

**Report prepared by**

Kathleen Mullan Harris

Carolyn Halpern

John Hussey

Eric Whitsel

Robert Hummer

John Knapp

# Wave V Birth Records Database – Release 1



CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

## Acknowledgment

This research uses data from Add Health, funded by grant P01 HD31921 (Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. The Office of Vital Statistics from the Ohio Department of Health (ODH) provided some data used in this study. Use of these data does not imply ODH agrees or disagrees with any presentations, analyses, interpretations, or conclusions.

## Suggested Citation

Citations of this Add Health User Guide should use the following format:

Harris, K., Halpern, C., Hussey, J., Whitsel, E., Hummer, B., and Knapp, J.. 2022. Add Health Birth Records Database. The National Longitudinal Study of Adolescent to Adult Health. Carolina Population Center, University of North Carolina at Chapel Hill. <https://doi.org/10.17615/fptr-9x43>.

## Introduction

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of over 20,000 adolescents who were in grades 7-12 during the 1994-95 school year, and have been followed for five waves to date, most recently in 2016-18 (see Harris et al. 2019 for more information about the study).

The Add Health Birth Records Database describes the birth event of Add Health sample members (AHSMs), born between 1974 and 1983. This database was constructed using birth certificate information and birth data of AHSMs reported by AHSMs and their parents. The Birth Records Database represents a subset of the larger Add Health study sample and is composed of AHSMs consenting to Add Health's use of their birth record data and born in states with agreements to make birth record data available to Add Health.

## Data Structure and Form

Transformations of the data into categorical variable formats were first done on the aggregate sample of *all births* that occurred in a state over a specific time interval. A subset of these statewide births was then specifically matched to AHSMs.

There were some challenges with transforming the original source data into categorical variables. First, each state's method for collecting and reporting birth data not only varied from one another, but also may have varied within states across years. As a result of these reporting differences, constructed variables in this database may be either direct transformations of source data single variables or composite variables constructed using two or more unique distinct source variables. Second, each state differed in what they would allow for data release. These state-specific regulations drove the decision to constrain *all* available

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

data to the strictest reporting requirements of one specific state. These guidelines served as the basis for overall reporting requirements and ultimately informed the Add Health variable construction and reporting formats. In some instances, birth record data from one state may have been available, but if the most restrictive state guidelines limited its version of the data, those cases were assigned a reserve code (described in the Reserve Codes section of this guide). Also, if another state's data could not be transformed into the most restrictive state's reporting format, a reserve code was assigned. For example, although the most restrictive state reported a specific data measurement, another state did not report the same measurement, or their method of collecting and reporting a similar measurement was in a format untranslatable to the most restrictive state's coding requirement. This is further detailed in the descriptions for reserve codes 95 and 96 below.

Before release, an overall deductive disclosure mitigation review of the database contents was performed.

## Data Limitations

The most significant limitation of these data is the variation among the states in the years that source data were available. In the AHSM population, the range of all possible birth years was from 1974 to 1983, and all source data focused specifically on this range of years. However, each state varied in the number of overall years of data provided within this range of birth years. Specifically, some states provided data that started later than 1974. Because Add Health data security policy does not disclose information related to the specific birth years of individual respondents or their states of birth, the distributions of birth years affected by non-reported or unavailable data cannot be inferred from these data. It is important to note that the AHSM birth records were only available from a subset of states and thus only a subset of the Add Health respondents. If birth records are used as an outcome variable, we recommend a careful analysis of missing data and use of the grand sample weight at Wave V (GSW5) in analysis.

## Pregnancy Complications Variables RMMA002, RMMB002, & RMMC002

Due to the reporting constraints and formatting of source data, the pregnancy complications variables had to be reported as a combination of three possible variables, RMMA002, RMMB002, and RMMC002, representing up to three possible conditions related to a pregnancy complication. For a record with only one relevant pregnancy condition code, only RMMA002 is used. For a record with two of the three possible condition codes, RMMA002 and RMMB002 are used in the record. For a record containing all three possible condition codes, RMMA002, RMMB002, and RMMC002 are used.

In this release, condition code 1 (diabetes), contained a low cell count that required suppression for deductive disclosure security. To retain some informative utility for these cases, rather than outright suppressing them, the records reporting diabetes were recoded to code 3, "Other." In addition to a matching a discrete "Other" code in the source data, code 3, "Other," in these data represents all possible condition codes not otherwise represented in the other categories. So, in addition to diabetes (which was recoded to the "Other" category in this release for security purposes), "Other" can represent conditions such as: placental complications, renal disease, heart disease, or anemia.

## Data Dictionary

Variable naming schema is summarized in Table 1 below. The first character "R" is constant throughout all the dataset variables. Characters 2–3 are used to group thematically similar variables together. The four

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

groupings are shown in the table below. Character 4 is used to differentiate versions of the same variable. As described earlier in the Data Structure and Form section, different variable versions were used when the format of the source data differed between years and necessitated a change to the categorical variable transformation schema to properly capture and represent what was present in the source data. Characters 5-7 indicate a variable number.

Table 1. Variable Naming Schema		
Character Position	Character	Description
1	R	Respondent
2 - 3	IW	Infant Wellbeing
	ID	Infant Demographics
	PD	Parental demographics
	MM	Maternal Medical Risk Factors
4	A	Version 1
	B	Version 2
	C	Version 3
5 - 7	1	Number
5 - 7	1	Number

A general outline of all the variables appearing in this dataset along with their definitions are listed in Table 2 below. Detailed information including frequencies of individual variable codes are in the accompanying codebook.

Table 2. Variable Definitions		
#	Name	Description
<b>Infant Wellbeing–IW</b>		
1	RIWA001	Birth weight, categorical, in grams
2	RIWA002	Gestational age at birth based on last menstrual period, completed weeks
3	RIWA003	Gestational age at birth based on obstetric estimate of gestation at delivery, completed weeks
4	RIWA004	Method of delivery
<b>Infant Demographics–ID</b>		
5	RIDA001	Parity
6	RIDA002	Singleton birth
<b>Parental demographics–PD</b>		
7	RPDA001	Mother’s nativity

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

8	RPDA002	Mother's age at birth
9	RPDA003	Father's age at birth
10	RPDA004	Mother's level of education at time of birth
11	RPDA005	Father's level of education at time of birth
<b>Maternal Medical Risk Factors–MM</b>		
12	RMMA001	Previous Live Births Now Deceased
13	RMMA002	Pregnancy Complications, 1
14	RMMB002	Pregnancy Complications, 2
15	RMMC002	Pregnancy Complications, 3

## Reserve Codes

Codes for missing values always begin with “9” and end with either “5”, “7”, or “8.” The interior is padded with enough additional “9”s to make the length exceed by one the maximum value of the variable. The following table illustrates the convention. Detailed breakdowns of missing codes and their occurrence frequencies can be found in the accompanying codebook.

<b>Table 3. Reserve Codes</b>	
<b>Code</b>	<b>Definition</b>
95, 995, 9995	No Basis for Calculation
96, 996, 9996	Missing in Source Data
98, 998, 9998	Unknown

**Reserve code 95, “No Basis for Calculation,”** was used when there was some form of available relevant data present in the source datasets, *however*, the data were reported in a format that could not be accurately transformed into the accepted Add Health categories. Code 95 was also used in instances where a source dataset had otherwise valid data, but some piece of relevant data required was missing from the record (e.g., when determining race/ethnicity, some source data split this into two or more variables. A 95 would be used in instances where ‘race’ was reported but ‘ethnicity’ was unknown, or vice versa.)

**Reserve code 96, “Missing in the Source Data,”** was used to indicate instances in which data were missing from a source dataset. Note, “Missing data” in this instance can represent two possible scenarios:

- 1) Source data had a relevant variable that could be transformed into the final variable categories, *but* the individual record in the source data was missing.
- 2) Source data contained no variables relevant to Add Health variable in question. (e.g., “Method of Delivery,” RIWA004, if a particular state’s source data simply did not report delivery method, code 96 would be used for that collection of records).

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

**Reserve code 98, “Unknown,”** was used in instances where a valid variable existed in the source data, but an individual record contained some form of special missing or reserve code indicating ‘unknown’ (e.g., ‘9999’, ‘-’, etc.) for that record’s response.

## References

Harris, Kathleen Mullan; Halpern, Carolyn Tucker; Whitsel, Eric A.; Hussey, Jon M.; Killeya-Jones, Ley A.; Tabor, Joyce; & Dean, Sarah C. (2019). Cohort profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). *International Journal of Epidemiology*. <https://doi.org/10.1093/ije/dyz115>