indicating a very strong alignment between our measure of assets at birth with the HPI's measure of broader community conditions.

[10] https://map.healthyplacesindex.org/

FIGURE 1. CORRELATION BETWEEN STRONG START INDEX SCORES AND HEALTHY PLACES INDEX SCORES (CENSUS-TRACT)



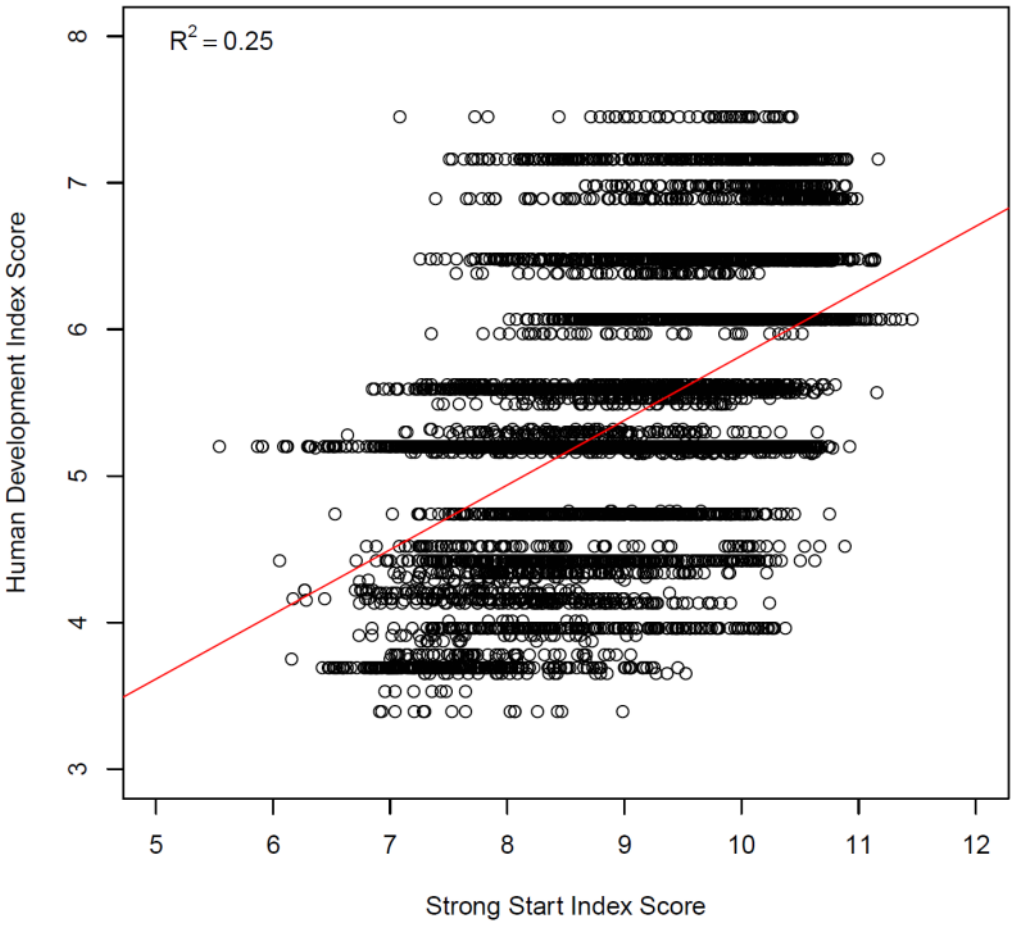
American Human Development Index (HDI)

The American Human Development Index (HDI) is an asset-based index that assesses, using a 10 point scale, how people are faring in terms of a long and healthy life, access to knowledge, and a decent standard of living at the California and county level. Because the HDI is a compilation of sample-based (i.e., survey based) information, individual indicators that comprise the total HDI score are collected at various time points.

For comparative purposes, we obtained county-level scores from A Portrait of California 2014–2015[11] and plotted them against census tract-level Strong Start Index scores. As reflected in the scatterplot

[11] http://www.measureofamerica.org/california2014-15/

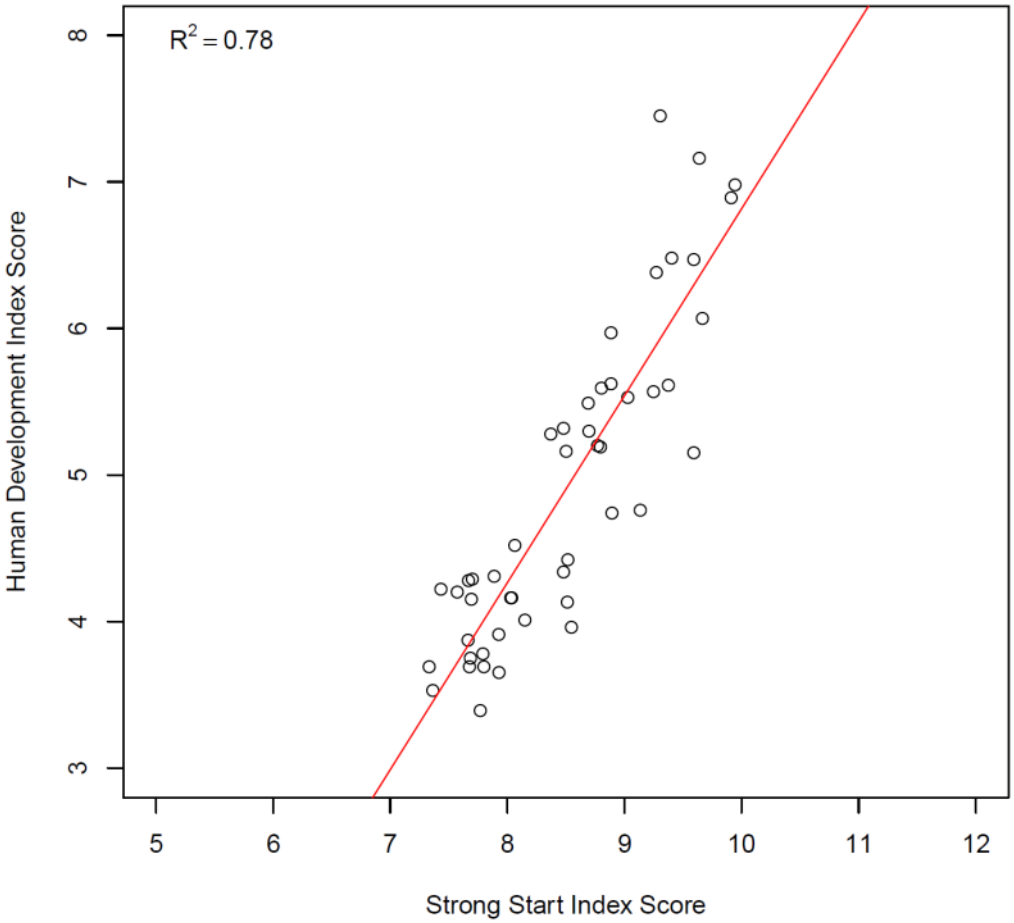
FIGURE 2. CORRELATION BETWEEN STRONG START INDEX SCORES AND HUMAN DEVELOPMENT INDEX SCORES (CENSUS TRACT).



Because HDI scores are not available at the census tract level, it is not surprising that we see a relatively low level of correlation (Figure 2, $R^2 = 0.25$). As such, we used our Strong Start county scores to produce Figure 3, which reflects a very high correlation between our measure of assets at birth with the HDI's measure of broader community conditions.

The difference in the R^2 values between the census tract-level vs county-level correlation analysis suggests that within-county variability substantial, and that aggregation at the county-level masks that variability. This finding reinforces the value of offering scores at the most local (i.e., census tract) level.

FIGURE 3. CORRELATION BETWEEN STRONG START INDEX SCORES AND HUMAN DEVELOPMENT INDEX SCORES (COUNTY)



Conclusions

Overall, the high, but not perfect, correlation between the Strong Start Index and HPI and HDI scores confirmed that the Strong Start Index was neither duplicative of these two published measures, nor exceedingly divergent.

CLIENT-LEVEL OUTCOMES

A large, and growing, body of literature affirms the importance of early childhood experiences in influencing adolescent and adult behavior. Assets present at birth lay the foundation for the emergence of protective factors during childhood that we know are tied to good outcomes and resilience throughout the lifecourse.[12] Although the Strong Start Index is not designed to predict any one particular outcome, it is intended to identify assets that are the underpinnings of future resilience and well-being. Given this, we examined evidence of a relationship between individual Strong Start Index assets and the absence of poor outcomes as proxies for later well-being. Specifically, we examined the correspondence of Strong Start Index assets with post neo-natal infant survival rates and the absence of a child protection system (CPS) referral of alleged abuse or neglect through age 5.

Per approved protocols and relevant data sharing agreements, we securely extracted CPS records reflecting children reported for possible maltreatment from the statewide child welfare information database (CWS/CMS), as well as statewide death record data. We then linked birth records to both death and child protection records using a probabilistic linkage methodology. Please see Appendix F. Record Linkage and Data Security for more information about the Children's Data Network's record linkage and data security process.

In order to assess the relationship between Strong Start Index scores at birth and subsequent CPS involvement and death within the first 5 years of life, both objectively poor outcomes that we would hope to prevent, and expect to be negatively related to asset scores at birth.

Child Protection Involvement

Three logistic regression models were fit using 'in-CWS' as the outcome, as defined as an alleged, investigated, or substantiated allegation of abuse or neglect for children less than five years old at the time of the referral.

The quality of model fit was assessed via Pseudo-R^2 and AUC, see Table 12.

- Model 1: Strong Start Index 12 dummies (e.g., non-teen, paternity, HS Educ, etc.)
- Model 2: Strong Start Index score of 1-12, as dummies (e.g., 1 if Strong Start Index=1, 1 if Strong Start Index=2), 1 if Strong Start Index=3, …, 1 if Strong Start Index=12)

- Model 2: Strong Start Index score of 1-12, as dummies (e.g., 1 if Strong Start Index=1, 1 if Strong Start Index=2), 1 if Strong Start Index=3, ..., 1 if Strong Start Index=12)
- Model 3: Strong Start Index Score of 1 to 12 (continuous)

Metrics of quality of fit indicated:

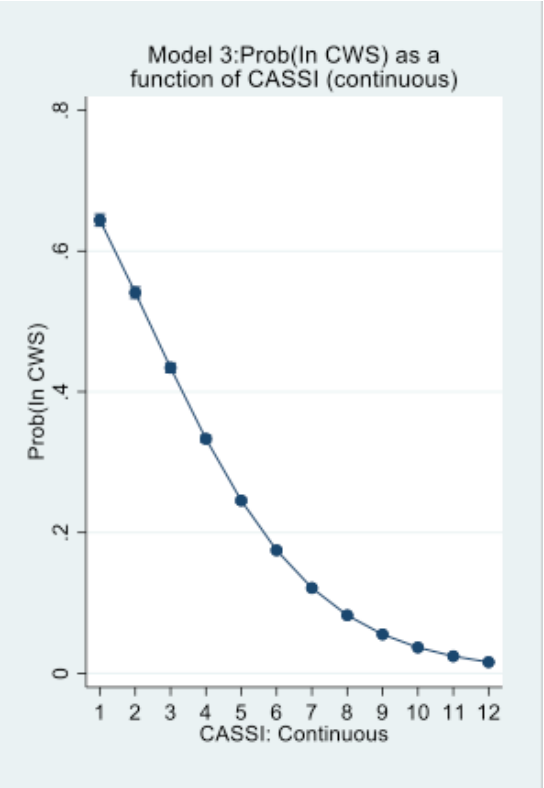
- Model 2 (using Strong Start Index as dummies) fit slightly poorer than Model 1 – Pseudo R^2 0.117 vs. 0.143; AUC=0.757 vs. 0.785
- Model 3 (using Strong Start Index as continuous) had a similar fit to Model 2.

TABLE 12. MODEL FIT (PSEUDO-R^2 & AUC) PREDICTING IN-CWS

MODEL	PSEUDO-R^2	AUC
1. Strong Start Index: 12 Items	0.143	0.785
2. Strong Start Index Count: 1-12 (indicators)	0.117	0.757
3. Strong Start Index Count: 1-12 (continuous)	0.113	0.757

The predicted probability of future child protection involvement by Strong Start Index for Model 3 is shown in Figure 4.

FIGURE 4. PROBABILITY (IN CWS) AS A FUNCTION OF STRONG START INDEX (MODEL 3)



Overall, the Strong Start Index demonstrated a strong, graded relation with the predicted probability of child protection involvement before age 5. In other words, this means that the more Strong Start assets a child was born with, the less likely they were to become involved with the child protection system in early childhood.

Death

Similarly using 2007 data, three logistic regression models were fit using 'in-Death' as the outcome. Death was restricted to non-neonatal deaths (from 1 month after birth until December 31, 2012) (n=933). 1,422 neonatal deaths, deaths within 1 month (30 days) of birth, were excluded from the analysis. As you can see from Tables 13 and 14, the rate of non-neonatal deaths is small (n=933, 0.18%), but enough to analyze.

TABLE 13. 2007 BIRTH COHORT NEONATAL AND NON-NEONATAL DEATH

	Neonatal Death		
Non-Neonatal Death	N	Y	Total
N	509,420	1,422	510,842
Y	933	0	933
Total	510,353	1,422	511,775

TABLE 14. 2007 BIRTH COHORT FREQUENCY OF NON-NEONATAL DEATH

		Frequency	Percent	Valid	Cumulative
Valid	N	510,842	99.82	99.82	99.82
	Y	933	0.18	0.18	100.00
	Total	511,775	100.00	100.00	

The analyses focus on the N=933 non-neonatal deaths.

We performed bivariate analyses using 1) tabulations with Fisher's exact tests, and 2) logistic regressions. Three models were estimated:

- Model 1: Strong Start Index 12 dummies (e.g., non-teen, paternity, HS Educ, etc.…)
- Model 2: Strong Start Index score of 1-12, as dummies (e.g., 1 if Strong Start Index=1, 1 if Strong Start Index=2), 1 if Strong Start Index=3, …, 1 if Strong Start Index=12)
- Model 3: Strong Start Index Score of 1 to 12 (continuous)

The quality of model fit was assessed via Pseudo-R^2 and AUC, see Table 15.

- Model 2 (using Strong Start Index as dummies) is a poorer fit than Model 1 – The Pseudo R^2 drops from 0.064 to 0.024 and the AIC drops from 0.737 to 0.658.
- Model 3 (using Strong Start Index as continuous) had a similar fit to Model 2.

TABLE 15. MODEL FIT (PSEUDO-R^2 & AUC) PREDICTING IN-CWS

MODEL	PSEUDO-R^2	AUC
1. Strong Start Index: 12 Items	0.064	0.737
2. Strong Start Index Count: 1-12 (indicators)	0.024	0.658
3. Strong Start Index Count: 1-12 (continuous)	0.024	0.658

Again, the Strong Start Index demonstrated a strong, graded relation with the predicted probability of non-neonatal death. In other words, this means that the more Strong Start assets a child was born with, the less likely they were to die before age 5.

CONCLUSIONS

The results of our external validation checks confirmed that the Strong Start Index, in its current form, adds unique insight into the conditions into which children are born in California. It is neither exceedingly divergent, but nor duplicative of other published measures. In addition, our initial analyses suggest that Strong Start Scores are related to at least two important indicators of child health and well-being (i.e., child protection involvement and death).

We are excited to have developed these data for the 2016 birth cohort and look forward to calculating Strong Start scores for successive cohorts. It is our sincere hope that these data will be used to better characterize young children and families, streamline processes for stakeholders, and, ultimately, change the conversation around investments for children and families. As such, we welcome comments, questions, and suggestions as we iterate and improve the Strong Start Index.

LITERATURE CITED

1. Beimers, D., & Coulton, C. J. (2011). Do employment and type of exit influence child maltreatment among families leaving Temporary Assistance for Needy Families? *Children and Youth Services Review, 33*(7), 1112-1119.
2. Buescher, P. A., Roth, M. S., Williams, D., & Goforth, C. M. (1991). An evaluation of the impact of maternity care coordination on Medicaid birth outcomes in North Carolina. *American Journal of Public Health, 81*(12), 1625-1629.
3. Cederbaum, J. A., Putnam-Hornstein, E., Sullivan, K., Winetrobe, H., & Bird, M. (2015). STD and abortion prevalence in adolescent mothers with histories of childhood protection involvement. *Perspectives on Sexual and Reproductive Health, 47*(4), 187-193.
4. Culhane, J. F., Webb, D., Grim, S., & Metraux, S. (2003). Prevalence of child welfare services involvement among homeless and low-income mothers: A five-year birth cohort study. *Journal of Sociology & Social Welfare, 30*, 79.
5. Doidge, J. C., Higgins, D. J., Delfabbro, P., & Segal, L. (2017). Risk factors for child maltreatment in an Australian population-based birth cohort. *Child Abuse & Neglect, 64*, 47-60.
6. Dubowitz, H., Kim, J., Black, M. M., Weisbart, C., Semiatin, J., & Magder, L. S. (2011). Identifying children at high risk for a child maltreatment report. *Child Abuse & Neglect, 35*(2), 96-104.
7. Figlio, D., Hamersma, S., & Roth, J. (2009). Does prenatal WIC participation improve birth outcomes? New evidence from Florida. *Journal of Public Economics, 93*(1-2), 235-245.
8. Finno-Velasquez, M., Palmer, L., Prindle, J., Tam, C., & Putnam-Hornstein, E. (2017). A birth cohort study of Asian and Pacific Islander children reported for abuse or neglect by maternal nativity and ethnic origin. *Child Abuse & Neglect, 72*, 54.
9. Folger, A. T. (2014). Maternal Chlamydia trachomatis infections and preterm birth: The impact of early detection and eradication during pregnancy. *Maternal and Child Health Journal, 18*(8), 1795-1802.
10. Johnson, H. L., Ghanem, K. G., Zenilman, J. M., & Erbelding, E. J. (2011). Sexually transmitted infections and adverse pregnancy outcomes among women attending inner city public sexually transmitted diseases clinics. *Journal of Sexually Transmitted Diseases, 38*(3), 167-171.
11. Haglund, B., & Cnattingius, S. (1990). Cigarette smoking as a risk factor for sudden infant death syndrome: A population-based study. *American Journal of Public Health, 80*(1), 29-32.

12. Hjern, A., Vinnerljung, B., & Lindblad, F. (2004). Avoidable mortality among child welfare recipients and intercountry adoptees: A national cohort study. *Journal of Epidemiology & Community Health, 58*(5), 412-417.
13. Liu, B., Roberts, C. L., Clarke, M., Jorm, L., Hunt, J., & Ward, J. (2013). Chlamydia and gonorrhoea infections and the risk of adverse obstetric outcomes: A retrospective cohort study. *Journal of Sexually Transmitted Infections, 89*(8), 672–678.
14. MacKenzie, M. J., Kotch, J. B., & Lee, L. C. (2011). Toward a cumulative ecological risk model for the etiology of child maltreatment. *Children and Youth Services Review, 33*(9), 1638-1647.
15. Mathews, T. J., & MacDorman, M. F. (2007). Infant mortality statistics from the 2004 period linked birth/infant death data set. *National vital statistics reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 55*(14), 1-32.
16. Murphey, D. A., & Braner, M. (2000). Linking child maltreatment retrospectively to birth and home visit records: An initial examination. *Child Welfare, 79*(6), 711.
17. Needell, B., & Barth, R. P. (1998). Infants entering foster care compared to other infants using birth status indicators. *Child Abuse & Neglect, 22*(12), 1179-1187.
18. Parrish, J. W., Young, M. B., Perham-Hester, K. A., & Gessner, B. D. (2011). Identifying risk factors for child maltreatment in Alaska. *American Journal of Preventive Medicine, 40*(6), 666-673.
19. Putnam-Hornstein, E. (2011). Report of maltreatment as a risk factor for injury death: A prospective birth cohort study. Child Maltreatment, 16(3), 163-174.
20. Putnam-Hornstein, E., Cederbaum, J. A., King, B., Cleveland, J., & Needell, B. (2013). A population-based examination of maltreatment history among adolescent mothers in California. *Journal of Adolescent Health, 53*(6), 794-797.
21. Putnam-Hornstein, E., Cederbaum, J. A., King, B., Eastman, A. L., & Trickett, P. K. (2015). A population-level and longitudinal study of adolescent mothers and intergenerational maltreatment. *American Journal of Epidemiology, 181*(7), 496-503.
22. Putnam-Hornstein, E., & Needell, B. (2011). Predictors of child protective service contact between birth and age five: An examination of California's 2002 birth cohort. *Children and Youth Services Review, 33*(8), 1337-1344.
23. Putnam-Hornstein, E., Needell, B., King, B., & Johnson-Motoyama, M. (2013). Racial and ethnic disparities: A population-based examination of risk factors for involvement with child protective services. *Child Abuse & Neglect, 37*(1), 33-46.
24. Putnam-Hornstein, E., Schneiderman, J. U., Cleves, M. A., Magruder, J., & Krous, H. F. (2014). A prospective study of sudden unexpected infant death after reported

maltreatment. *The Journal of Pediatrics, 164(*1), 142-148.

25. Putnam-Hornstein, E., Simon, J. D., Eastman, A. L., & Magruder, J. (2015). Risk of re-reporting among infants who remain at home following alleged maltreatment. *Child Maltreatment, 20*(2), 92-103.
26. Putnam-Hornstein, E., Webster, D., Needell, B., & Magruder, J. (2011). A public health approach to child maltreatment surveillance: Evidence from a data linkage project in the United States. *Child Abuse Review, 20*(4), 256-273.
27. Reed, M. M., Westfall, J. M., Bublitz, C., Battaglia, C., & Fickenscher, A. (2005). Birth outcomes in Colorado's undocumented immigrant population. *BMC Public Health, 5*(1), 100.
28. Resnick, M. B., Gueorguieva, R. V., Carter, R. L., Ariet, M., Sun, Y., Roth, J., ... & Mahan, C. S. (1999). The impact of low birth weight, perinatal conditions, and sociodemographic factors on educational outcome in kindergarten. *Pediatrics, 104*(6), e74.
29. Schnitzer, P. G., & Ewigman, B. G. (2005). Child deaths resulting from inflicted injuries: household risk factors and perpetrator characteristics. *Pediatrics, 116*(5), e687-e693.
30. Slack, K. S., Berger, L. M., DuMont, K., Yang, M. Y., Kim, B., Ehrhard-Dietzel, S., & Holl, J. L. (2011). Risk and protective factors for child neglect during early childhood: A cross-study comparison. *Children and Youth Services Review, 33*(8), 1354-1363.
31. Williams, G., Tonmyr, L., Jack, S. M., Fallon, B., & MacMillan, H. L. (2011). Determinants of maltreatment substantiation in a sample of infants involved with the child welfare system. *Children and Youth Services Review, 33*(8), 1345-1353.
32. Wu, S. S., Ma, C. X., Carter, R. L., Ariet, M., Feaver, E. A., Resnick, M. B., & Roth, J. (2004). Risk factors for infant maltreatment: a population-based study. *Child Abuse & Neglect, 28*(12), 1253-1264.
33. Zhou, Y., Hallisey, E. J., & Freymann, G. R. (2006). Identifying perinatal risk factors for infant maltreatment: an ecological approach. *International Journal of Health Geographics, 5*(1), 53.

APPENDICES

APPENDIX A. OTHER LOCAL, STATE, AND NATIONAL DATA SOURCES AND TOOLS

CALIFORNIA COUNTY SCORECARD OF CHILDREN'S WELL-BEING

[California]
https://www.childrennow.org/portfolio-posts/2018scorecard/

The California County Scorecard of Children's Well-Being evaluates children's well-being by examining 28 indicators across California's 58 counties, over time, and by race and ethnicity. It leverages administrative and survey data to provide county-level data visualizations. The indicators are organized into three domains: Education, Health, and Child Welfare & Economic Well-Being. The scorecard uses the domains to rate each county's relative performance. The relative performances are grouped by child population density by race/ethnicity, percentage of families with children who can afford basic living expenses, average family income, and percentage of children living at or below poverty.

The star ratings can be used to compare two or more counties and highlight strengths and areas for improvement and to better understand the regional demographics of each county. In addition, the scorecard can be used to compare a county's performance to the state's performance, rank sorting, and allow counties to track progress over time.

KIDS COUNT

[National, Includes California]
https://datacenter.kidscount.org/
http://www.aecf.org/m/resourcedoc/AECF-KIDSCOUNTIndex-2012.pdf

KIDS COUNT, a project of the Annie E. Casey Foundation, evaluates the well-being of children in the United States over time. The 16 indicators are grouped into four domains: Economic Well-Being, Education, Health, and Family & Community. The data are organized by:

- State
- Topics
  - Demographics
  - Economic Well-Being: employment and income, public assistance, housing, poverty

- Education: children with disabilities, early childhood, school age, young adults
- Family & Community: community environment, family structure, voting, other family and community
- Health: birth outcomes, health insurance, vital statistics, dental health
- Safety & Risky Behaviors: child abuse and neglect, juvenile justice, out of home placement, public safety

HEALTHYCITY.ORG

[California]
http://www.healthycity.org/

HealthCity.org is an online, community-based tool used to access data, create maps and service referrals, and allow individuals to collaborate toward a common goal. It incorporates demographic and secondary information from surveys and end-users. Data are categorized by race, gender, categories, and statistics based on categories. Categories include: Age, citizenship status, educational attainment, health care access, household size, income, language spoken, length of residence, marital status, means of transport, non-driver/carless, place of birth, population, race/ethnicity, sexual orientation, transportation, and usual source of care.

AskCHIS

[California]
https://healthpolicy.ucla.edu/Pages/home.aspx

UCLA's AskCHIS™ is an online health query system that allows users to find health statistics on California state, counties, regions, Los Angeles Service Planning Districts, and San Diego Health Districts. The data reflects the responses of 20,000 Californians interviewed each year by the California Health Interview Survey (CHIS). The tool can be used to compare health statistics by age, race/ethnicity, and poverty level. It can be used to examine trends in data from 2001 and across time.

AMERICA'S CHILDREN: KEY NATIONAL INDICATORS OF WELL-BEING, 2017

https://www.childstats.gov/americaschildren/

America's Children: Key National Indicators of Well-Being, 2017 uses Federal data to identify

the key indicators that affect children's well-being and monitors trends over time. The goals of the report are to improve reporting of Federal data on children and families and make these data readily available. The report identified 41 key indicators and organized them into seven domains: Family and social environment, economic circumstances, health care, physical environment and safety, behavior, education, and health.

PORTRAIT OF CALIFORNIA

[California]
http://www.measureofamerica.org/california2014-15/

Measure of America's A Portrait of California uses the human development framework and index to examine the well-being of children and families across California. The report uses a ten-point scale, the American Human Development (HD) Index to examine three domains: a long and healthy life, access to knowledge, and a decent standard of living. Inequalities in these domains divide communities into five different "Californias:" One Percent California, Elite Enclave California, Main Street California, Struggling California, and Disenfranchised California. "Californias" are defined by well-being and access to opportunity. The report can be used to compare California's counties, cities, 265 Census Bureau-defined areas, women and men, racial and ethnic groups, and examine changes over time.

PORTRAIT OF LOS ANGELES COUNTY

http://www.measureofamerica.org/los-angeles-county/

Measure of America's A Portrait of Los Angeles County uses the human development framework and index to examine the well-being and equity of LA County residents. The report uses a ten-point scale, the American Human Development (HD) Index to examine three domains: a long and healthy life, access to knowledge, and a decent standard of living. Inequalities in these domains divide communities into "Five LA Counties:" Glittering LA, Elite Enclave LA, Main Street LA, Struggling LA, and Precarious LA. The report includes indexes for 106 cities and unincorporated areas in LA County, and 35 community plan areas with the LA for demographic groups. It examines a range of issues including health, education, living standards, environmental justice, housing, homelessness, violence, and inequality.

THE NCVHS MEASUREMENT FRAMEWORK FOR COMMUNITY HEALTH WELL-BEING, V4

https://ncvhs.hhs.gov/wp-content/uploads/2018/03/NCVHS-Measurement-Framework-V4-Jan-12-2017-for-posting-FINAL.pdf

The National Committee on Vital and Health Statistics (NCVHS) measures determinants of

community health and well-being through equity and life-course perspectives. The framework uses ten domains and subdomains. The domains include: Community vitality, demographics, economy, education, environment, food and agriculture, health, housing, public safety, and transportation. This framework allows communities to access and choose data at the local level, identify key issues and best practices, and make comparisons against peers. In addition, it allows each sector to view how they are accomplishing outcomes and accomplishing collective impact.

MASSACHUSETTS COMMUNITY HEALTH INFORMATION PROFILE (MassCHIP)

http://www.mass.gov/eohhs/researcher/community-health/masschip/

MassCHIP provides community-level data to access health needs, track health status indicators, and assess health programs. Indicators are grouped in seven domains: Demographic indicators, all perinatal and child health indicators, infectious disease indicators, injury indicators, chronic disease indicators, substance abuse indicators, and hospital discharges. Data are available at a variety of geographic levels including: Cities and towns, CHNA, counties, Healthy Start Regions, state totals, and EOHHS regions. The data presented by MassCHIP is accessible to health care providers, state and federal agencies, universities, community health centers, and local boards of health.

VITAL STATISTICS DATA

[National, Includes California]
https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm

Centers for Disease Control and Prevention's (CDC) Division of Vital Statistics provides data files on birth, period linked birth-infant death, birth cohort birth – infant death, mortality multiple cause, and fetal death. The online data access tools include tables, data files, and reports. Through WONDER, users are able to query population data.

APPENDIX B. EVIDENCE FOR INCLUSION IN THE CALIFORNIA STRONG START INDEX, BY INDICATOR

	Lower incidence of child maltreatment / child protection system involvement	Lower risk of infant mortality / stillbirth	Lower risk of early childhood mortality	Lower risk of mortality	Lower risk ofpreterm birth / low birth weight	Reduced prevalence of STDs	Lower risk of adverse educational outcomes
FAMILY							
Legal parentage established at birth	4; 6; 8; 16; 17; 18; 22; 23; 25; 26; 30; 31; 32; 33	2; 15; 24	19; 29		9	3	28
Born to non-teen parents	1; 4; 5; 6; 8; 14; 16; 17; 18; 22; 23; 25; 26; 32; 33	2; 15; 24	19; 29		9; 10		28
Born to parents with at least a HS degree	1; 4; 5; 6; 8; 14; 16; 17; 23; 25; 26; 30; 32; 33	2; 15	19; 29		9; 10		28
HEALTH							
Healthy birth weight	1; 5; 8; 17; 18; 22; 23; 25; 30; 32; 33	15; 24				3	28
Absence of congenital anomalies, abnormalities, or complications at birth	22; 23; 25; 26			12; 15; 19	10; 13		28
Absence of transmissible (mother-to-child) infections	3	13			10; 13		
SERVICE							
Access to and receipt of timely prenatal care	1; 8; 16; 17, 22; 23; 25; 26; 32; 33	2	29		9;10	3	28
Receipt of nutritional services (WIC) if eligible	1; 6; 14; 30; 32				7		
Hospital with high percentage of births with timely prenatal care	1; 8; 16; 17, 22; 23; 25; 26; 32; 33	2	29		9; 10	3	28
FINANCIAL							
Ability to afford and access healthcare	1; 8; 14; 22; 23; 25; 30; 32; 33	2; 24	19; 29		7; 9		
Born to a parent with a college degree	23; 25; 26	15	19				
Born to parents with employment history	1; 5; 18; 30						

APPENDIX C. ASSESSMENT OF INCLUSION OF HEALTHY PLACES INDEX (HPI)

Summary

Five logistic regression models were fit using "in-CWS" as the outcome. The outcomes "in-CWS" indicates an alleged, investigated, substantiated allegation of abuse or neglect for children less than five years old at the time of the referral. These analyses were conducted for children born in California in 2007.

- Model 1: Strong Start Index 12 dummies (e.g., non-teen, paternity, HS Educ, etc.)
- Model 2: Strong Start Index score of 1-12, as dummies (e.g., 1 if Strong Start Index=1, 1 if Strong Start Index=2), 1 if Strong Start Index=3, ..., 1 if Strong Start Index=12)
- Model 3: Strong Start Index Score of 1 to 12 (continuous)
- Model 4: Strong Start Index Score of 1 to 12 (continuous) + Healthy Place Index (total score)
- Model 5: Strong Start Index Score of 1 to 12 (continuous) + Healthy Place (8 Subscales)

Results

The quality of model fit was assessed via Pseudo-R^2 and AUC, see Table 1.

- Model 2 (using Strong Start Index as dummies) fit slightly poorer than Model 1 – Pseudo R^2 0.117 vs. 0.143; AUC=0.757 vs. 0.785.
- Model 3 (using Strong Start Index as continuous) had a similar fit to Model 2.
- The fit of Model 4 vs. 3 is very slightly better. The healthy place index adds almost nothing in terms of association.
- The fit of Model 5 vs. 3 is very slightly better. The 8 subscales of the healthy place index add almost nothing in terms of association.

TABLE 1. MODEL FIT (PSEUDO-R^2 & AUC) PREDICTING IN-CWS

MODEL	PSEUDO-R^2	AUC
1. Strong Start Index: 12 Items	0.143	0.785
2. Strong Start Index Count: 1-12 (indicators)	0.117	0.757
3. Strong Start Index Count: 1-12 (continuous)	0.113	0.757
4. Strong Start Index Count: 1-12 (continuous) + Healthy Place Index (Total Score)	0.120	0.765
5. Strong Start Index Count: 1-12 (continuous) + Healthy Place (8 Subscales)	0.121	0.765

Determination

After exploring publicly available community-level data, we decided to restrict the Strong Start Index variables to person-level, birth record data because of (1) it was consistent with our goal of straightforward index and (2) with the addition of community-level variables (as represented by the Healthy Start Index variables), there were limited or no gains in correlation with our proxy outcomes for well-being.

APPENDIX D. STRONG START SCORE DE-IDENTIFICATION RECOMMENDATIONS

Overview

I examined the Strong Start Index dataset to determine if and when the data could be used to ascertain characteristics of the births from which the index is calculated. Of the six geographic metrics, I identified two which could potentially lead to identifying the characteristics of individual births – 1) census tract, and 2) county. Both of these geographic units had regions with a small number of births (e.g., less than 10 births). The remaining four geographic units (i.e., State Assembly District, State Senate District, LA SPA, and LA Supervisory District) all had regions with large numbers of births, at least 4,000 per unit. Thus, census tract and county units were examined to determine if/whenbirth characteristics of individual births could be identified. The next two pages describe the analysis of census tracts and counties and the recommendations for preserving the privacy of all individuals born within these areas.

Analyses and Recommendations

Census Tracts

At the census tract level, there are three forms of summary data – 1) The average STRONG START INDEX score, 2) The Strong Start Index quintile, and 3) the counts of the number of births by the number of potential assets (i.e., count of births with 12 assets, 11 assets, … 2 assets, and 1 asset). I will refer to these, collectively, as Strong Start Index summary statistics.

My first analysis concerned identifying tracts with small numbers of births. There were 227 tracts (out of 7,969, 2.9%), which had 10 or fewer births.

Recommendation #1. I recommend suppressing the reporting of any Strong Start Index summary statistics for any census tract with 10 or fewer births. This would result in suppressing the data for 227 census tracts.

My second analysis concerned whether the asset counts (e.g., ASSET_12 count of births with 12 assets) could be used to reveal the characteristics of a specific birth.

Example #1. Suppose there is only 1 birth in a census tract, and that birth is categorized as having 12 assets. We can clearly infer that all 12 STRONG START INDEX characteristics were coded as “yes” for that birth. The data for such a census tract would need to be suppressed to protect the privacy of the child born in that census tract.

Example #2. Suppose there were 11 births in a census tract, and 10 of those 11 births had 12 assets. For any birth in that census tract, we have an 11 out of 12 (92%) chance of choosing a birth with 12 assets. We would have a 92% chance of saying that any given birth in this census tract reflects a birth with 12 assets present.

Recommendation #2. I recommend that Strong Start Index summary statistics should be suppressed when the data could be used to say, with 80% certainty or greater, that a birth reflected a particular Strong Start Index characteristic.

Table 1 shows the results obtained of applying recommendations 1 and 2. A total of 244 (out of 7,969, 3.1%) of census tracts are flagged for data suppression, of which 206 are flagged solely for having 10 or fewer observations in the census tract.

TABLE 1. FLAGS VARIABLES REFLECTING POTENTIAL FOR IDENTIFICATION IN STRONG START INDEX DATASET.

flagged	flag_b10	flag_asset12	flag_asset11	flag_asset11up	_Freq_	_Perc_
0	0	0	0	0	7725	96.94
1	0	0	0	1	16	0.20
1	0	0	1	1	1	0.01
1	1	0	0	0	206	2.59
1	1	0	0	1	4	0.05
1	1	0	1	1	14	0.18
1	1	1	0	1	3	0.04

LEGEND

flag_b10 -> flagged due to N<=10 in tract.
Flag_asset12 -> flagged due to 12 asset indicator giving 80+% chance of revealing birth characteristics.
Flag_asset11 -> flagged due to 11 asset indicator giving 80+% chance of revealing birth characteristics.
Flag_asset11up -> flagged due to 11 and 12 asset indicators (together) giving 80+% chance of revealing birth characteristics.
flagged -> combination flag. Observation flagged for suppression of STRONG START INDEX summary statistics.

Considering the observations that are not flagged (in Table 1), I investigated whether the mean Strong Start Index score could be used to say, with 80% certainty or greater, that a birth reflected a particular characteristic. No other observations were found where such an inference could be made.

Counties

At the county level, there is just one form of summary data – The average STRONG START INDEX score. My first analysis concerned identifying counties with small numbers of births. There were 6 (out of 70, 8.6%) which had 10 or fewer births.

Recommendation #3. I recommend suppressing the reporting of any Strong Start Index summary statistics for any county with 10 or fewer births. This would result in suppressing the data for 1 county (specifically, county "003") see below.

```
. list if (BIRTH_CNTY <= 10), abb(30)

     +-----------------------------------+
     | CENS_CNTY   MEAN_CNTY   BIRTH_CNTY |
     |-----------------------------------|
  2. |        003    9.333333           3 |
     +-----------------------------------+
```

Considering the observations that are not flagged (via recommendation 3), I investigated whether the mean Strong Start Index score could be used to say, with 80% certainty or greater, that a birth reflected a particular characteristic. No observations were found where such an inference could be made.

APPENDIX E. PRECISION OF STRONG START SCORES

CENSUS TRACT

I calculated the average Strong Start score by census tract. The average was 8.99 (SD=1.02, min=3, max=12). I then computed the margin of error associated with each census tract mean. The margin of error can be used to compute the 95% confidence interval for the mean as… [mean Strong Start score] + [margin of error]

So, if the mean score is 8.2 and the margin of error is 0.7, the 95% CI is (7.5,8.9).
We have previously masked 244 (out of 7,969, 3.1%) of census tracts to prevent identification. The tabulations of the margin of error, after accounting for these 244 masked values are shown in Table 1.

Table 1. Tabulation of Margins of Error

		Freq.	Percent	Valid	Cum.
Valid	+/- 1.5->2.0	15	0.19	0.19	0.19
	+/- 1.0->1.5	128	1.61	1.66	1.85
	+/- 0.5->1.0	2743	34.42	35.51	37.36
	+/- min->0.5	4839	60.72	62.64	100.00
	Total	7725	96.94	100.00	
Missing	Masked: Deid	244	3.06		
Total		7969	100.00		

Table 1 shows

1. In yellow: that 0.19% of census tracts have a margin of error of 1.5 or greater.
2. In blue: 1.61% of census tracts have a margin of error of 1.0-1.5
3. In yellow and blue: 1.8% of census tracts have a margin of error of 1.0 or greater
4. In pink: 3.06% are masked to prevent identification.

For numeric/tabular presentations, we might want to include the margin of error. Further, we might want to call attention to Strong Start means by census tract where the margin of error is 1.0 or greater. For graphical presentations, we might want to consider masking Strong Start means by census tract where the margin of error is 1.0. If we did this – 1) This would mask a total of 4.86% of census tract means (3.06% for identification, 1.8% for imprecision), 2). The margin of error for the graphical display of census tract means would all be less than 1.

COUNTY

After masking for identification, all margins of error are less than 1. The county with N=25 has a margin of error of 0.9.

APPENDIX F. RECORD LINKAGE AND DATA SECURITY

Overview

At the Children's Data Network, we use 'record linkage' to connect person-level information from separate state and county databases. Through this process, we are able to create encrypted "linkage keys" to connect service, program, and outcome information for individuals.

The CDN relies on an open-source software program from Choice Maker LLC to probabilistically match records and generate linkage keys in a scientific, secure, and robust manner.

Data Transfer

The CDN receives data files / records from our agency partners via Secure File Transfer Protocol (SFTP), which uses standard encryption processes to protect information. Database analysts on our team are responsible for extracting and setting-up the encrypted data on a non-networked server. This non-networked server is hosted in a high-security facility (e.g., building security, facility biometric screening) at USC, with access restricted to authorized CDN personnel.

Data Cleaning

In order to improve our record linkage efficiency, we go through a process of data 'hygiene' checks, cleaning, and standardization. This process involves selecting fields relevant to linkage, such as addresses and birthdates, and standardizing them in SQL data tables. We currently have 10 major SQL data tables – broadly conceptualized as containing information organized around children (including both guardian and address information) and individuals (including address information). Auxiliary tables also exist to store additional relevant information to linkages.

Linkage

Our linkage model compares selected fields of two records at a time, one record each from different data sources. The model can also compare records from the same data source to identify duplicates (multiple records for the same client) within a program. For each field in the pair of records, the model applies a set of logical instructions, called clues, to assess whether values of the selected field (such as First Name) agree, disagree or do not have enough information to be compared. In addition, each clue is categorized as to whether it indicates that two records may represent the same person (a Match clues) or differen persons (a Differ clue). CDN has customized our model with many such clues to respond to various scenarios among our linkage variables.

Match Probabilities

Each clue in the model has an associated weight, which represents the significance of its contribution to the final decision of whether the records match. Weights are computed from training data based on the Maximum Entropy machine learning algorithm. The weights of all activated, or "fired", clues are tallied by Match or Differ category and then combined to yield a single Match Probability between zero and one for each pair of records.

This numerical value represents the level of confidence that the pair of records is a match. A probability value of 0.8 and above typically indicates sufficient matching information. The closer the probability gets to 1, the stronger the clues identify a Match decision.

Model Performance

By examining the general pattern of clues "fired" for the entire universe of linkage, we choose record pairs for manual review that represent patterns that occur with high frequency, focusingparticularly on patterns of clue firings that produce questionable probabilities.

Experts in data linkage mark the pairs as either a Match, Differ, or Hold. The marked pairs are added to the existing training data and used to adjust the clue weights. In addition, data experts look for cases where additional clues might be needed, in order to capture the experts' judgments about the record pairs. The human training process may repeat for several iterations until researchers are satisfied with ChoiceMaker's decisions using the reviewed pairs as training data. The threshold of 0.8 mentioned previously was selected after many iterations of this process, where ChoiceMaker had "learned" to correctly label pairs sufficiently and be labeled a stable and mature model.