Published in final edited form as:
Matern Child Health J. 2014 January ; 18(1): 316–325. doi:10.1007/s10995-013-1236-7.

Development of a Linked Perinatal Data Resource From State Administrative and Community-Based Program Data

Eric S. Hall,
Perinatal Institute and Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH, USA

Neera K. Goyal,
Perinatal Institute, Cincinnati Children's Hospital, Cincinnati, OH, USA

Robert T. Ammerman,
Every Child Succeeds, Cincinnati Children's Hospital, Cincinnati, OH, USA

Megan M. Miller,
Perinatal Institute, Cincinnati Children's Hospital, Cincinnati, OH, USA

David E. Jones,
Biostatistics and Epidemiology, Cincinnati Children's Hospital, Cincinnati, OH, USA

Jodie A. Short, and
Every Child Succeeds, Cincinnati Children's Hospital, Cincinnati, OH, USA

Judith B. Van Ginkel
Every Child Succeeds, Cincinnati Children's Hospital, Cincinnati, OH, USA

Eric S. Hall: eric.hall@cchmc.org; Neera K. Goyal: neera.goyal@cchmc.org; Robert T. Ammerman: robert.ammerman@cchmc.org; Megan M. Miller: megan.miller_2@cchmc.org; David E. Jones: david.jones@cchmc.org; Jodie A. Short: Jodie.Short@cchmc.org; Judith B. Van Ginkel: judith.vanginkel@cchmc.org

Abstract

To demonstrate a generalizable approach for developing maternal-child health data resources using state administrative records and community-based program data. We used a probabilistic and deterministic linking strategy to join vital records, hospital discharge records, and home visiting data for a population-based cohort of at-risk, first time mothers enrolled in a regional home visiting program in Southwestern Ohio and Northern Kentucky from 2007 to 2010. Because data sources shared no universal identifier, common identifying elements were selected and evaluated for discriminating power. Vital records then served as a hub to which other records were linked. Variables were recoded into clinically significant categories and a cross-set of composite analytic variables was constructed. Finally, individual-level data were linked to corresponding area-level measures by census tract using the American Communities Survey. The final data set represented 2,330 maternal-infant pairs with both home visiting and vital records data. Of these, 56 pairs (2.4 %) did not link to either maternal or infant hospital discharge records. In a 10 % validation subset (n = 233), 100 % of the reviewed matches between home visiting data and vital records were true matches. Combining multiple data sources provided more comprehensive details of perinatal health service utilization and demographic, clinical, psychosocial, and behavioral characteristics than available from a single data source. Our approach offers a template for leveraging disparate sources of data to support a platform of research that evaluates the timeliness and reach of home visiting as well as its association with key maternal-child health outcomes.

Correspondence to: Eric S. Hall, eric.hall@cchmc.org.

Dr. Hall and Dr. Goyal are co-first authors who contributed equally to this work.

Keywords

Home visiting; Early childhood development; Data linking

Background and Motivation

Perinatal data resources, including state vital records, hospital discharge data, and records from community-based programs like home visiting, each contain unique elements important for maternal-child health research, and yet these data sets lack interoperability. Previous models of linked databases, such as the Massachusetts's Pregnancy to Early Life Longitudinal (PELL) system and other statewide efforts in California and Pennsylvania, utilize a linked maternal-child data core. Core elements originating from vital statistics and hospital discharge records are generated for the entire population and include maternal and child demographics, diagnoses, and outcomes. These elements may serve as a core for linkage to subsequent data sets representing subpopulations with local program data [1–4]. While such linkages have enabled maternal-child health research for a range of state-level outcomes, additional research is needed to further leverage information collected by community-based programs like home visiting.

While the effectiveness of home visiting has been demonstrated for many outcomes, including infant development and parenting, the impact on preterm birth is currently not well understood. The utility of a maternal-child data core derived from administrative records can be enhanced by capturing detailed measures of prenatal program use including gestational age at enrollment, intensity of participation, and content of visits. Such further conceptualization and measurement of home visiting may be particularly important for preterm birth, where factors such as nutrition and health behaviors may only be amenable to intervention if exposure begins early and is sustained at a sufficiently high intensity.

The current aim is to demonstrate a generalizable approach to supplement state administrative records with local program data through the development of a maternal-child health resource. This paper outlines the process of linking data from home visiting to statewide records (birth and death vital records, hospital discharge records) and area-based measures of health. Our rationale is that linked data will expand the potential for evaluating population-level outcomes, enabling more detailed, hypothesis-driven analyses of perinatal outcomes in a real world setting. Given recent investments in home visiting through the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program, [5] population-based, linked systems are particularly relevant for evaluating the impact and timeliness of home visiting services provided to at-risk families.

Methods

Study Design and Population

This retrospective, descriptive analysis includes a population of women enrolled in Every Child Succeeds (ECS), an established, regional home visiting program serving Greater Cincinnati who gave birth at an Ohio hospital in 2007–2010. Since its inception in 1999, ECS has conducted over 400,000 home visits with 17,000 families in Southwestern Ohio and Northern Kentucky. Eligible participants must be first-time mothers with at least 1 of 4 risk characteristics: unmarried, low income (<300 % of poverty level, receipt of medicaid, or reported financial concerns), <18 years of age, or suboptimal prenatal care. Participants enroll during pregnancy or before their child reaches 3 months of age. Home visits are provided by social workers, child development specialists, nurses, or para-professionals, beginning weekly or more-frequent and tapering to fewer visits as the child ages until 3

years of age. A web-based data entry system is used to bill and document service provision data; home visitors collect detailed information for each maternal-infant pair, including content and frequency of home visits, household demographics, and screening instruments including the home observation for measurement of the environment (HOME), Kempe Family Stress Inventory, and Ages and Stages Questionnaires [6–9]. Participants enrolled in ECS have consented to data use for the purposes of quality assurance, benchmarking, and research. This study has resulted from collaboration and partnership between stakeholders including care providers, public health agencies, and local home visiting programs which have established common aims and data sharing agreements. The study was reviewed and approved by the Ohio Department of Health (ODH) and Cincinnati Children's Hospital Medical Center Institutional Review Boards.

A maternal-child health data resource was generated by linking individual home visiting records with statewide vital and Ohio Hospital Association (OHA) records and census tract data from the American Community Survey (ACS) using established linkage strategies [10–12]. Contributions of each data set are listed in Table 1. The process included the following steps, further detailed below: (a) Source data were consistently formatted in preparation for linking; (b) Variables from each source were prescreened and evaluated for "discriminating power" to identify the set best supporting individual-level record linkage; (c) Individual-level data sources were linked using a probabilistic and deterministic approach; (d) Unresolved links were manually resolved; (e) Variables were recoded into clinically significant categories; (f) Individual-level records were linked to corresponding area-level measures; (g) A cross-set of composite analytic variables was constructed.

Data Preparation

Vital records obtained from the ODH represented 597,000 live births within Ohio during 2007–2010. Approximately 6.5 million hospital discharges were represented by the OHA data for the corresponding time period. Newborn infants were identified by limiting OHA records to those with a birth date matching the date of hospital admission. In addition to 611,345 newborn records, 620,929 records representing maternal hospitalizations corresponding to birth events were identified by restricting the OHA set to records listing delivery related Diagnosis Related Group (DRG) codes (370–384). We identified 3,902 women enrolled in ECS who gave birth in 2007–2010. Of those, 2,476 (63.4 %) listed an Ohio residence while 1,426 (36.5 %) listed a Kentucky residence. Because a considerable number of mothers from Northern Kentucky give birth in Ohio hospitals and receive an Ohio birth certificate, all women included in the home visiting database were eligible for linkage.

Records shared no universal identifier for linking; instead, common identifying elements were selected and formatted to facilitate linkage. Formatting of names, dates, and zip codes was standardized and insurance status, infant gender, delivery route, and preterm birth status were consistently encoded across data—relevant OHA DRG codes indicated delivery route (370–375) and preterm birth status (386–391).

Prescreening

Elements in each set were evaluated for discrimination power—a calculated score indicating usefulness in distinguishing between records. For example, unique medical record numbers enable perfect discrimination among records. Conversely, a variable for which all records share the same value (e.g. birth state within a single state's vital records) cannot be used to discriminate between records. Discrimination power was calculated using the LINKS record linkage package, a SAS™ (SAS Institute Inc., Cary, NC, USA) macro developed at the University of Manitoba [13]. Table 2 presents the variable combinations used in linking

various individual-level data sets to the vital record (which served as a linkage hub to which all other records were matched) sorted in order of variable discriminating power.

Individual Records

The LINKS tool implemented probabilistic and deterministic linkage by evaluating common data elements to calculate likelihood of matches. Upon execution, variables with the greatest discriminating power were required to agree (deterministic) while other common elements contributed to a match likelihood score (probabilistic). Record pairs were deemed matched when a likelihood score threshold was exceeded. For linkage between home visiting and vital records, either maternal or infant birth date values were required to agree between all matched pairs. For linkage between vital and OHA records, birth hospital was required to agree. Additionally, infant birth date on the vital record was required to agree with birth date on the matched OHA newborn record and required to fall between the dates of admission and discharge on the matched OHA maternal record. A set of records with medium-to-high likelihood scores falling below the match threshold was produced for manual review. Frequent causes for inconclusive matches were inconsistent name spelling, address disagreement, and transposed numbers within dates. Under manual review such discrepancies were easily reconciled and comprised less than 2 % of the final data set.

The final mother-centric perinatal data resource represented each woman with a linked ECS-vital record pair. For multiple gestation pregnancies a single home visiting record potentially linked to multiple vital records. In these cases, data from all mothers were retained, but only data from the firstborn infant were included in the final set. For validation, a 10 % random sample of linked records (generated using SAS) was subjected to manual expert review to determine accuracy of matches by examining agreement between identifying elements.

Hospital Discharge Data

Indicator variables for relevant co-morbidities, risk factors, and complications were derived from the OHA data using International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes; categorization was modeled from previous studies utilizing similar data [14, 15]. Each hospital record contained a primary diagnosis, up to 14 additional diagnoses, a primary procedure, and up to 8 additional procedures. “Appendix” lists codes used to categorize hospital discharge variables.

Area-Level Measures

Geocoding ECS addresses produced latitude–longitude coordinate pairs for each observation. We linked individual records to five-year data estimates from the 2010 American Community Survey (ACS) containing aggregate sociodemographic measures by area of residence at the census tract level [16]. Area-level measures were selected on the basis of theoretical relevance for a range of maternal-child health outcomes and on the basis of previous empirical research [17].

Results

The final data set represented 2,330 women with linked home visiting and vital record pairs. Basic demographics and enrolment timing are described in Table 3. Although birth certificates belonging to Ohio residents delivering out-of-state were unavailable for linkage; the final set contained 2,183 records listing an Ohio residence representing 88.2 % of the Ohio-based (ECS) population during the study period. For each of the 2,330 linked home visiting and vital record pairs, an attempt was made for linkage to two OHA records—one representing maternal hospitalization and one representing infant hospitalization. Of these linked pairs, 88.9 % were able to be linked to a maternal OHA discharge record; 85.9 %of

the linked pairs were able to be linked to an infant OHA discharge record; 2.4 %were not linked to either OHA discharge record. In the 10 % validation subset (n = 233), 100 % of the reviewed matches between home visiting and vital records were determined to be true matches.

Analytic Variables

Table 4 presents final analytic variables as constructed through a combination of data derived from home visiting, vital records, and OHA data sets. The final linked set represented a broad range of measures including demographics, clinical and psychosocial indicators, behavioral risk factors, measures of home environment and program utilization, and birth outcomes.

Discussion

Home visiting represents an early opportunity to improve pregnancy and child health outcomes through care coordination, education, and social support in at-risk populations [18–20]. A particular focus for investment in home visiting through the MIECHV program is prenatal care management and risk reduction for preterm birth, which occurs with profound sociodemographic and geographic disparity [5, 21, 22]. Research focused on community-based prevention of preterm birth remains challenging due to the complexity of biological, cultural, and socioeconomic risk factors associated with this outcome [23]. Our study demonstrates how combining data obtained through home visiting with a core of data from vital and hospital discharge records provides a data platform supporting perinatal epidemiologic research. These methods can serve as a template for developing similar regional data resources as states and communities implement or expand home visiting.

Previous strategies linking vital records using sociodemographic variables have similarly found match rates above 80 % [24]. Other statewide linked resources have enabled investigations for a range of maternal-child outcomes [3, 4, 25–27]. The PELL system, which uses the same LINKS tool utilized in the current study, has demonstrated the utility of linked maternal-child data in Early Intervention program evaluation [1, 2, 27]. The current study builds upon previously established capabilities by supplementing population-level data with maternal information obtained through home visiting and with area-based measures of health. This detailed data resource integrates previously disparate measures enabling innovative epidemiological research involving clinical, social, and community factors which contribute to adverse perinatal outcomes.

An example of the enhanced utility of this data resource is the ability to evaluate of the effect of engagement level, or treatment dosage, in prenatal home visiting. Specifically, we will test the hypothesis that preterm birth and birth weight are associated with both timing of prenatal enrollment by weeks of gestation and frequency of home visits, adjusting for maternal risk factors. This dataset will also enable inclusion of covariates such as the percent living below poverty level by census tract, an important area-level measure associated with health outcomes including preterm birth [28]. Further applications of area-level measures may include the use of an ecological framework to test the effect of community characteristics on program retention, level of engagement, and program effectiveness [29].

Limitations

Previous analyses have identified deficiencies in the completeness, accuracy, and content of research data containing administrative codes such as ICD-9-CM codes [30]. Limitations in the number of coded diagnoses and procedures may crowd out relevant codes among highly complicated patients or may be biased toward maximizing financial reimbursement [31–34].

We attempted to mitigate the impact of undercoding by generating, whenever possible, composite variables from multiple sources. Future use of this data may require validation of measures like maternal complications and congenital anomalies using length of stay, procedures, and co-diagnoses to minimize misclassification. Previous vital record validation efforts have demonstrated that such data elements, when reported, are highly accurate and concordant with independently generated clinical data [35–38]. As an example of preliminary validation in our dataset, of 108 maternal-infant pairs identified as having chorioamnionitis, over 75 % were associated with a co-diagnosis of maternal or infant infection, treatment with maternal antibiotics, labor induction in the setting of fetal distress, or prolonged hospitalization and neonatal intensive care use.

The final data set also reflects some limitations of source data. Although composite variables reduced the effects of missing source data, missingness was highest among variables originating from a single data set, including body mass index (7.96 %missing in vital records), household income(15.46 % missing in home visiting data), and infant length of stay (14.27 % missing based on hospital discharge data). Finally, an important consideration in development of community-based, locally-linked data is the limitation of population and program size that may impact generalizability.

Conclusions

Population-based, linked data systems are essential for evaluating the impact and reach of home visiting to at-risk families, a relevant topic given federal investments through MIECHV. The current study demonstrates a generalizable approach for integrating community-based program data with state data including vital and hospital discharge records to enable hypothesis-driven analysis pertinent to perinatal outcomes. Following this model, researchers elsewhere may develop similar resources that leverage available data and inform stakeholders as they tailor policies to local maternal-child populations.

Acknowledgments

Dr. Goyal was supported by the BIRCWH K12 program, co-funded by the Office of Research on Women's Health and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Award Number 5K12HD051953-07. Dr. Ammerman was supported by Grant R01MH087499 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not represent the official views of the NICHD, the NIMH, or the NIH. The authors acknowledge support of the United Way of Greater Cincinnati, Kentucky H.A.N.D.S., and Ohio Help Me Grow, and technical assistance from Ted Folger.

References

1. Barfield WD, Clements KM, Lee KG, et al. Using linked data to assess patterns of early intervention (EI) referral among very low birth weight infants. Maternal and Child Health Journal. 2008; 12(1): 24–33.10.1007/s10995-007-0227-y [PubMed: 17562149]
2. Clements KM, Barfield WD, Kotelchuck M, et al. Birth characteristics associated with early intervention referral, evaluation for eligibility, and program eligibility in the first year of life. Maternal and Child Health Journal. 2006; 10(5):433–441.10.1007/s10995-006-0080-4 [PubMed: 16710765]
3. Herrchen B, Gould JB, Nesbitt TS. Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. Computers and Biomedical Research. 1997; 30(4):290–305. [PubMed: 9339323]
4. Srinivas SK, Fager C, Lorch SA. Evaluating risk-adjusted cesarean delivery rate as a measure of obstetric quality. Obstetrics and Gynecology. 2010; 115(5):1007–1013.10.1097/AOG.0b013e3181d9f4b6 [PubMed: 20410776]

5. Health Resources and Services Administration. [Accessed November 1, 2012] Maternal, infant, and early childhood home visiting program. 2012. http://mchb.hrsa.gov/programs/homevisiting/index.html
6. Ammerman RT, Putnam FW, Kopke JE, et al. Development and implementation of a quality assurance infrastructure in a multisite home visitation program in Ohio and Kentucky. J Prev Interv Community. 2007; 34(1–2):89–107.10.1300/J005v34n01_05 [PubMed: 17890195]
7. Bradley RH, Caldwell BM. Home observation for measurement of the environment: A validation study of screening efficiency. American Journal of Mental Deficiency. 1977; 81(5):417–420. [PubMed: 66872]
8. Kempe, RS.; Kempe, CH. Child abuse. Cambridge, Mass: Harvard University Press; 1978.
9. Squires, J.; Bricker, DD.; Twombly, E. Ages and stages questionnaires: A parent-completed child monitoring system. 3. Baltimore: Paul H. Brooks Pub. Co; 2009.
10. Black, C.; Roos, LL. Linking and combining data to develop statistics for understanding the population's health. In: Friedman, DJ.; Hunter, EL.; Parrish, RG., editors. Health statistics: Shaping policy and practice to improve the population's health. New York, NY: Oxford University Press; 2005. p. 214-240.
11. Roos LL, Wajda A. Record linkage strategies. Part I: Estimating information and evaluating approaches. Methods of Information in Medicine. 1991; 30(2):117–123. [PubMed: 1857246]
12. Roos LL Jr, Wajda A, Nicol JP. The art and science of record linkage: Methods that work with few identifiers. Computers in Biology and Medicine. 1986; 16(1):45–57. [PubMed: 3948494]
13. Manitoba Centre for Health Policy. [Accessed October 2, 2012] Concept: LINKS: A record linkage package. 2012. http://mchpappserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1029
14. Goyal NK, Fager C, Lorch SA. Adherence to discharge guidelines for late-preterm newborns. Pediatrics. 2011; 128(1):62–71.10.1542/peds.2011-0258 [PubMed: 21690121]
15. Lorch SA, Baiocchi M, Ahlberg CE, et al. The differential impact of delivery hospital on the outcomes of premature infants. Pediatrics. 2012; 130(2):270–278.10.1542/peds.2011-2820 [PubMed: 22778301]
16. U.S. Census Bureau. [Accessed November 12, 2012] American community survey. 2012. http://www.census.gov/acs/www/
17. Krieger N, Chen JT, Waterman PD, et al. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The public health disparities geocoding project (US). Journal of Epidemiology and Community Health. 2003; 57(3):186–199. [PubMed: 12594195]
18. American Academy of Pediatrics Council on Child and Adolescent Health. The role of home-visitation programs in improving health outcomes for children and families. Pediatrics. 1998; 101(3 Pt 1):486–489. [PubMed: 9499196]
19. Kitzman H, Olds DL, Henderson CR Jr, et al. Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated child-bearing. A randomized controlled trial. JAMA, the Journal of the American Medical Association. 1997; 278(8):644–652. [PubMed: 9272896]
20. Sweet MA, Appelbaum MI. Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. Child Development. 2004; 75(5):1435–1456.10.1111/j.1467-8624.2004.00750.x [PubMed: 15369524]
21. Muglia LJ, Katz M. The enigma of spontaneous preterm birth. The New England Journal of Medicine. 2010; 362(6):529–535.10.1056/NEJMra0904308 [PubMed: 20147718]
22. Kramer MR, Hogue CR. Place matters: Variation in the black/white very preterm birth rate across U.S. metropolitan areas, 2002–2004. Public Health Reports. 2008; 123(5):576–585. [PubMed: 18828412]
23. Behrman RE, Butler AS, Institute of Medicine (U.S.). Preterm birth: Causes, consequences, and prevention. Washington, D.C: National Academies Press; 2007. Committee on understanding premature birth and assuring healthy outcomes.
24. Herman A, McCarthy B, Bakewell J, et al. Data linkage methods used in maternally-linked birth and infant death surveillance data sets from the United States (Georgia, Missouri, Utah and

Washington State), Israel, Norway, Scotland and Western Australia. Paediatric and Perinatal Epidemiology. 1997; 11(S1):5–22.10.1046/j.1365-3016.11.s1.11.x [PubMed: 9018711]

25. Declercq E, Barger M, Cabral HJ, et al. Maternal outcomes associated with planned primary cesarean births compared with planned vaginal births. Obstetrics and Gynecology. 2007; 109(3): 669–677.10.1097/01.AOG.0000255668.20639.40 [PubMed: 17329519]
26. Manning SE, Davin CA, Barfield WD, et al. Early diagnoses of autism spectrum disorders in Massachusetts birth cohorts, 2001–2005. Pediatrics. 2011; 127(6):1043–1051.10.1542/peds.2010-2943 [PubMed: 21576313]
27. Shapiro-Mendoza CK, Tomashek KM, Kotelchuck M, et al. Risk factors for neonatal morbidity and mortality among “healthy”, late preterm newborns. Seminars in Perinatology. 2006; 30(2):54–60. [PubMed: 16731277]
28. DeFranco EA, Lian M, Muglia LA, et al. Area-level poverty and preterm birth risk: A population-based multilevel analysis. BMC Public Health. 2008; 8:316.10.1186/1471-2458-8-316 [PubMed: 18793437]
29. McGuigan WM, Katzev AR, Pratt CC. Multilevel determinants of retention in a home-visiting child abuse prevention program. Child Abuse and Neglect. 2003; 27(4):363–380. [PubMed: 12686322]
30. Cheung NT, Fung V, Chow YY, et al. Structured data entry of clinical information for documentation and data collection. Studies in Health Technology and Information. 2001; 84(Pt 1): 609–613.
31. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Medical Care. 1994; 32(1):81–90. [PubMed: 8277803]
32. Iezzoni LI, Foley SM, Daley J, et al. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? JAMA, the Journal of the American Medical Association. 1992; 267(16):2197–2203. [PubMed: 1556797]
33. Hsia DC, Krushat WM, Fagan AB, et al. Accuracy of diagnostic coding for medicare patients under the prospective-payment system. The New England Journal of Medicine. 1988; 318(6):352–355.10.1056/NEJM198802113180604 [PubMed: 3123929]
34. Simborg DW. DRG creep: A new hospital-acquired disease. The New England Journal Of Medicine. 1981; 304(26):1602–1604.10.1056/NEJM198106253042611 [PubMed: 7015136]
35. Reichman NE, Hade EM. Validation of birth certificate data. A study of women in New Jersey’s HealthStart program. Annals of Epidemiology. 2001; 11(3):186–193. [PubMed: 11248582]
36. Reichman NE, Schwartz-Soicher O. Accuracy of birth certificate data by risk factors and outcomes: analysis of data from New Jersey. American Journal of Obstetrics and Gynecology. 2007; 197(1):32, e1–8.10.1016/j.ajog.2007.02.026 [PubMed: 17618747]
37. DiGiuseppe DL, Aron DC, Ranbom L, et al. Reliability of birth certificate data: A multi-hospital comparison to medical records information. Maternal and Child Health Journal. 2002; 6(3):169–179. [PubMed: 12236664]
38. Vinikoor LC, Messer LC, Laraia BA, et al. Reliability of variables on the North Carolina birth certificate: A comparison with directly queried values from a cohort study. Paediatric and Perinatal Epidemiology. 2010; 24(1):102–112.10.1111/j.1365-3016.2009.01087.x [PubMed: 20078836]

Appendix

See Table 5.

Table 1

Data categories and elements included in the linked data resource by each contributing source data set

	Vital records	Hospital discharge	Home visiting program	American Community Survey[a]
Demographics	X	X	X	
Maternal substance use	X	X	X	
Household and parental assessments			X	
Maternal comorbidities		X		
Complications of pregnancy	X	X		
Complications of labor and delivery	X	X		
Maternal body mass index	X			
Infant gestational age	X			
Infant birth weight	X			
Infant congenital anomalies	X	X		
Postnatal complications	X	X		
Infant mortality	X	X		
Hospital charges and length of stay		X		
Home visiting program utilization			X	
Area-based measures by census tract				X

[a] Available as five-year estimates through the U.S. Census Bureau

Table 2

List of data elements used in linking the various individual-level data sets, sorted in order of variable discriminating power

Rank based on discriminating power	Description of variable	Data sets containing the linkage variable (in addition to vital records)
1	Mother's date of birth	Home visiting program, hospital discharge—maternal
2	Street number of residence	Home visiting program
3	Mother's last name	Home visiting program
4	Infant's last name	Home visiting program
5	Mother's maiden name	Home visiting program
6	Street name of residence	Home visiting program
7	Infant's date of birth	Home visiting program, hospital discharge—maternal, hospital discharge—infant
8	Mother's first name	Home visiting program
9	Zip code of residence	Home visiting program, hospital discharge—maternal, hospital discharge—infant
10	Birth hospital	Hospital discharge—maternal, hospital discharge—infant
11	County of residence	Home visiting program
12	Insurance status	Home visiting program, hospital discharge—maternal, hospital discharge—infant
13	Infant gender	Home visiting program, hospital discharge—infant
14	Delivery route	Hospital discharge—maternal
15	Preterm birth status	Hospital discharge—infant

Table 3

Demographic characteristics and enrollment timing of women with linked home visiting and vital records data

	Among total population (N = 2,330)	%
Ohio residence	2,183	93.7
Kentucky residence	147	6.3
Prenatal enrollment	1,128	48.4
Enrolled during 1st trimester of pregnancy	134	5.8
Enrolled during 2nd trimester of pregnancy	521	22.4
Maternal age (years)		
< 18	580	24.9
18–25	1,493	64.1
25–35	231	9.9
Maternal race		
African American	1,330	57.1
White	928	39.8
Hispanic ethnicity	173	7.4
Unmarried	2,163	92.8
Maternal education		
< High school diploma or equivalent	1,066	45.8
High school completion	638	27.4
Some college	502	21.6
Medicaid or other public insurance	1,737	74.5
Enrolled in WIC	1,790	76.8

Table 4

Select variables, origin data sources, and percent missingness in the final combined data resource

Description of variable	Data source	Percent missing (%)
Birth weight	Vital Statistics	0.26
Cervical abnormalities	Hospital Discharge Data	0.00
Chorioamnionitis	Vital Records Hospital Discharge Data	0.00
Cleft lip or palate	Vital Records Hospital Discharge Data	0.00
Congenital cardiac anomaly	Vital Records Hospital Discharge Data	0.00
Congenital pulmonary anomaly	Vital Records Hospital Discharge Data	0.00
Congenital spinal anomaly	Vital Records Hospital Discharge Data	0.00
Maternal diabetes	Vital Records Hospital Discharge Data	0.00
Fetal Distress	Vital Records Hospital Discharge Data	0.00
Gestational age at birth	Vital Records	0.05
Gestational age at enrollment	Vital Records Home Visiting Program	0.00
Home visiting agency	Home Visiting Program	0.00
Home visit frequency	Home Visiting Program	0.00
Home visiting referral source	Home Visiting Program	0.00
Household income	Home Visiting Program	15.46
Human immunodeficiency virus	Hospital Discharge Data	0.00
Induction of labor	Vital Records Hospital Discharge Data	0.00
Infant age at death	Hospital Discharge Data Vital Records	0.00
Infant length of stay	Hospital Discharge Data	14.27
Infant mechanical ventilation	Vital Records Hospital Discharge Data	0.00
Insurance type	Vital Records Hospital Discharge Data	0.31
Intrauterine growth retardation	Hospital Discharge Data	0.00
Intraventricular hemorrhage	Hospital Discharge Data	0.00
Living arrangement	Home Visiting Program	8.12
Marital status	Vital Records Home Visiting Program	0.21
Maternal age	Vital Records	0.00
Maternal body mass index	Vital Records	7.96
Maternal contact with baby's father	Home Visiting Program	9.88
Maternal education	Vital Records Home Visiting Program	0.26
Maternal ethnicity	Vital Records Home Visiting Program	0.00
Maternal hypertension	Vital Records	0.00

Description of variable	Data source	Percent missing (%)
	Hospital Discharge Data	
Maternal hypothyroidism	Hospital Discharge Data	0.00
Maternal race	Vital Records Home Visiting Program	0.00
Method of delivery	Vital Records Hospital Discharge Data	0.00
Multiple gestation pregnancy	Vital Records	0.00
Necrotizing enterocolitis	Hospital Discharge Data	0.00
NICU admission	Vital Records	0.00
Oligohydramnios	Hospital Discharge Data	0.00
Percent below poverty level, by census tract[a]	American Community Survey	0.00
Placental disorders	Hospital Discharge Data	0.00
Preeclampsia	Hospital Discharge Data	0.00
Preterm labor	Hospital Discharge Data	0.00
Previous fetal loss, stillbirth or neonatal death	Vital Records Hospital Discharge Data	0.00
Premature rupture of membranes	Vital Records Hospital Discharge Data	0.00
Retinopathy of prematurity	Hospital Discharge Data	0.00
Sexually transmitted infections	Vital Records Hospital Discharge Data	0.00
Substance use—alcohol	Home Visiting Program Hospital Discharge Data	0.00
Substance use—cigarettes	Home Visiting Program Vital Records Hospital Discharge Data	0.00
Substance use—other drugs	Home Visiting Program Hospital Discharge Data	0.00
Total infant hospital charges	Hospital Discharge Data	15.05
Townsend index score[a,b]	American Community Survey	0.00

Clinical variables from hospital discharge and/or vital records data were constructed as binary indicator variables based on the presence of available data to confirm the diagnosis

[a] Available as five-year estimates through the U.S. Census Bureau

[b] The Townsend Index score is calculated as a sum of the standardized scores for four census tract-level variables: (1) percentage unemployed, (2) percentage of households without access to a car, (3) percentage of all households renting, and (4) percentage of households with crowded housing. Higher Townsend Index scores reflect higher levels of deprivation and social disadvantage

Table 5

International classification of diseases, 9th revision, clinical modification (ICD-9-CM) codes used to categorize hospital discharge diagnoses

Diagnosis	ICD-9-CM codes
Cigarette use	649.01, V15.82
Alcohol use	305.00, 305.03
Marijuana use	305.20, 305.21, 305.23, 305.31
Other drugs of abuse	304.00, 305.50, 305.90, 304.71, 648.31, 304.01, 304.21, 305.60, 304.23, 305.61
Maternal mental health diagnosis	311.0, 300.4, 296.20, 296.50, 296.80, 296.90, 300.00, 300.11, 309.81, 313.81, 648.41, 648.42, V62.84, 307.51, 295.30, 295.90, 296.7, 295.32, 300.01
Previous poor birth outcome	V23.49, V23.5
HIV	V08
Sexually transmitted infection	647.11, 098.0, 098.15, 647.01, 054.9, 054.10, 054.19, 131.9, 614.9, 616.0, 131.01, 646.61, 647.21, 131.00, V08, 090.2, 090.9
Hypertension	401.9, 642.01, 642.31, 642.71, 642.91, 642.32, 642.21, 760
Obesity	278.00, 278.01, 649.11, V85.34, V85.37, V85.4, V85.36, V85.54, V85.32, V85.39
Hypothyroidism	244.1, 244.9, 648.11, 244.0, 244.8
Asthma	493.20, 493.90
Maternal transfusion	99.04, 99.07
Maternal cardiac complications	413, 414, 410.1, 410.2, 410.4, 421.0, 424.0, 426.0, 427.89, 648.61, 648.62, V12.53, V45.01, 426.7, 438.89, 412, 427.9
Maternal anemia	280.0, 280.9, 285.9, 648.21, 648.22, 282.41, 282.49
Maternal sickle cell disease	282.60, 282.63
Epilepsy	345.90, 649.41, 345.2
Disorders of placentation	656.71, 641.21, 641.01, 656.91
Cervical abnormalities	649.71, 654.51, 654.61
Uterine abnormalities	218.9, 615.0, 615.9, 617.9, 621.8, 654.01, 654.11, 654.41, 752.34, 654.12, 665.22, 752.2
Fetal malpresentation	652.21, 652.31, 652.61, 652.71, 652.81, 660.01, 652.91
Cephalopelvic disproportion or dystocia	653.01, 653.11, 653.41, 660.11, 660.21, 660.31, 660.41, 660.71, 660.81, 653.51
Cord abnormality	663.11, 663.21, 663.31, 663.41, 663.81, 663.01
Maternal hemorrhage or shock	285.1, 666.12, 666.22, 669.12, 669.21, 666.02, 669.22
Group B streptococcus carrier	V02.51
Preeclampsia or eclampsia	642.41, 642.42, 642.51, 642.61
Preterm Labor	644.21
Oligohydramnios	658.01, 761.2
Intrauterine growth retardation	656.51, 764.93, 764.07, 764.08, 764.09
Maternal diabetes	250.00, 648.01, 648.81, 790.21, 249.00, 250.01, V58.67, 775
Induction of labor	73.01, 73.1, 73.4
Chorioamnionitis	658.41, 670.12, 670.02, 670.82, 762.7
Meconium stained amniotic fluid	770.12, 779.84, 770.11
Fetal distress	656.31, 659.71
Vacuum or forceps instrumentation	72, 72.1, 72.21, 72.4, 72.71, 72.79, 73.3, 72.29, 72.6, 72.9
Premature rupture of membranes	658.11, 658.21
Birth injury	767.2, 767.3, 767.6, 767.8

Diagnosis	ICD-9-CM codes
Congenital spinal anomaly	756.1, 756.19
Congenital pulmonary anomaly	748.4, 277, 748.69, 748.3
Congenital gastrointestinal anomaly	751.3, 751.69, 756.73, 756.79
Cleft palate or lip	749.04
Congenital renal condition	593.89, 753.15, 753.29, 753.3
Congenital cardiac anomaly	416.8, 423.9, 745.4, 745.5, 747, 746.89, 427.9, 745.11, 747.49, 747.83, 745.1, 747.3
Metabolic, genetic, or chromosomal disorders	V29.3, 758, 759.81
Intraventricular hemorrhage	772.11, 772.14, 772.12
Necrotizing enterocolitis	777.5, 777.53
Neonatal infection	771.81, 041.04, 771.89, 322.9, 041.1, 041.19, 041.85, 482, V29.0, 041.01, 482.32, 770, 771.1, 771.82, 771.83, 008.45, 995.92
Retinopathy of prematurity	362.23, 362.2, 362.22