BMJ Nutrition, Prevention & Health

Obesity Prevention in Early Life (OPEL) study: linking longitudinal data to capture obesity risk in the first 1000 days

Erika R Cheng , Sami Gharbi, Tammie L Nelson , Sarah E Wiehe

To cite: Cheng ER, Gharbi S, Nelson TL, *et al*. Obesity Prevention in Early Life (OPEL) study: linking longitudinal data to capture obesity risk in the first 1000 days. *BMJ Nutrition, Prevention & Health* 2024;**0**:e000671. doi:10.1136/bmjnph-2023-000671

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjnph-2023-000671).

Department of Pediatrics, Indiana University School of Medicine, Indianapolis, Indiana, USA

Correspondence to
Dr Erika R Cheng;
echeng@iu.edu

Received 26 April 2023
Accepted 16 December 2023

ABSTRACT

To develop robust prediction models for infant obesity risk, we need data spanning multiple levels of influence, including child clinical health outcomes (eg, height and weight), information about maternal pregnancy history, detailed sociodemographic information of parents and community-level factors. Few data sources contain all of this information. This manuscript describes the creation of the Obesity Prevention in Early Life (OPEL) database, a longitudinal, population-based database that links clinical data with birth certificates and geocoded area-level indicators for 19 437 children born in Marion County, Indiana between 2004 and 2019. This brief describes the methodology of linking administrative data, the establishment of the OPEL database, and the clinical and public health implications facilitated by these data. The OPEL database provides a strong basis for further longitudinal child health outcomes studies and supports the continued development of intergenerational linked clinical-public health databases.

INTRODUCTION

Overweight and obesity impact >40 million children under the age of 5.[1 2] The 'first 1000 days' from a woman's pregnancy to her child's second birthday is a critical period for addressing obesity.[3] Numerous risk factors for obesity exist during this time,[4] but little is known about the joint predictive performance of such factors, as population-based datasets often lack sociodemographic data alongside measured heights and weights across pregnancy and early childhood, and birth cohorts may not consistently capture maternal data. Further, while the built environment's influence on childhood obesity is recognised[5–7]; relatively little is known about these relationships because geographical data are often unavailable.

This paper presents the Obesity Prevention in Early Life (OPEL) database, a longitudinal, epidemiological data repository that combines birth certificate, contextual-level and health outcome data for children born in Marion County, Indiana between 2004 and 2019.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ While epidemiological studies have identified numerous risk factors for obesity during pregnancy and early life, risk prediction based on single factors is likely to be incomplete. A better approach would target multiple levels of influence, but existing population-based data sources tend to contain information on maternal risk factors separately from risk factors during infancy and from measures of height and weight across childhood.

WHAT THIS STUDY ADDS

⇒ This paper describes the development of a population-based database containing clinical data linked to children's birth certificates and geocoded area-level indicators created to study determinants of children's obesity risk in the first 1000 days.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The Obesity Prevention in Early Life (OPEL) linked database provides a strong basis for longitudinal studies on children's early-life obesity risk and supports the continued development and use of linked clinical-public health-geographical databases.

⇒ This paper reports on: (1) the construction of the linked OPEL database; (2) the methodological assessment of the linkage and (3) the clinical, epidemiological and public health implications of the linked OPEL database.

METHODS

A complete list of variables available in the OPEL database is presented as online supplemental appendix A.

Data systems

1. The Child Health Improvement through Computer Automation (CHICA) system was a paediatric primary care clinical decision support system that operated in five Indianapolis community health centres from 2004 to 2019.[8] The database contains EHR data and comprehensive information collected from parents using a customised 20-item prescreening form. This information covers measured height and weight,

insurance status, demographics (eg, child sex, age and race/ethnicity) and social factors (eg, parent health literacy, food/housing insecurity, parental depression and infant feeding practices).
2. Information regarding live births in Marion County, Indiana is gathered through the Indiana Standard Certificate of Live Birth (ie, 'birth certificate) and stored at the Marion County Public Health Department (MCPHD). Birth certificate data provide legal and sociodemographic details about the child's mother and father, along with information on prenatal care, labour and delivery, neonatal conditions and procedures, and discharge.
3. The Social Assets and Vulnerabilities Indicators (SAVI) database, sponsored by the Polis Centre, compiles geocoded data on communities in the 11-county Indianapolis metropolitan statistical area. Data are drawn from >30 federal, state and local providers, and are linked to the smallest available geographical level of aggregation. Variables include welfare, education, health, public safety, housing, demographics and locations of health facilities.

Data sharing and security procedures

We developed a secure, multistep data sharing process, initiating with the transfer of files containing limited protected health information variables for linkage. Confidential data from MCPHD and the CHICA team were encrypted and transferred to a protected server using Slashtmp, an Indiana university-specific secure data transfer system. After the record linkage process, matched records were assigned a unique identifier. Core data files retained their unique link identifier and were only linked when generating analytical files.

Data were stored, linked and analysed on a password-protected and encrypted server with two-factor authentication. The server resides on a private IP address behind two layers of firewalls and network monitoring, with no internet access. Access to data is restricted to key personnel on this study.

Inclusion/exclusion criteria and study subjects

CHICA contains data for 63 741 children from birth to age 18, inclusive of 461 201 well-child visits conducted between 2004 and 2019. Focusing on the age group of 0–4 years, we identified 22 689 children with at least 2 well-child visits between 0 and 6 months, excluding those without a Marion County zip code (constituting <5% of the CHICA cohort). A well-child visit was defined based on the utilisation of an appropriate International Classification of Diseases diagnosis code indicative of preventive care.[9] This longitudinal dataset incorporates outcome data at ages 24, 36 and 48 months through sequential records for individual children. The MCPHD dataset contained 274 306 birth certificate records.

Dataset linkage using deterministic and probabilistic record linkage

Record linkage was performed using unique identifiers present across data sources, including social security number, first/middle/last name, first/middle/last initial, phonetic expression of first/last name (Soundex), month/day/year of birth, gender, race/ethnicity, street number of address, zip code (five digits) and phone number. Before linkage, these identifiers were cleaned and standardised, including removal of hyphens and parentheses from social security and phone numbers, elimination of prefixes, suffixes, hyphens, commas and other non-alphabetic characters from names, limiting zip codes to five digits, and standardising race/ethnicity. Frequencies were checked for unexpected observations such as numeric data for name.

Raw data from CHICA (N=22 689) and birth certificate records (N=274 306) were cleaned and coded using Stata/MP V.14.1.

We identified linkage pairs through deterministic matching, probabilistic matching and manual review (online supplemental appendix B).[10] The deterministic matching used a conservative automated approach that included three algorithms to exactly match different combinations of a subset of identifiers. We hand-validated a subset of matches using first/last name and date of birth (DOB), substituting mother's last name and father's last name in addition to the baby's last name. We refined algorithms to achieve the highest and most robust linkage rates. The first deterministic algorithm used last name, first name and DOB. Father's and mother's last names were used in subsequent deterministic algorithms.

We then used RecMatch software to enhance the likelihood of matching CHICA and birth certificate records using probabilistic algorithms.[11] We performed multiple probabilistic matches using a combination of different blocking and matching schemes to capture additional pairs.

The study team (ERC, SG and SEW) performed manual review using conservative thresholds for acceptance to identify true matches. True positives were determined by manual review, the gold standard.[12] Records representing the same individual were connected via transitive property across pairwise matches (eg, A=B and B=C would connect A, B and C), and assigned a unique study identifier.

Geocoding and geotagging procedures

We then linked individual-level health data from CHICA and birth certificates to SAVI. We cleaned (eg, removing apartment numbers, PO boxes, and nonsensical addresses) and geocoded addresses from the birth certificate. We then used ArcGIS[13] to geotag children's addresses with SAVI data at various levels (eg, street, 50 m buffer, zip code).

Starting with 17 712 unique addresses, we conducted geocoding in two runs: first, using street address and zip code and then street only. Run parameters included spelling sensitivities of 80, minimum candidate scores of 10 and minimum match scores of 80; parameters for Polis were set to 80/10/85, respectively.

Table 1 Sequential deterministic and probabilistic linking for the 22 689 records in Chica against 274 306 Marion County birth certificates

Matching phase	Type of match	# Linked	Cumulative # Linked	Cumulative % Linked	Remaining # Unlinked
1	Deterministic: LN, FN, DOB	17 240	17 240	76.0%	5449
1	Deterministic: Father LN, FN, DOB	350	17 590	77.5%	5099
1	Deterministic: Mother LN, FN, DOB	582	18 172	80.1%	4517
2	Probabilistic variations	1471	19 643	86.6%	3046
3	Manual review	214	19 857	87.5%	2832

DOB, date of birth; FN, first name; LN, last name.

RESULTS

Individual-level data linkage

We linked 87.5% (N=19 857) of patient records between data sources (table 1). Our deterministic algorithms linked 80.1% of all records; 94.9% of which were matched with the first algorithm using last name, first name and DOB. Subsequent algorithms contributed a few additional matches, 350 (1.8% of total records) and 582, respectively (2.9% of total records), with more success matching on mothers' names than on fathers'.

Probabilistic matching linked an additional 1471 records. We used 18 probabilistic algorithms, with the highest yield matching on last or first name, Soundex of last or first name and sex, when blocked by last or first name and DOB (online supplemental appendix B). Our manual review linked 214 more records.

Figure 1 presents the proportion of new matched pairs by linkage phase.

Contextual-level data geocoding and geotagging

We geocoded 92.5% (N=16 396) of unique addresses, or 94.5% of our cohort of babies (N=19 437), with 88.0% successfully geocoded in the first run using street address and zip code and an additional 4.5% geocoded in the second run. Remaining linkages (by transitive property) were determined via an in-house daisy chain process in

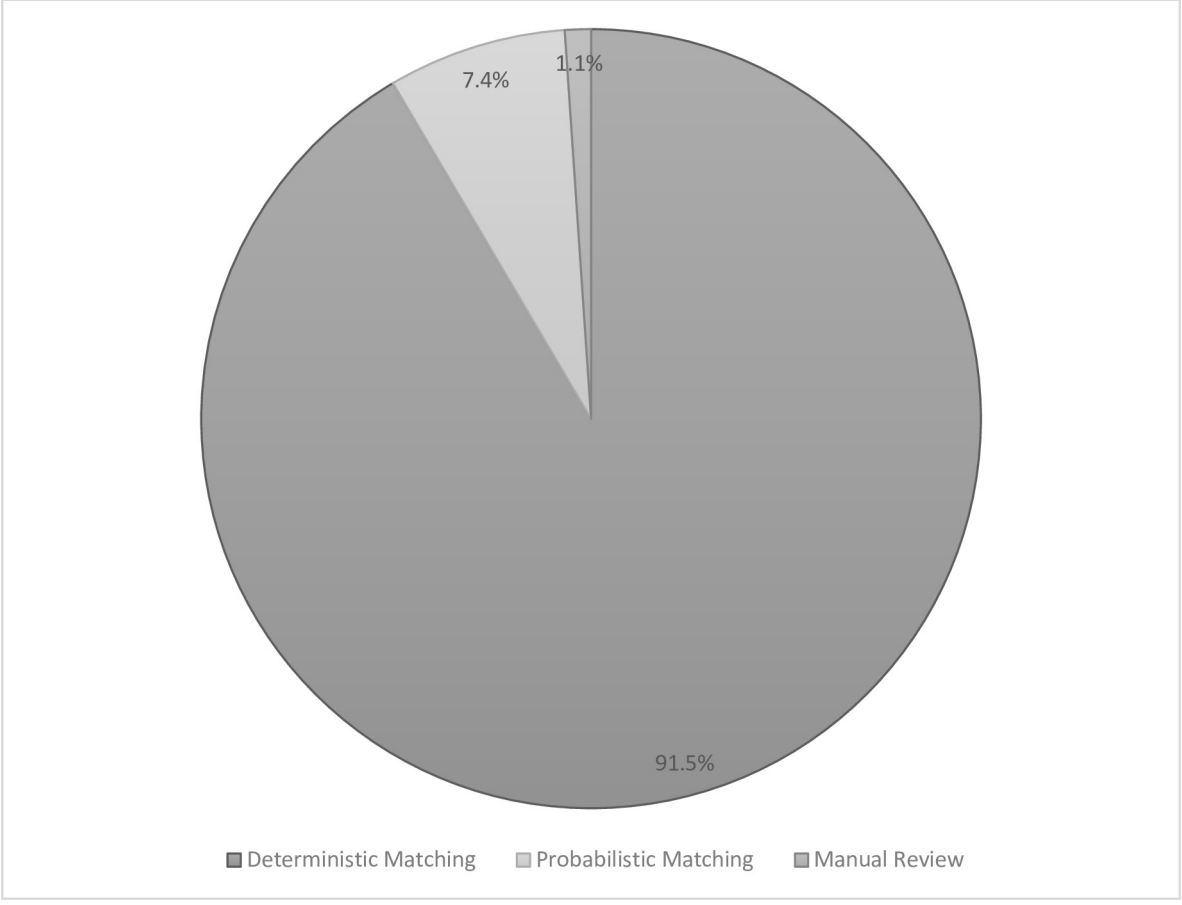

Figure 1 Proportion of matches from deterministic, probabilistic and manual record linkage processes.

STATA. The number of phases was determined by evaluated incremental increases in the number of pairs.

DISCUSSION

This paper introduces the OPEL database, a novel and comprehensive longitudinal data repository. This database, born out of the necessity to address the critical period of the first 1000 days, integrates birth certificate, contextual-level and health outcome data for children in Marion County, Indiana, born between 2004 and 2019. Linking data from CHICA, birth certificates and SAVI enables a multidimensional analysis, covering sociodemographic, clinical and geographic factors. Notably, the study achieves an impressive 87.5% linkage of clinical records and 95% successful geocoding, providing a robust foundation for future investigations into childhood obesity aetiology and related outcomes. If replicated in different states or health systems, this approach could offer valuable insights into the interplay of factors influencing obesity, informing research,[14 15] public health interventions and programming.

Applying our algorithms to different states and health systems may yield different linkage rates. Limitations include constraints of existing data (eg, missingness, data quality), potential misclassification from self-reported, and omission of other correlates of early child health (eg, genetics, paternal factors) not collected from these sources. Attrition may cause selection bias.

Strengths include the use of three matching processes to link more records than prior approaches and our use of primary data collection, EHR and contextual data sources. OPEL spans the first 1000 days and employs replicable methods, making it valuable for investigators in other geographic areas where similar data linkages are possible. OPEL serves as a foundation for additional longitudinal studies that link maternal and paternal information, public health programming and other contextual data that will allow us to comprehensively examine the aetiology of childhood obesity and to track obesity-related outcomes.

Contributors ERC helped conceptualise the study, interpreted data and drafted the manuscript. SG carried out the analyses and interpretation of results. TLN helped with interpretation of data and drafting of the manuscript. SEW provided feedback on the research question, analysis plan and interpretation of the data. All authors critically reviewed the manuscript for important intellectual content and approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Funding This research was funded by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (K01DK114383).

Disclaimer The sponsors had no role in the study design; collection, analysis and interpretation of data; writing of report; or decision to submit for publication.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures were approved by the Indiana University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed by Dr. Emma Derbyshire, Nutritional Insight, UK.

ORCID iDs

Erika R Cheng http://orcid.org/0000-0001-8289-5087
Tammie L Nelson http://orcid.org/0000-0003-4204-7555

REFERENCES

1 Friedrich MJ. Global obesity epidemic worsening. *JAMA* 2017;318:603.
2 Collaborators GO. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med* 2017;377:13–27.
3 Brown T, Moore TH, Hooper L, *et al*. Interventions for preventing obesity in children. *Cochrane Database Syst Rev* 2019;7:CD001871.
4 Woo Baidal JA, Locks LM, Cheng ER, *et al*. Risk factors for childhood obesity in the first 1,000 days: a systematic review. *Am J Prev Med* 2016;50:761–79.
5 Papas MA, Alberg AJ, Ewing R, *et al*. The built environment and obesity. *Epidemiol Rev* 2007;29:129–43.
6 Dunton GF, Kaplan J, Wolch J, *et al*. Physical environmental correlates of childhood obesity: a systematic review. *Obes Rev* 2009;10:393–402.
7 Lovasi GS, Hutson MA, Guerra M, *et al*. Built environments and obesity in disadvantaged populations. *Epidemiol Rev* 2009;31:7–20.
8 Anand V, Biondich PG, Liu G, *et al*. Child health improvement through computer automation: the CHICA system. *Stud Health Technol Inform* 2004;107:187–91.
9 Texas Children's Health Plan. HEDIS quick reference for well-child visits. Secondary HEDIS quick reference for well-child visits. 2020. Available: https://www.texaschildrenshealthplan.org/sites/default/files/pdf/PR-2005-028_HEDIS_Wellchild.pdf
10 Dusetzina SB, Tyree S, Meyer A-M, *et al*. Linking data for health services research: a framework and instructional guide. Rockville, MD Agency for Healthcare Research and Quality (Prepared by the University of North Carolina at Chapel Hill under Contract no.290-2010-000141); 2014.
11 Grannis S, Egg J, Ribeka N. RecMatch: probabilistic patient record matching; 2008.
12 Grannis SJ, Overhage JM, McDonald CJ. Analysis of Identifier performance using a deterministic linkage algorithm. Proceedings / AMIA… Annual Symposium. AMIA Symposium; 2002:305–9
13 Esri Inc. ArcGIS pro 2.8 version. program; 2021.
14 Cheng ER, Cengiz AY, Miled ZB. Predicting body mass index in early childhood using data from the first 1000 days. *Sci Rep* 2023;13:8781.
15 Cheng ER, Steinhardt R, Ben Miled Z. Predicting childhood obesity using machine learning: practical considerations. *BioMedInformatics* 2022;2:184–203.

Appendix A. Complete list features available in the OPEL database, by data source

Name	**Description**	**Data Source**
weight	child's weight at visit	CHICA
wtcentile	child's weight percentile	CHICA
height	child's height at visit	CHICA
htcentile	child's height percentile	CHICA
insurance	What kind of insurance, if any, the patient has at time of visit	CHICA
any_household_members_smoke	Do any of the people that live with the child smoke?	CHICA
car_seat_position_01	Does the child use a car seat, and if so, which way is it facing?	CHICA
fluoride_supplemented	Does the child have fluoride supplemented somehow through consumption?	CHICA
has_smoke_detector	Does the child's living area have a smoke detector?	CHICA
hc	child's head circumference in centimeters	CHICA
hccentile	child's head circumference percentile	CHICA
know_how_to_save_choking_child	Do the child's caregivers know how to perform the Heimlich maneuver on a choking child?	CHICA
left_alone_in_water	Is the child left alone in water?	CHICA
lg_failed	What question of the language developmental test did the child fail on?	CHICA
maternal_depression_concern	Based on a questionnaire, is there a concern that the mom might be depressed?	CHICA
medicationallergies	Does the child have any medication allergies and have the allergies been confirmed by a doctor or only reported by the family?	CHICA
painqualitative	Is the child in pain, yes, no or NA?	CHICA
ps_passed	What is the highest passed question for the psychosocial developmental test?	CHICA
sleeps_on_side_or_back	Does the child sleep on their side or back?	CHICA
slept_on_stomach_ever	Does the child ever sleep on their stomach?	CHICA
uses_walker	Does the child use a walker?	CHICA
baby_left_alone_could_fall	Is the baby ever left alone where they could fall?	CHICA
sleeps_unsafe_soft_surface	Does the child sleep on an unsafe soft surface such as a mattress that they can suffocate on if they sleep facedown?	CHICA
tested_smoke_detector	If the child's living place has a smoke detector, has it been tested as working?	CHICA
abdomen_exam	If the child's abdomen is examined, is it abnormal or normal?	CHICA
back_exam	If the child's back is examined, is it abnormal or normal?	CHICA
chestlungs_exam	If the child's chest or lungs are examined, is it abnormal or normal?	CHICA
extgenitalia_exam	If the child's external genitalia is examined, is it normal or abnormal?	CHICA
extremities_exam	If the child's extremities (hands, feet, nose, ears) are examined, are they normal or abnormal?	CHICA
fm_passed	For the fine motor skills developmental test, what is the highest question passed?	CHICA
general_exam	If the child had a general exam, was it normal or abnormal?	CHICA

gm_passed	For the gross motor skills developmental test, what is the highest passed question?	CHICA
head_exam	If the child's head is examined, is it normal or abnormal?	CHICA
heartpulses_exam	If the child's heart and pulse are examined, is it normal or abnormal?	CHICA
lg_passed	For the language developmental test, what is the highest scoring passed question?	CHICA
neuro_exam	If a neurological battery is done, was it normal or abnormal?	CHICA
nodes_exam	If the lymph nodes are checked, were they normal or abnormal?	CHICA
nosethroat_exam	If the nose and throat are examined, are they normal or abnormal?	CHICA
skin_exam	If the child's skin is examined, was it normal or abnormal?	CHICA
teethgums_exam	If the child's teeth and gums are examined, were they normal or abnormal?	CHICA
preferred_language	Does the child have a preferred language and if so, is it English or Spanish?	CHICA
burns_knowledge	Does the caregiver have knowledge of how to take care of burns?	CHICA
firearms_at_home	Are there any firearms in the home?	CHICA
firearms_where_visits	Are there any firearms where the visit is taking place?	CHICA
has_stairway_gates	Are there child safety gates over the stairways?	CHICA
household_products_out_of_reach	Are household cleaning products such as bleach out of the reach of children?	CHICA
matches_lighters_safe	Are matches and lighters kept in a safe manner? childproof wheel, out of reach, etc.	CHICA
play_area_fenced	Is the child's play area fenced in?	CHICA
pool_at_house	Is there a pool the child can access?	CHICA
chica_devscreen_status	This is a developmental screening that states whether the child is developing normally or if they are developmentally delayed and indicate which developmental screenings have been done.	CHICA
seen_dentist	Has the child ever been seen by a dentist? This is unlikely to be true until after the child has teeth.	CHICA
taking_medications	Is the child on any medications and if so, has this list of medications been confirmed to be accurate?	CHICA
tv_in_room	Is there a TV in the child's bedroom?	CHICA
tv_over_2hrs	Does the child watch TV for more than two hours every day?	CHICA
uses_bottle	Does the child use a bottle to eat?	CHICA
asthmastatus	Does the child have any asthma symptoms and if so, are they persistent, intermittent, uncontrolled/controlled?	CHICA
chica_devscreen_sx	Are there any developmental concerns?	CHICA
lye_drain_cleaners_in_house	Are there any lye, drain, or other more dangerous cleaners in the house?	CHICA
ps_failed	What question of the psychosocial test did the child fail on?	CHICA
stop_at_curb	Does the child stop at curbs or run straight without stopping?	CHICA
wears_bike_helmet	Does the child wear a bike helmet for activities where one is recommended?	CHICA
insurancename	What kind of insurance does the child have?	CHICA
parents_confident_filling_out_	Do the parents appear confident filling out forms?	CHICA

parents_need_help_reading	Do the parents need help reading forms?	CHICA
ten_childrens_books_in_home	Are there at least 10 children's books in the home available to the child?	CHICA
visittype	Is this a visit because the child is sick?	CHICA
chica_adhd_sx	Is the child having symptoms of ADHD?	CHICA
constipation_sx	Is the child having symptoms of constipation?	CHICA
firearms_kept_unloaded	Are any firearms kept unloaded in the household?	CHICA
look_both_ways	Does the child look both ways before crossing the street?	CHICA
unsupervised_near_water	Is the child left unsupervised near water?	CHICA
firearms_discussed	Has firearm safety been discussed with the child?	CHICA
grades_dropped_lately	Has the child's school grades dropped recently?	CHICA
knows_how_to_swim	Does the child know how to swim?	CHICA
rides_bike_in_street	Does the child ride their bike in the street?	CHICA
school_suspension_this_year	Has the child been suspended from school this year?	CHICA
snoring	Have parents noticed that the child snores?	CHICA
special_education_classes	Does the child attend special education classes?	CHICA
escape_plan_for_fire	Has the family discussed a house fire escape plan with their child? Older children version of smoke_alarm_knows_what_to_do	CHICA
informant	What household member is answering the questions?	CHICA
smoke_alarm_knows_what_to_do	Does the child know what to do when the smoke/fire alarm is triggered? Younger children version of escape_plan_for_fire	CHICA
specialneeds	Does the child have special needs or accomodations? Such as ear defenders, speech therapist, etc...	CHICA
visit_attendee	What household member is attending the visit but not necessarily the informant?	CHICA
hot_water_heater_adjusted	Has the water heater been adjusted so the water can only be heated to 120 degrees farenheit? This is a scalding concern.	CHICA
plastic_wrappers_secured	Are plastic wrappers in the environment secured or left in an accessible area? This is a suffocation hazard.	CHICA
taking_solid_food	Is the child eating solid food yet?	CHICA
cutting_food_bite_size	Are the child's solid foods being cut into bite size pieces before being given to the child? If no, this is a choking/suffocation hazard.	CHICA
carries_hot_liquids	Is the child allowed to carry hot liquids? This is a burn hazard.	CHICA
play_area_cooking	Does the child have an area to play and be safely in away from cooking area while caregiver is cooking? This is a burn risk if not.	CHICA
safety_latches_installed	Have safety latches been installed in the house?	CHICA
car_seat_inspection	Has the child's car seat been inspected and if so, is it forward or rear facing? Rear facing is the safer option.	CHICA
developmental_referral	Has the child been referred to developmental testing and if so, have only the first steps been taken or has the appointment been made?	CHICA
fm_failed	What difficulty of the fine motor skills test did the child fail on?	CHICA
correctedvision	Does the child wear glasses or contact lenses?	CHICA
firearms_friends	Does the child go to friend's houses which have firearms?	CHICA
plays_dangerous_items	Does the child play with dangerous items?	CHICA
wears_sports_protective_gear	Does the child wear protective gear while playing sports?	CHICA
safety_caps_on_bottles	Are there child safety caps on pill bottles around the child?	CHICA

wears_life_jacket	Does the child wear a life jacket in situations where that is recommended?	CHICA
bedtime_media	Does the child use media products at bedtime?	CHICA
daytime_sleepiness	Is the child sleepy during the day?	CHICA
questionnaireinformants	Which caregiver filled out the questionnaire?	CHICA
sleep_quantity	Does the child get sufficient or insufficient sleep?	CHICA
chica_t2dm_fh	Does the child's medical records include family history?	CHICA
chica_t2dm_gdm	Did the child's mother have gestational diabetes?	CHICA
chica_t2dm_lga	Was the child large for their gestational age during pregnancy?	CHICA
epilepsy_history	Is there a family-reported family history of epilepsy?	CHICA
breast_feeding_help_needed	Does the mother need help breastfeeding?	CHICA
oral_exam	Has the child's mouth been examined and if so, was it normal or abnormal?	CHICA
bp_eval	Has the child's blood pressure been evaluated and if so, was it elevated once or repeatedly elevated? There was no option for hypotensive in this variable.	CHICA
empty_container_after_use	Do caregivers empty bathwater container immediately after use? This is a drowning risk if no.	CHICA
well_water	Does the child's household run off well-water? Well-water is a contamination concern.	CHICA
lowliteracyrisk	Is the child at risk of low literacy and if so, have they gone to a clinic to help?	CHICA
morning_headaches	Does the child have headaches in the morning or wake up with a headache?	CHICA
nocturnal_enuresis	Does the child wet the bed/pee during sleep? This question is for kids who are out of diapers.	CHICA
stops_breathing_at_night	Does the child's caregiver know if the child stops breathing during the night?	CHICA
trouble_breathing_at_night	Does the child's caregiver know if the child has trouble breathing during the night?	CHICA
wakes_with_snort	Does the child caregiver know if the child wakes up with a snort?	CHICA
rides_after_dark	Does the child ride the bike after sunset?	CHICA
knows_rules_of_road	Does the child know traffic rules?	CHICA
swims_fast_moving_water	Does the child swim in fast-moving water such as a river?	CHICA
chica_adhd_dx	Does the child have an ADHD diagnosis?	CHICA
doors_secure	Are the doors in the child's home secure?	CHICA
sharp_edged_furniture	Are there sharp-edged furniture in the child's home?	CHICA
pulseox	What was the child's pulse oxygenation percentage at visit?	CHICA
has_window_guards	Does the child home have window guards?	CHICA
play_equipment_protected	Does the child play on safe playground equipment?	CHICA
asthmasymptoms	Does the child have symptoms of asthma?	CHICA
gm_failed	What gross motor test did the child fail on?	CHICA
chica_adhd_side_effects	Does the child experience side effects from their ADHD medication?	CHICA
irondeficiencyscreenqualitativ	Has the child been checked for iron deficiency and if so, what were the results?	CHICA
chica_devscreen_management	Is the child part of activities specifically made for children?	CHICA
normal_newborn_screen	Did the child have the normal newborn screen and if so, what were the results?	CHICA

vaccine_given	Has the child had the HPV, Tdap, or meningococcal vaccine given?	CHICA
anhedonia_past_few_weeks	Has the child been anhedonic/apathetic the last few weeks?	CHICA
cigarettes_snuff_friend	Does the child's friend or friend's household use cigarettes or snuff?	CHICA
cigarettes_snuff_live_with	Does someone the child lives with use snuff?	CHICA
ever_use_tobacco	Has the child ever used tobacco?	CHICA
has_drunk_alcohol	Has the child drunk alcohol at all?	CHICA
has_gotten_high	Has the child used an illicit substance?	CHICA
has_had_forced_sex_act	Has the child experienced a forced sex act?	CHICA
has_had_intercourse	Has the child had intercourse?	CHICA
sad_past_few_weeks	Has the child been sad in the past few weeks?	CHICA
suicide_concerns	Is there a concern of suicidality for the child?	CHICA
used_marijuana	Has the child used marijuana?	CHICA
interested_birth_control	Is the child interested in contraception?	CHICA
ready_to_quit	Is the child ready to quit smoking cigarettes?	CHICA
watches_tv	Does the child watch TV?	CHICA
sleep_problems	Does the child have problems sleeping?	CHICA
nobp	Child did not cooperate in visit; Could not check blood pressure.	CHICA
nohearing	Child did not cooperate in visit; Could not perform hearing exam.	CHICA
risk_based_hearing_screen	Has the child undergone a hearing screen that was ordered based on high risk?	CHICA
chica_devscreen_treatment	Does the child have a written care plan or access to family support services?	CHICA
anxiety_status	Does the child have an anxiety diagnosis, or has this questionnaire been deferred?	CHICA
phq9_score	What was the mother's depression score on the phq9?	CHICA
driven_with_drunk	Has the child driven while drunk?	CHICA
drunk_and_activity	Has the child been drunk while doing an activity?	CHICA
drunk_last_month	Has the child been drunk in the last month?	CHICA
family_substance_abuse	Does the child's family abuse any substances?	CHICA
happy_how_things_going	Is the child happy with life?	CHICA
uses_drugs	Does the child use drugs?	CHICA
sudep_risk_counseling	Is the child at risk for sudden unexpected death from epilepsy? If so, is the risk high or low?	CHICA
surgical_hx	Has the child had their tonsils and adenoids removed?	CHICA
feed_at_night	Does the child eat at night?	CHICA
contraceptive_method_discussed	Has birth control been discussed with the child such as condoms and hormonal birth control?	CHICA
abuse_otc	Does the child abuse over the counter medication?	CHICA
abuse_steroids	Does the child abuse steroids drugs?	CHICA
criticized_for_drinking	Has the child been criticized for drinking?	CHICA
friends_use_drugs	Has the child's friends used drugs (other than alcohol/caffeine) in the last month?	CHICA
friend_drunk_last_month	Has the child's friends been drunk in the last month?	CHICA
fun_in_past_two_weeks	Does the child think they've had fun in the last two weeks?	CHICA
bike_has_coaster_brakes	Does the child's bike have coaster brakes? Coaster brakes allow you to pedal backwards to brake.	CHICA
past_depression_or_suicide	Has the child had any previous history of depression or suicidality?	CHICA
immune_compromise	Is the child immuno-compromised?	CHICA

prescription_for_cessation	Is the child on a prescribed nicotine replacement drug?	CHICA
intercourse_past_year	Has the child had intercourse in the last year?	CHICA
might_be_pregnant	Could the child be pregnant?	CHICA
medication	Does the child have a Ritalin prescription?	CHICA
depression_workup	Is there a developed safety plan for the child's depression?	CHICA
chica_autism_risk	Is the child at a higher risk of autism due to family history?	CHICA
tooth_erupted	Has the child had a tooth erupt from beneath the gums yet?	CHICA
autism_behavior_problems	Does the child have autism related behavior problems?	CHICA
autism_cam	Does the child use complementary alternative medicine for autism?	CHICA
autism_financial_concerns	Are there financial concerns related to the child's autism such as paying for therapy?	CHICA
autism_parent_needs_respite	Is the child's caregiver in need of a break? i.e. showing symptoms of caregiver burnout	CHICA
patient_in_mental_health	Is the child undergoing mental health care?	CHICA
food_insecurity	Is the child's caregiver worried about getting enough food and if so, has this been MD confirmed or resolved?	CHICA
rental_status	Is the child's rental home clean & safe vs having issues, and has this been confirmed by an MD?	CHICA
snapdeniedlast30days	Has the child's SNAP(food stamps) been denied in the last 30 days?	CHICA
utility_status	Has the child's household had one of their utilities (water, power, heat, gas) shut off? Yes, no, or yes but not heat.	CHICA
mlp_condition_type	Is the child's family going through an eviction, on the SNAP program, or renting?	CHICA
wakes_up_one_or_more_times_a_n	Does the child wake up at least once during the night?	CHICA
wakes_up_and_needs_help_to_sleep	Does the child wake up at night and need help getting back to sleep?	CHICA
sleeps_on_back	Does the child sleep on their back?	CHICA
slept_on_stomach_side_ever	Does the child ever sleep on their stomach or side?	CHICA
abuse_concern	Is there a concern that the child is being abused?	CHICA
constipation_dx	Has the child been diagnosed with constipation?	CHICA
parent_thinks_child_has_sleep_pr	Do the caregivers think that the child has problems with their sleep?	CHICA
eyesvision_exam	Did the child have a normal or abnormal vision exam?	CHICA
breastfed	Is the child being breastfed at this time?	CHICA
psfsicklecell	Result of pre-screening form on tablet for sickle cell anemia.	CHICA
negativeenvironmentalhistory	Was the child potentially exposed to something negative in their environment such as tuberculosis or lead?	CHICA
negativenutritionhistory	Did the child have nutrition problems such as early introduction to cow milk or needing low iron formula?	CHICA
negativepastmedicalhistory	Did the child have a low birth weight?	CHICA
cholesterol_screen	Is the child at risk of high cholesterol based on parental history?	CHICA
earshearing_exam	Did the child have a normal or abnormal hearing exam?	CHICA
hearingleft	Does the child have full or partial hearing in their left ear?	CHICA
hearingright	Does the child have full or partial hearing in their right ear?	CHICA
ppd_result	What was the result of the mother's post-partum depression assessment?	CHICA
venousbloodleadqualitative	How much lead was in the child's blood, if tested?	CHICA

mother_bmi	Maternal body mass index	MCPHD
PNC_Clinic_Type	Type of prenatal care clinic	MCPHD
Sex	Child's sex	MCPHD
FATHER_OCCUP_DSCRP	Is child's father employed at time of birth?	MCPHD
MomNativeAm	Is child's mother Native American?	MCPHD
Mother_Weight_Gain_P	How many pounds the mother has gained during pregnancy.	MCPHD
MARRIED_NOW	Are child's parents married at time of birth?	MCPHD
APGAR5	Appearance, Pulse, Grimace, Activity, and Respiration at five minutes post birth. Score of 10 is good; one is bad.	MCPHD
BIRTH_WEIGHT_GRAM	Birth weight in grams from modern birth certificate	MCPHD
finalroute	How was the child delivered?	MCPHD
HEP_B_TEST	Was hepatitis B vaccine given at birth?	MCPHD
Apgar1	Appearance, Pulse, Grimace, Activity, and Respiration at 1 minute post birth. Score of 10 is good; one is bad.	MCPHD
Dad_Race9Eth	race of child's father	MCPHD
Mom_Race9Eth	race of child's mother	MCPHD
PREN_VISIT_NBR	number of prenatal care visits	MCPHD
EST_GEST	estimated gestation in weeks	MCPHD
MOTHER_AGE	age of the mother at birth in years	MCPHD
FATHER_AGE	age of the father at birth in years	MCPHD
PREVIOUS_LIVE_NBR	How many living babies has the mother giving birth to before?	MCPHD
plurality	Is this a plural or singleton birth? (twins)	MCPHD
BREAST_FED	Was the child breast-fed at hospital release?	MCPHD
MOTHER_ED	mother's education level in years	MCPHD
FATHER_ED	father's education level in years	MCPHD
LD_MECONIUM	delivery complication: was there meconium present at delivery?	MCPHD
LD_NONE	no delivery complications	MCPHD
LD_NON_VERTEX	delivery complication: child in non- vertex position	MCPHD
firstpnc	prenatal care initiated in first trimester	MCPHD
wtgrams	child's birth weight in grams	MCPHD
PREV_BIRTH_TOTAL	number of previous live births – all birth certificates	MCPHD
Kotelchuck	adequacy of prenatal care index	MCPHD
mdpsmoke	Did the mother smoke during pregnancy?	MCPHD
abcond	Were abnormal conditions present at birth?	MCPHD
anomaly	Was a congenital anomaly found?	MCPHD
infect	maternal infections	MCPHD
labdel	labor and delivery	MCPHD
mmorb	maternal morbidity	MCPHD
methdel	method of delivery	MCPHD
oblab	obstetrical labor	MCPHD
obproc	obstetrical procedures	MCPHD
risk	maternal risk factor	MCPHD
RACE	race of the child	CHICA
ETHN	ethnicity of the child	CHICA
wic_ever	Has the child ever been in the WIC program?	CHICA/ MCPHD
PERINPOVN1	persons living in poverty as percentage of population	SAVI
VIOLENTN2	violent crime (including simple assaults) per 1,000 people	SAVI
VIOLNSTN2	violent crime (not including simple assaults) per 1,000 people	SAVI
AGGVASLTN2	aggravated assaults per 1,000 people	SAVI

ROBBERYN2	robberies per 1,000 people	SAVI
PROPERTYN2	property crime per 1,000 people	SAVI
THFTVHN2	vehicle thefts per 1,000 people	SAVI
BURGLARYN2	burglaries per 1,000 people	SAVI
WALKSCORE	walkability score	SAVI
FRRDTRAN1	free and reduced lunch program participants as percentage of enrollment	SAVI
POVB185N1	population below 185% poverty (proxy for reduced lunch)	SAVI
POVB125N1	population below 125% poverty (proxy for free lunch)	SAVI
RESNEWPEN1	total residential building permits per 100 housing units	SAVI
COMMALLPN1	total commercial building permits per 100 housing units	SAVI
TREE_CANOPY	tree canopy as percentage of land area	SAVI
PCT_POP_FOOD_DESERT	percentage of population far from grocery stores	SAVI

Appendix B. Deterministic and probabilistic matching elements

ALGORITHM		BLOCKING VARIABLES	MATCHING VARIABLES	THRESHOLD	# MATCHED BABIES (INCREMENTAL)
DETERMINISTIC MATCHING	DM1	LN/FN/DOB			17,240
	DM2	LN=Father_LN/FN/DOB			350
	DM3	LN=Mother_LN/FN/DOB			582
PROBABILISTIC MATCHING	PM1	LN/DOB	FN/FN_Soundex/SEX	-31	523
	PM2	FN/DOB	LN/LN_Soundex/SEX	-26	795
	PM3	DOB	LN/FN/MN/LN_Soundex/FN_Soundex/SEX	7	35
	PM4	YB/FI	LN/FN/MN/MB/DB/LN_Soundex/FN_Soundex/SEX	13	46
	PM5	MB/FI	LN/FN/MN/YB/DB/LN_Soundex/FN_Soundex/SEX	19	12
	PM6	DB/FI	LN/FN/MN/YB/MB/LN_Soundex/FN_Soundex/SEX	19	3
	PM7	(LN=Mother_LN)/DOB	FN/FN_Soundex/SEX	-1	29
	PM8	FN/DOB	(LN=Mother_LN)/LN_Soundex/SEX	11	17
	PM9	DOB	(LN=Mother_LN)/FN/MN/LN_Soundex/FN_Soundex/SEX	23	1
	PM10	YB/FI	(LN=Mother_LN)/FN/MN/MB/DB/LN_Soundex/FN_Soundex/SEX	12.3	-
	PM11	MB/FI	(LN=Mother_LN)/FN/MN/YB/DB/LN_Soundex/FN_Soundex/SEX	12.3	-

	PM12	DB/FI	(LN=Mother_LN)/FN/MN/YB/MB/LN_Soundex/FN_Soundex/SEX	12.67	-
	PM13	(LN=Father_LN)/DOB	FN/FN_Soundex/SEX	-6	3
	PM14	FN/DOB	(LN=Father_LN)/LN_Soundex/SEX	-51	7
	PM15	DOB	(LN=Father_LN)/FN/MN/LN_Soundex/FN_Soundex/SEX		-
	PM16	YB/FI	(LN=Father_LN)/FN/MN/MB/DB/LN_Soundex/FN_Soundex/SEX		-
	PM17	MB/FI	(LN=Father_LN)/FN/MN/YB/DB/LN_Soundex/FN_Soundex/SEX	12	-
	PM18	DB/FI	(LN=Father_LN)/FN/MN/YB/MB/LN_Soundex/FN_Soundex/SEX		-
			Manual Review		214
				Total	19,857