
AN OVERVIEW OF THE CHILDHOOD
INTEGRATED LONGITUDINAL DATA SYSTEM

Nina Lalich

Elizabeth Anthony

Francisca Richter

Claudia Coulton

Robert Fischer

Center on Urban Poverty and Community Development

Case Western Reserve University

April 2015

1. Introduction

Governmental and nonprofit agencies serving children and families routinely gather administrative records. These records are often used internally to describe patterns of service use, risk factors, costs and outcomes. But, individuals who use agency programs often use other programs at the same time, traversing multiple systems as they move along in their development. The on-going integration of administrative records across agencies and time has the potential to provide new types of information that can be used to evaluate outcomes, drive decision making, target resources and gain an understanding of how the collective work of agencies and systems are addressing the needs and concerns in their communities.

The ChildHood Integrated Longitudinal Data (CHILD) System at the Center on Urban Poverty and Community Development (CUPCD) is among the most complete integrated data systems (IDS) available for children within a county. The CHILD System is composed of linked administrative records of children in Cuyahoga County, Ohio, beginning with the 1992 birth cohort to the present. The linkage of records across time and systems is performed via deterministic and probabilistic matching techniques. The records contain geographic information that enables aggregation to the neighborhood, city, county or other jurisdiction level and also allows linking with other data systems at various levels of geographies such as parcel, address, or census tracts. The end result is a longitudinal data system in which children are observed if and when they are served by one of the 16 administrative systems that compose CHILD.

The CHILD System began in the late 90's as part of the Cuyahoga County Invest in Children (IIC) initiative¹. IIC is a community-wide, public-private partnership of government leaders and agencies, non-profit organizations and local foundations. The purpose of IIC is to assure that all the County's young children and their families receive the supports they need so that they and their families, communities, and schools are ready for them to enter kindergarten and succeed. Since no one agency or system within the initiative can achieve this outcome on their own, the CHILD System was required to support joint planning and the evaluation of collective impact. Since its inception, the CHILD System has gone from comprising information across 7 to 16 administrative systems (see Figure 1).

The CHILD System contains information on the following observable events: birth, infant mortality, home visiting receipt, early childhood mental health service receipt, public assistance

¹ Fischer, R. L., Lalich, N., & Coulton, C. (2008). Taking it to scale: Evaluating the scope and reach of a community-wide initiative on early childhood. *Evaluation and Program Planning*, 31, 199-208.

benefits (e.g., Medicaid, Cash Assistance, SNAP), lead exposure, child welfare involvement, subsidized child care, special needs child care, public pre-school attendance, kindergarten readiness, public school attendance and test scores for children enrolled in public school in several districts, juvenile court involvement, jail entrance, and homelessness (see table 1).

2. Structure, Matching and Updating Procedures

On a regular schedule, new data become available from the various administrative entities. Matching programs are used to link new records over time and across two or more systems, producing what are called Event Tables. Event Tables can always be linked to each other via a child identification number, but may vary in the way time is recorded. Event tables also include a few key variables pertinent to the particular source and anticipated uses of the data (see Figure 2).

To link new data to the CHILD System, the new data are geocoded and standardized². A third-party SAS macro, LinkPro³ performs deterministic and probabilistic matching to determine whether the new records match those already in the system. Variables that are used for matching vary from one dataset to another, but can include child's and mother's names, birth dates, social security numbers, family address, child's race and gender, and Soundex variables for names (phonetic indexing of name by sound to compensate for some of the inconsistencies found in misspelled names). LinkPro applies probabilistic weights to estimate the likelihood that a pair of records from separate files refers to the same individual. The estimated likelihood scores are used to determine two thresholds such that matches with scores above the higher threshold are considered true matches and those below the lower threshold are deemed false. Matches with scores within the thresholds are considered 'possible matches.' These, and other matches that LinkPro tags as 'unresolved' are manually reviewed.

Continuous evaluation of these methods guarantees that the linkages are at acceptable levels of reliability and completeness. After all new data have been matched and the system has been updated, data in the CHILD system are "self-matched" to eliminate possible duplicate records. All children in the CHILD System who have only one ID from one data source, are

² Standardizing the data entails using common abbreviations for addresses, ensuring everything is written in uppercase, and removing punctuation.

³ Roos Leslie L., Walld Randy, Wajda Andre, et al. (1996) Record Linkage Strategies, Outpatient Procedures, and Administrative Data. Medical Care, Vol. 34, No. 6 , pp. 570-582.

matched to the remaining children. The resulting matches identify duplicate records that need to be combined into one record.

3. Population Represented

Representation in the CHILD System depends on the variables of interest (see Figure 3). Any cross section of the data includes all children born in 1992 and later who experienced the event of interest in the specific time period queried. The administrative data can track one-time events such as births, or recurrent events such as home visits; however, the majority of events represented in the CHILD System occur more than once and are aperiodic. As the schedule of appearance in administrative records is highly variable across children, researchers routinely create count variables of service receipt over a specified period of time. For some children, birth certificate data will be the only time they appear in the CHILD System, either because they move out of Cuyahoga County or because they have not experienced any of the events measured in the administrative records. The largest group with near full representation in the CHILD System is that of children who were born in the county, have resided in the county since birth, and attended a public school in the county.

The CHILD System contains records from administrative entities that follow three different calendars. Some administrative records follow a traditional calendar year (for example, public assistance data); others follow an academic year (for example, public school data). The last set of providers maintain records based on the dates of service or the period of involvement in a system.

3.1. ACS, Census comparisons

After a slight increase between the 1990 and 2000 census, the population of children in Cuyahoga County has decreased in recent years, from 347,990 children under 18 in 2000 to 290,262 in 2010 and 274,640 in 2013. The number of births in the county has steadily declined since 1990, from 22,560 in 1990 to 18,895 in 2000 and 15,100 in 2010. Since about 2008, in a given year, the CHILD System includes some type of information for at least half of the children living in the county (see Figure 3), with as many as 58% included in 2013. Most of these children are Medicaid recipients (49% of all children in 2012). The remaining children not receiving Medicaid or other forms of public assistance might be included in the CHILD System through their participation in other programs such as universal pre-kindergarten, attendance at public

schools⁴, early childhood home visiting, early childhood mental health, or special needs child care. A child might also enter the CHILD system because of blood-lead testing, or having been a victim of child maltreatment, been in foster care, or been in juvenile court. Currently, the CHILD System includes almost 100% of the birth records for children born in Ohio and the goal is to expand coverage within and across counties with additional administrative data⁵.

Figure 3 shows the percent of children born in Cuyahoga County who continue to be observed in the CHILD System over the course of their childhood. These percentages do not include children born elsewhere who moved to Cuyahoga County. Up to age 5, about half of the children in each birth cohort (especially the more recent cohorts) continue to be represented in CHILD. This figure is consistent with the fact that the share of county children in public health care programs went from 38% in 2001 to 51% in 2009⁶. Among the older birth cohorts in the system, the representation declines as children get older, dropping to 35% in 2010 for the 1992 birth cohort. This dropoff may be due to migration out of the county, or a decrease in use of public assistance by age 18.

4. Linking the CHILD System to Neighborhood or Address Level Data

The addresses and geocodes in the CHILD System provide a means to link data from CHILD with data from other systems that contain information at the address or neighborhood level (using geocodes such as block group, census tract, or other geography). For example, CUPCD has a free and publicly accessible longitudinal database of social, economic, and property data called NEO CANDO⁷. NEO CANDO allows users to access data for the entire 17 county Northeast Ohio region, at various geographic levels. Neighborhood level variables can be extracted from the NEO CANDO system and appended to individual records based on geocodes. Examples include poverty rate, teen birth rate, foreclosure rate, violent crime counts or rates, or juvenile delinquency rates.

⁴ Currently, CHILD only has a complete set of education records for children in CMSD, but other school systems are in the process of joining.

⁵ A small number of birth records are suppressed each year for legal reasons, so coverage is not 100% for each birth cohort.

⁶ 2010 Ohio Kids Count Report. Children's Defense Fund. July, 2011. Available at <http://www.childrensdefense.org/library/data/2010-ohio-kids-count-report.html>

⁷ NEO CANDO system, Center on Urban Poverty and Community Development, MSASS, Case Western Reserve University (<http://neocando.case.edu>)

Individual addresses can be linked to another CUPCD database, the Neighborhood Stabilization Team integrated parcel information system (NST)⁸. Historical data from the NST include information at the address level on housing type, housing conditions, housing values, land use, mortgage originations, deed transfers, foreclosure filings and completions, vacancies, code violations, demolitions, tax delinquencies, building permits, community development investments and crime reports for all properties in Cleveland, Ohio.

5. Confidentiality and Sharing Identifiable Information

Creating an IDS remains a difficult endeavor despite their growing presence in the field of human services. In particular, the process of accessing data from agencies is governed by federal regulations such as the Privacy Act, HIPAA, and FERPA, as the release of individually-identifiable information is a risk to personal privacy and confidentiality⁹. Federal, state, and local regulations dictate strict limitations as to what identifiable information can be released. Due to university policies, university legal staff and not the CUPCD typically work with lawyers from data contributing agencies to enter into a process of data sharing bound by a legal document, often called a data use agreement (DUA) or memorandum of understanding (MOU), governing acceptable uses of the data. Most IDS, including the CHILD System, also submit protocols and receive approval from an Institutional Review Board (IRB) charged with the protection of human subjects.

The CHILD system is protected following a strict set of procedures. Only a small number of staff who are certified in human subjects protection and have signed oaths of confidentiality work with identifiable records. All work with these records is done on a highly secure server and personal identifiers are stored separately and linked through a random ID. Researchers and analysts, who are also certified by the IRB, only work with deidentified data to produce summary statistics.

⁸ Hirsh, A., Schramm, M., & Coulton, C. (August 2012). Neighborhood Stabilization Team Web Application. Briefly Stated No. 12-04. Center on Urban Poverty and Community Development. Retrieved from: http://blog.case.edu/msass/2012/09/13/Briefly_Stated_No_12-04_NST_Web_App.pdf

⁹ Petrila, John (2011). Legal Issues in the Use of Electronic Data Systems for Social Science Research. Actionable Intelligence for Social Policy (AISP), University of Pennsylvania. Retrieved from: http://www.sp2.upenn.edu/aisp_test/wp-content/uploads/2012/12/0033_12_SP2_Legal_Issues_Data_Systems_000.pdf

6. Limitations

The CHILD System is not a designed longitudinal panel tracking *all* county children at established points in time. As with all integrated administrative data systems, the richness and quality of the data is dependent on the access to- and quality of administrative records, incidence of service use, and mobility patterns of the population. The following limitations should be noted.

First, all data in the CHILD System were originally collected for program purposes rather than research. As a result, the attention to accuracy and reliability are not as high as would be expected for data collected in controlled research settings. Second, the CHILD System does not include data from all providers of a particular service. For example, at this time, CHILD contains Head Start records for children using subsidies but not children whose families pay with private dollars. Third, selection biases, which plague the ability to estimate causal effects, are difficult to model given the ambiguity of non-event observations. For example, with respect to children's participation in early childhood programming, it is important to acknowledge unobserved parental or environmental factors that may have simultaneously influenced enrollment decisions and educational outcomes. However, no sign of enrollment in the data may come from very different decisions. Parents could have placed the child in a private program or simply not have sent the child to school. Moreover, we cannot be sure whether the lack of an event means that it did not happen or that it was not reported. For example, not all child maltreatment or juvenile delinquency events are known to the authorities.

Fourth, every year about 1,000 birth records, approximately 6.7% of all county births are suppressed for legal reasons, most commonly adoptions and paternity disputes. Fifth, given the lag time associated with data extract receipt, data in the CHILD System cannot be used for real time decision making. Rather, analyses highlight trends and patterns in service receipt over time for planning and decision making going forward.

Lastly, many analyses using the CHILD System are performed over proportions or ratios. For example, the proportion of a birth cohort that receives a specific service or the saturation of a public assistance program within a particular geography. Birth certificate data provide one estimate of the size of the birth cohort at the outset, but as the birth cohort ages, migration begins to have an effect. Birth cohorts can be adjusted upward for in-migration and downward for out-migration using child population estimates and projections from the U.S. Census; however, migration, whether in or out accumulates over time. The question remains as to when and how to adjust for migration rates to produce a valid denominator.

7. Examples of the CHILD System in Use

Planning. In 2011, the Sisters of Charity Foundation of Cleveland engaged CUPCD in a planning process to create a Promise Neighborhood in one of Cleveland's most challenged neighborhoods. Data from the CHILD System were used to describe the magnitude of need in the community, which was used to identify players to participate in revitalizing the neighborhood. With a collaborative team in place, data from the CHILD System were used to describe demographic, educational, and social service involvement changes in the neighborhood over time.

Program Improvement. The Ohio Department of Education implemented the 3rd grade reading guarantee beginning in the 2013-2014 school year. Accordingly, students must meet a minimum score on the state reading test to advance to the 4th grade. Using the CHILD System, CUPCD was able to describe the early childhood experiences including birth characteristics, public service receipt, home visiting, preschool enrollment, child welfare involvement, and lead exposure of three groups of students tested during the first year of implementation: 1) students who passed the state reading test and advanced to the 4th grade; 2) students who failed the state reading test during the school year but passed after summer remediation and were ultimately promoted to the 4th grade; 3) students who failed the state reading test during the school year and summer and were retained in the 3rd grade. The purpose of this analysis was to examine the characteristics of these three groups of students to better understand the challenges they faced and identify predictors of nonpassage that could be useful in targeting supports in the early childhood and preschool period.

Program Evaluation. The academic success of children in the early grades has been found to be closely tied to their experiences in early childhood. Although experiences in early childhood, both positive and negative, are vitally important to predict kindergarten readiness, research has been limited by the absence of integrated and longitudinal data covering experiences from birth. Using the CHILD System, the CUPCD examined how early childhood services—home visits and high quality preschool—affected kindergarten readiness and third grade literacy accounting for individual, family, and neighborhood characteristics, social service receipt and residential mobility.

Policy Innovation. The CHILD System was used to inform the County's decision to explore performance-based contracting (Pay for Success or PFS) for financing human services. Through a partnership between government, nonprofit, philanthropic organizations, and private investors, PFS provides immediate capital to implement innovative social programs targeting

entrenched social issues. If successful, governments see cashable savings and a return to investors. Using the CHILD System, the CUPCD identified several high risk groups known to engender high costs for the county. Ultimately, after thorough exploration of service receipt patterns, the County launched Partnering for Family Success to more quickly reunite parents experiencing homelessness with their children who were placed in out-of-home foster care.

8. Future Directions

Since its founding, CUPCD has been committed to finding better solutions to problems associated with urban poverty by using data and research to inform decision making. The CHILD System was initiated to support this purpose. It is clear that data on two-generations, linking children to their parents or guardians and following them into adulthood, would allow for richer analyses. This data system could not only begin to identify key events that are associated with generational poverty but more importantly opportunities for interrupting this cycle over the life course. Two-generation data could also be used to better understand the ripple effect of a major life event throughout a family system. The integration of post-secondary and workforce data into our system would allow tracking early 1990s cohorts into adulthood and linking them to their children's records.

To date, the CHILD System has grown through the careful matching and updating of Event Tables out of which the required analyses have been performed. As the demands for data from the CHILD system grow, it will be necessary to consider the creation of a Master Event Table or panel including direct or derived variables from all systems at once. In this panel, each record would code information at the [child/month/year] level and a geographic location would be associated with each record as long as any one system contained such information for that time period. Once formal policies and procedures have been established for use and governance, this panel could be available to a wider research community. Currently, only researchers from the CUPCD have access to the integrated data. Providing access to others will require negotiating approved uses of data with administrative entities.

Agency or Data Provider	Dataset(s)	Years (and ages)
	Individual Level Data	
Ohio Department of Health	Births	1992-2012
Ohio Department of Health	Deaths	1992-2012
Ohio Department of Health	Blood-lead testing	2000-present (0-6)
Cuyahoga County Job and Family Services	Medicaid eligibility, Food Stamps (SNAP), Cash Assistance	1992-present (starts at birth)
Cuyahoga County Job and Family Services	Subsidized Child Care	1997-present (starts at birth)
Cuyahoga County Children and Family Services	Child Maltreatment, Foster Care	1990-present (0-18)
Help Me Grow	Newborn, Ongoing, and Early Intervention Home Visiting	1999-present (prenatal to kindergarten)
Cleveland Department of Public Health	MomsFirst Home Visiting	2007-present (prenatal to 2)
Cuyahoga County Board of Health	Newborn Home Visiting	2010-present (first months of life)
Starting Point	Special Needs Child Care	2000-present (0-6)
Starting Point	Universal Pre-Kindergarten	2007-present (3-5)
Multiple Providers	Early Childhood Mental Health	2008-present (0-6)
Cleveland Metropolitan School District	Attendance, KRA-L, Proficiency tests	2005-present (pre-k to 12 th)
Cuyahoga County Office of Homeless Services	Homeless services (shelter stays, permanent supportive housing, etc.)	2010-present (0 to adulthood)
Cuyahoga County Sheriff's Office	County Jail Inmates	2002-present (18 and older)
Cuyahoga County Juvenile Court	Juvenile Court Filings	2000-present (10 to 17)
	Provider Level Data	
Starting Point	Family Child Care Homes	1999-present
Starting Point	Annual Child Care Slots and Enrollment	2002-present

Table 1. Sources of individual-level administrative data linked within individuals and over time to conform CHILD.

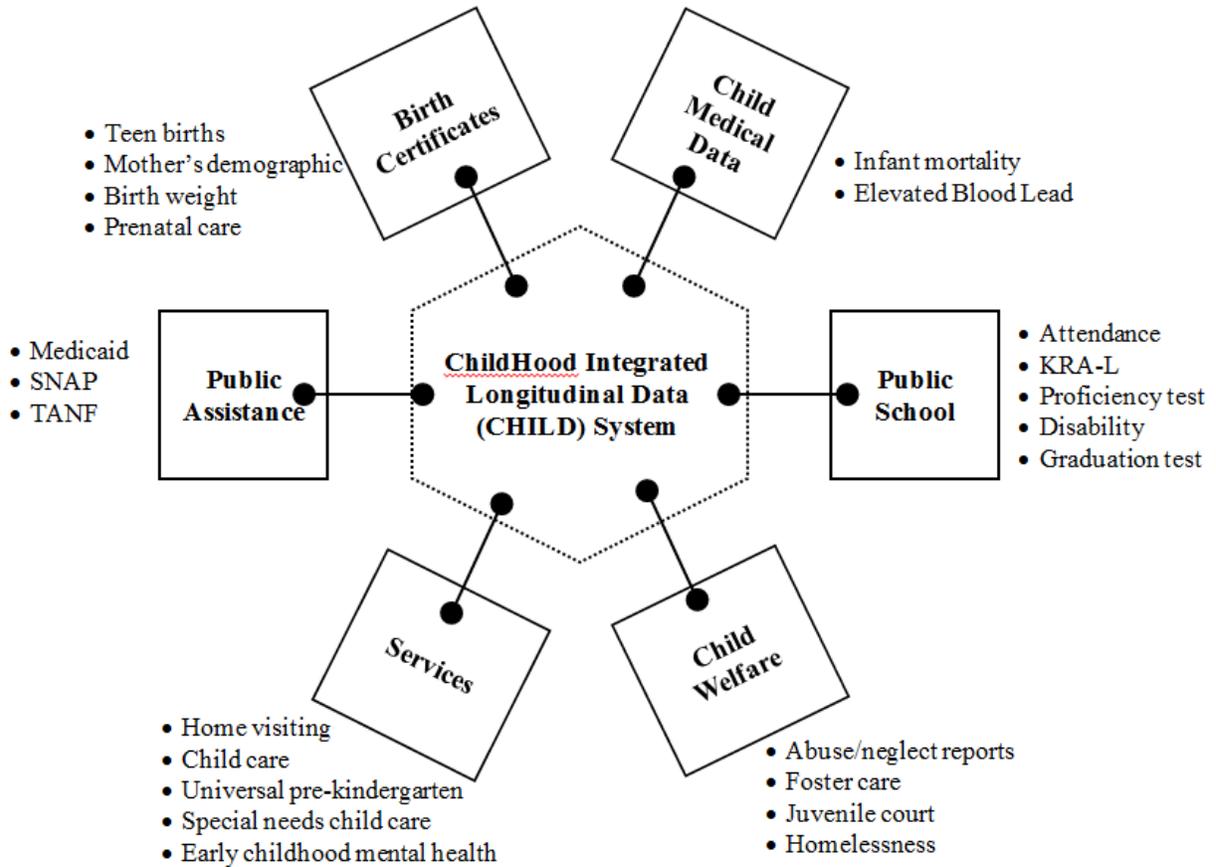


Figure 1. CHILD integrates data from 16 different administrative systems represented in six broad categories: Birth Certificates, Medical Data, Public Assistance, Services, Child Welfare, and the Cleveland Metropolitan Schools.

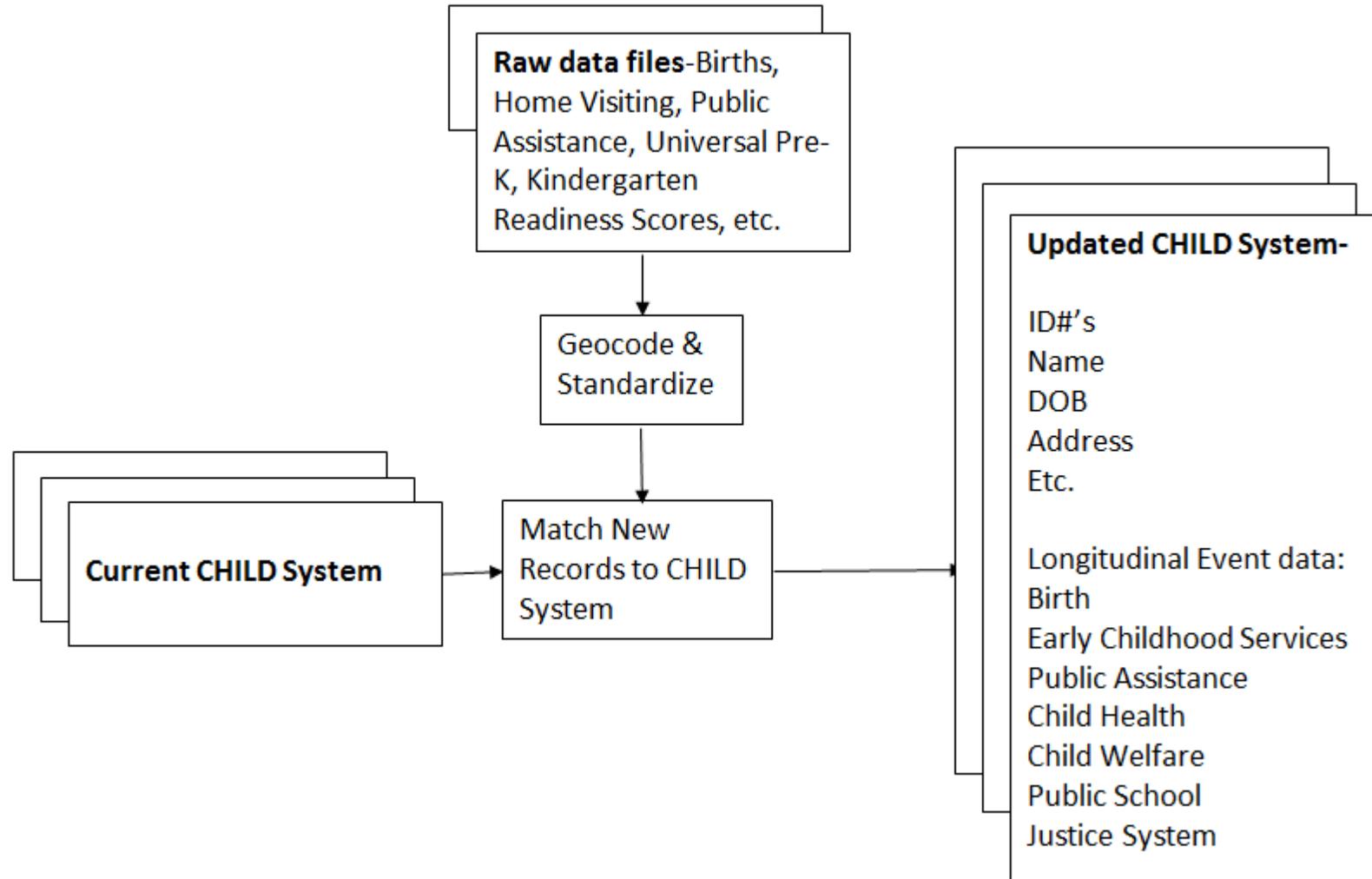


Figure 2. The linking and updating of CHILD produces Event Tables which include variables derived from two or more administrative system (far right layered boxes).

Coverage of CHILDS*											
Number and % observed in:	2000	Main events observed	2004	Main events observed	2008	Main events observed	2010	Main events observed	2013	Main events observed	
1992 Birth Cohort (n=22,087)	7,169 (32%)	PUB, MAL	8,309 (38%)	PUB, MAL	7,907 (36%)	PUB, MAL, SCH, JUV	7,680 (35%)	PUB, MAL, SCH, JAI	5,600 (25%)	PUB, JAI	
2000 Birth Cohort (n=18,895)	18,686 (99%)	BIR, PUB, CC, MAL, HV, LEA**	8,487 (45%)	PUB, CC, MAL, SPN, LEA**	7,648 (40%)	PUB, MAL, SCH	7,486 (40%)	PUB, MAL, SCH	7,075 (37%)	PUB, MAL, SCH, JUV	
2008 Birth Cohort (n=16,246)					15,281 (94%)	BIR, PUB, CC, MAL, HV, LEA, ECM	9,125 (56%)	PUB, CC, MAL, HV, SPN, LEA, HOM, ECM	8,040 (49%)	PUB, CC, MAL, SPN, LEA, UPK, HOM, SCH	
2010 Birth Cohort (n=15,100)							14,091 (93%)	BIR, PUB, CC, MAL, HV, LEA, HOM, ECM	7,667 (51%)	PUB, CC, MAL, HV, SPN, LEA, UPK, HOM, ECM	
Cross Section: All children ages 0-17 living in Cuyahoga County with events in selected year	69,468 out of 347,990 (20%)	BIR, PUB, CC, MAL, HV, SPN, LEA**	117,129 out of 321,071 (36%)	BIR, PUB, CC, MAL, HV, SPN, JUV, LEA**	151,753 out of 300,438 (51%)	BIR, PUB, CC, MAL, HV, SCH, SPN, LEA, UPK, ECM, JUV	160,189 out of 290,262 (55%)	BIR, PUB, CC, MAL, HV, SCH, SPN, LEA, UPK, ECM, JUV, HOM	159,501 out of 274,640 (58%)	BIR, PUB, CC, MAL, HV, SCH, SPN, LEA, UPK, ECM, JUV, HOM	

*CHILD currently includes birth cohorts 1992 and later. It also includes 42,965 children born in 1989-91 for whom selected data was matched for a special study (child maltreatment, Cleveland public school data, homeless services, and juvenile court). Since not all data have been matched for these cohorts (public assistance, child care, etc) they are excluded from this table.

** Lead data for 2000-2004 are available but have not yet been added to CHILDS, and are not included in the percent observed for 2000 and 2004.

Event acronyms:

- BIR=Birth certificate
- CC=Subsidized child care
- ECM=Early Childhood Mental Health
- HOM=Homeless Services
- HV=Home visiting
- JAI=Jail
- JUV=Juvenile court
- LEA=Blood lead testing
- MAL=Child maltreatment & foster care
- PUB=Public assistance (medicaid, SNAP, and cash assistance)
- SCH=Cleveland public school attendance & test scores
- SPN=Special needs child care
- UPK=Universal pre-kindergarten

Figure 3. Representation of Cuyahoga County children in CHILDS for various birth cohorts and cross-sectionally.