2024

Add Health Wave VI Documentation

Add Health
The National Longitudinal Study of Adolescent to Adult Health

Add Health as a Resource for the Science of the Exposome

Rural-Urban Commuting Area (RUCA) Codes

Report prepared by

Andrea N. Goodwin

James D. Stewart

Robert A. Hummer

Eric A. Whitsel

UNC | CAROLINA POPULATION CENTER

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

https://doi.org/10.17615/9ygq-g064

This user guide is one in a set of user guides focusing on the built, environmental, and natural features of geopositioned/geocoded Add Health respondent locations over Waves I-VI. Collectively, they describe exposomic measures in the following three domains:

Built Domain	**Environmental Domain**	**Natural Domain**
Commuting Area	Ambient Air	Altitude
Land Use	Indoor Air	Meteorology
Roadway Proximity/Density	Noise	Green space
	Waterborne Lead	Blue space
	Nighttime Light Pollution	
	Solar Irradiation	

Under the Built Domain, this particular user guide summarizes the rationale for the latest construction and assignment of rural-urban commuting area (RUCA) codes. It also documents how the RUCA source data were acquired, as well as the protocol for quality controlling their assignment and classification across waves. Whenever possible, construction, assignment, and classification were harmonized to ensure temporal comparability, although important inter-wave differences exist and are grey-highlighted herein.

Acknowledgement

Data for Wave VI of Add Health was supported by two cooperative agreements from the National Institute on Aging (1U01AG071448, principal investigator Robert A. Hummer, and 1U01AG071450, principal investigators Robert A. Hummer and Allison E. Aiello) and a special supplement (U01-AG071450-02S1, principal investigators Robert A. Hummer, Allison E. Aiello, and Eric A. Whitsel) to the University of North Carolina at Chapel Hill. Co-funding for Wave VI was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institute on Minority Health and Health Disparities, the National Institute on Drug Abuse, the NIH Office of Behavioral and Social Science Research, and the NIH Office of Disease Prevention. Data from Waves I-V of Add Health are from the Add Health Program Project, grant P01 HD31921 (Kathleen Mullan Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Add Health was originally designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. Add Health is currently directed by Robert A. Hummer; it was previously directed by Kathleen Mullan Harris (2004-2021) and J. Richard Udry (1994-2004). Information on obtaining Add Health data is available on the project website (https://addhealth.cpc.unc.edu).

Citation for User Guide

Goodwin AN, Stewart JD, Hummer RA, Whitsel EA. Add Health as a Resource for the Science of the Exposome**:** *Rural-Urban Commuting Area (RUCA) Codes,* Wave VI Data Documentation. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available from: https://doi.org/10.17615/9ygq-g064

Table of Contents

1. Introduction

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of U.S. adolescents who were in grades 7-12 during the 1994-1995 school year. Using a complex, school-based cluster-sampling frame, researchers selected high school and feeder school pairs from 80 communities across the United States and drew a sex- and grade-stratified random sample of 20,745 adolescents for inclusion in the study. This sample has been followed from adolescence into early midlife across six waves of data collection to date, with the most recent wave of data collection (Wave VI) taking place between 2022 and 2025 when respondents were ages 39 to 49.

Over the years, Add Health has collected a wealth of information from respondents and their parents about demographic characteristics, familial structures, social relationships, health behaviors, cognition, physical and mental health status, medication usage, and health care access. Add Health also has collected anthropometric, cardiovascular, metabolic, renal, hepatic, inflammatory/immune, infectious, neurodegenerative, and multi-omic biomarkers from respondents. In addition, Add Health has merged multilevel contextual data about the economic, school, neighborhood, policy, and environmental contexts in which the respondents are embedded to the core survey and biological data at each wave. The Add Health dataset thereby provides researchers with rich opportunities to explore the causes and consequences of health status across multiple contextual domains as individuals age across the life course.

This user guide is one in a series documenting the latest contextual and environmental data assembled under the exposome supplement introduced in the preceding acknowledgment. Collectively, the supplemental data and documentation enable researchers to examine a broader array of built, environmental, and natural exposures linked to accurately geopositioned/geocoded Add Health respondent residences from Wave I through Wave VI. Because Wave VI data are not ready for geocoding or dissemination at present, this user guide and the associated data are focused on Wave I-V linkages. The Add Health Team will update this data set and user guide when Wave VI data are available for dissemination.

2. General Overview

The aircraft noise measures include standard metrics based on data generated with support from the Federal Aviation Administration (FAA grant 13-C-AJFE-BU, principal investigator Junenette L. Peters) under the *FAA Center of Excellence for Alternative Jet Fuels and the Environment* (ASCENT) as well as proxies generated by the University of North Carolina at Chapel Hill (UNC) for spatiotemporally characterizing aircraft noise exposures on a national scale. The rationale for and utility of acquiring the aircraft noise measures is described below.

2.1 Rationale

Since its inception, Add Health has continued amassing and disseminating contextual data files across multiple levels of geography, thus resulting in an increasingly comprehensive and diverse set of contextual measures in a nationally representative study spanning adolescence to mid adulthood. In general, these data have been provided to establish infrastructure for research addressing the role of diverse exposures across multiple levels and across the life course in the etiology and disparities of our most pressing health issues. The data collectively position Add Health as a central resource for scientists to more effectively operationalize and study the exposome and its consequences for population health across the life course, with particular attention to disparities across population subgroups.

2.2 Utility

The aircraft noise measures described herein expand the contextual data available to Add Health researchers, enhancing their capacity to examine the social, environmental, and biological dimensions of the exposome and how they contribute to U.S. population health and disparities. They may be valuable to researchers who study the population burden of and trends in aircraft noise exposures;[1] demographic and socioeconomic disparities among them;[2] their associations with cardiovascular and neurocognitive health and disease;[3,4] as well as the putative roles of hearing, stress, and sleep disturbance in the latter.[4-6]

3. Processing Details

The two types of aircraft noise measures in this file include those based on daytime, nighttime, and day-night noise contours over large geographic areas surrounding 90 major U.S. airports and proxies based on count of, distance to, and enplanements at approximately 900 additional airports.

3.1 Aircraft Noise Measures around 90 Major Airports

3.1.1 Equivalent Sound Level for a 15-Hour Day (LAEQD) **[RMENEXEQD001]**
Equivalent Sound Level for a 9-Hour Night (LAEQN) **[RMENEXEQN001]**
Day-Night Level (DNL) **[RMENEXDNL001]**

Airport-specific, aircraft noise contours for the years 1995-2015 (in 5-year intervals) were estimated using FAA's Aviation Environmental Design Tool based on aircraft performance, temporal changes in airport configurations, and atmospheric and geological conditions by U.S. Department of Transportation John A. Volpe National Transportation Systems Center / Wyle Laboratories. Specifically, airport decks (airport runway locations and utilization data) from Volpe and Wyle were used to estimate daytime, nighttime, and day-night level noise measures around 90 major airports. These modeled estimates were provided to UNC as contour lines ranging from 45 to 75 decibels (dB) in 1 dB increments.

These source data were used to assign LAEQD, LAEQN, and DNL (dB) to geopositioned/geocoded Add Health study respondent locations, but were originally obtained to assign them to non-Add Health participant locations, at which time the data were subjected to several pre-processing quality control steps. First, a count of study participants within each airport's 45 dB contours for the year 2000 was generated using participant coordinates on January 1 of that year. Then, potential changes in contours from one source data year to the next were evaluated by comparing the geographic areas of 45 dB contours for the years 2000, 2005, and 2010, and observing that contour areas decreased over time. Next, a subset of airports was identified for which year 2000 contours were available from both Volpe and Wyle using different sets of assumptions (e.g., grid cell size and atmospheric conditions). The two sets of contours were used to estimate measures at geocoded participant locations on January 1, 2000, and those measures were compared using Deming regression parameter estimates, interclass correlation coefficients, and scatter plots. The identified differences in those measures were attributable to underlying differences in Volpe and Wyle assumptions. Volpe addressed these differences by re-generating contours for all Wyle airports using Volpe assumptions and Wyle airport decks. Additionally, contour lines that crossed were identified because they can lead to errors in participant-specific estimation. Volpe addressed contours that crossed by re-generating them for affected airports and years. Then, the revised contours were integrated into the UNC file repository and a quality control document was generated showing side-by-side comparisons of contour maps by study year for all 90 major airports. Leveraging thoroughly vetted source data, LAEQD, LAEQN, and DNL were estimated at geocoded Add Health respondent addresses using a series of geoprocessing tasks in R followed by data interpolation and summarization steps in SAS.

3.1.2 Geoprocessing for LAEQD, LAEQN, and DNL

The next geoprocessing steps relied primarily on spatial R commands within R version 4.1.3 (03-10-2022 -- "One Push-Up") for 64-bit Windows. First, the Universal Transverse Mercator (UTM) zone was identified for each of the 90 major airports. Second, each of the 45-75 dB LAEQD, LAEQN, and DNL contours for the 90 major airports was reprojected from geographic coordinates (longitude and latitude in decimal degrees) to UTM coordinates (eastings and northings in meters) for each of the five study years (1995, 2000, 2005, 2010, and 2015) to enable more accurate estimation and distance measurement. Third, the UTM-projected 45-75 dB LAEQD, LAEQN, and DNL contours for each airport were merged into a collection of mutually exclusive rings for each of the five study years to facilitate spatial joins of individual LAEQD, LAEQN, and DNL to geocoded/geopositioned Add Health respondent addresses. Next, geocoded/geopositioned Add Health respondent address locations were reprojected from geographic to UTM coordinates and LAEQD, LAEQN, and DNL were estimated at those locations by UTM zone and year of source data (1995, 2000, 2005, 2010, and 2015). Lastly, individual outputs from UTM zones were merged into single files for each year of source data. For more detail on the geoprocessing tasks, see Appendix A.

3.1.3 Data Interpolation and Summarization for LAEQD, LAEQN, and DNL

SAS version 9.4 for 64-bit Windows and the year-specific files generated using spatial R were used to interpolate LAEQD, LAEQN, and DNL at all available geocoded/geopositioned respondent addresses and dates between 1994 and 2015. After interpolation of LAEQD, LAEQN, and DNL, data redundancy in the output files was eliminated by consolidating address date ranges for individual respondents when LAEQD, LAEQN, and DNL within those date ranges did not change. All 20,745 Add Health Wave I respondents were included in the final analysis files, even when geographic coordinates were unavailable in the master address file. See Table 1 for a description of the replacement codes used to identify missing LAEQD, LAEQN, or DNL.

Table 1. Replacement Codes for Missing Values of LAEQD, LAEQN, and DNL

Missing Value	Description
-9989	Less than 45 dB: Respondent coordinates available for time period, but located outside the 45 dB (geographically largest) contour lines for the 90 major airports.
-9990	Coordinates missing in Add Health master address file for time period.
-9992	Missing LAEQD, LAEQN, or DNL source data for time period (applies to all years after 2015).

3.1.4 Quality Control Checks for LAEQD, LAEQN, and DNL

Extensive quality control measures were performed after the processing steps detailed above. Data inputs and outputs were verified at each step of LAEQD, LAEQN, and DNL estimations. Post-processing automated checks were used to confirm the number of unique respondent IDs in each output file, the absence of gaps in continuous date ranges, minima and maxima, and replacement codes for missing values. An additional battery of post-processing checks were executed to verify the integrity of assignments from start to finish. For LAEQD, LAEQN, and DNL, the verification relied on an intermediate file that contained all data in the final analysis file plus key source data, thereby enabling seamless verification of inputs and outputs. Verifications involved vetting measures for full address histories of stratified random samples of respondents based on key characteristics as follows:

- 5 random respondents with measures set to replacement code -9989 (< 45 dB).
- 5 random respondents with measures set to replacement code -9990 (missing coordinates).
- 5 random respondents with measures set to replacement code -9992 (missing data).
- 5 random respondents each from 6 different UTM zones (30 respondents total).
- 5 random respondents with no UTM zone assignment (interchangeable with -9990 above).
- 5 random respondents with coordinates that fell within overlapping contours, thereby necessitating calculation of measures based on formulae that linearly rescale (logarithmic) decibels before adding them:

  LAEQD = 10 * (log10 (sum (10**(LAEQD_FIRST/10), 10**(LAEQD_SECOND/10))));
  LAEQN = 10 * (log10 (sum (10**(LAEQN_FIRST/10), 10**(LAEQN_SECOND/10))));
  DNL = 10 * (log10 (sum (10**(DNL_FIRST/10), 10**(DNL_SECOND/10))));

- 20 random respondents with non-missing values (values do not equal -9989, -9990, or -9992), which provided the opportunity to verify mixtures of time intervals with missing and non-missing measures.

Targeted checks were also performed to ensure that measures were correct for (1) a single respondent located in a water body, which required "snapping" (moving) the respondent location to the closest point on land and (2) respondents located within noise contours that straddled a UTM zone boundary and which therefore required attachment to respondents in two different UTM zones. All post-processing quality control checks were successful.

3.2 Aircraft Noise Proxies for Other Airports

3.2.1 Airport Count **[RMENEXCNT001]**
Mean Distance **[RMENEXDIS001]**
Mean Total Enplanements **[RMENEXENP001]**

The aircraft noise proxies for approximately 900 additional airports were based on a spatiotemporally broad and readily available array of airport layers and associated variables, primarily from the Data and Maps (D&M) collection compiled by the Environmental Systems Research Institute, Inc. (ESRI: www.esri.com), but also directly from the National Transportation Atlas Database (NTAD), which is published by the Bureau of Transportation Statistics (BTS) within the U.S. Department of Transportation (DOT). The proxies are the count of, mean distance to, and mean total enplanements at airports within approximately 22.2 miles of each geocoded respondent's address, a search radius equal to the largest 45 dB DNL contour in UTM coordinates among the 90 major airports introduced in Section 3.1. Airport counts and distances were calculated using a combination of ESRI D&M, NTAD, and military airports. Total (i.e., large certified, commuter, air taxi, foreign, and in-transit) enplanements per year were available only for ESRI D&M and NTAD. Although the data source years were not entirely concordant with the *LAEQD*, *LAEQN,* and *DNL* source files (1995 to 2015 in 5-year intervals), they included annual data for non-military airports during calendar years 1994, 2000, 2003, 2008, and 2015 which were chosen to approximate those of the 90 major airports used in the estimation of LAEQD, LAEQN, and DNL. Because the airports layer in the oldest ESRI D&M data set used (ESRI D&M 2000) excluded airports with fewer than 250 passenger enplanements per year, the same exclusion was applied to airport layers for all subsequent years. For more details, see Table 2.

Table 2. Non-Military Data Sources Used to Estimate Proxies for Aircraft Noise from Other Airports

Study Year (Target)	Airport Coordinates Source	Source Date for Airport Coordinates	Airport Count (n)*	Total Enplanements Source In Metadata	Closest Total Enplanements Calendar Year
1995	ESRI D&M 2000 acais_airports_2000.shp	publication date: 02-16-1999, 1998, 1997; **ground condition: 1994-1996**	846	ACAIS/FAA via National Atlas of the United States and the United States Geological Survey	CY 1994
2000	ESRI D&M 2004 airports_2004.shp	publication date: 10-2001; **ground condition: 2000, 1999**	901	ACAIS/FAA via National Atlas of the United States and the United States Geological Survey	CY 2000
2005	ESRI D&M 2012 airports_2012.shp	publication date: 10-2001, 1998; **ground condition: 2000, 1999, 2003**	880	ACAIS/FAA	CY 2003
2010	ESRI D&M 2014 airports_2014.shp	publication date: 07-01-2012 public-use extract of NTAD **2010**	940	ACAIS/FAA	CY 2008
2015	National Transportation Atlas Database (NTAD) 2017	effective date: 06-22-2017	19,765	ACAIS/FAA: https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/media/cy15_all_enplanements.xls	CY 2015

ACAIS: Air Carrier Activity Information System. **FAA:** Federal Aviation Administration. **NTAD:** National Transportation Atlas Database (Department of Transportation, Bureau of Transportation Statistics). ***Note:** Airport counts reflect the total number of non-major airports before filtering to exclude those with less than 250 total enplanements. Airports for ESRI D&M 2000 had already been filtered.

3.2.2 Geoprocessing Airport Count, Mean Distance, and Mean Total Enplanements

Airport count, mean distance, and mean total enplanements were generated for each Add Health respondent with geographic coordinates using a series of geoprocessing tasks in R followed by data interpolation and summarization steps in SAS. The geoprocessing tasks relied primarily on spatial R commands within R version 4.1.3 (03-10-2022 -- "One Push-Up"). The overall geoprocessing workflow in R was as follows:

1. Reprojected ESRI D&M and NTAD airport locations from geographic coordinates (longitude and latitude in decimal degrees) to UTM coordinates (eastings and northings in meters) for closest available source data calendar years (1994, 2000, 2003, 2008, and 2015) to facilitate more accurate distance measurements for proxy assignments.
2. Isolated the largest 45 dB DNL contour in UTM coordinates to identify the search radius to use for calculating airport count, mean distance, and mean total enplanements in relation to Add Health respondent locations (approximately 22.2 miles).
3. Added in military airports to supplement ESRI D&M and NTAD.
4. Removed overlap between military airports and those already in ESRI D&M and NTAD, and implemented filtering of airports based on total enplanements.
5. Extracted airport count, mean distance, and mean total enplanements by source data year and UTM zone for individual airports within approximately 22.2 miles of respondent locations.
6. Merged outputs from individual UTM zones into single files for each source data year.

For more detail on geoprocessing tasks, see **Appendix B**.

3.2.3 Data Interpolation and Summarization for Airport Count, Mean Distance, and Mean Total Enplanements

SAS version 9.4 was used to leverage year-specific proxy files generated using spatial R to interpolate proxies for all available respondent locations and dates between 1993 and 2015. After interpolation of each proxy, data redundancy was eliminated in the output files by consolidating address date ranges for proxy values that did not change over time for individual respondents. The differences in data consolidation by proxy resulted in three separate files for analysis, one each for airport count, mean distance, and mean total enplanements.

To provide a full accounting for all 20,745 Add Health study respondents at Wave I (baseline), final analysis files were generated that included records for respondent locations even when there were no geographic coordinates in the master address file, no airports within the specified search radius, or no values for total enplanements in the source data. See **Table 3** for a description of the replacement codes used to identify missing values.

Table 3. Replacement Codes for Missing Values of Mean Total Enplanements

Missing Value	Description
-9987	No airports within the specified search radius. This replacement code applies to mean distance when airport count = 0 and mean total enplanements = 0.
-9990	Coordinates missing in Add Health master address file for time period. This replacement code applies to all proxies.
-9992	Missing source data for time period. This replacement code applies to (1) all variables for years after 2015 when geographic coordinates were available and (2) total enplanements for all years between 1994 and 2015 when airport count > 0 and mean distance > 0, but total enplanements = 0.

3.2.4 Quality Control Checks for Airport Count, Mean Distance, and Mean Total Enplanements

Data inputs and outputs were verified at each step of proxy estimation. Post-processing automated checks confirmed the number of unique respondent IDs in each output file, the absence of gaps in continuous ranges of days since randomization, proxy minima and maxima, and replacement codes for missing values. An additional battery of post-processing spot checks was executed to verify the integrity of proxy assignments from start to finish. The verification relied on an intermediate file for each proxy that contained all data in the final analysis file plus key source data, thereby enabling seamless comparisons of inputs and outputs. Verifications involved vetting proxies for full address histories of stratified random samples of respondents based on key characteristics as follows:

- 5 random respondents with proxies set to replacement code -9990 (missing coordinates).
- 5 random respondents with proxies set to replacement code -9992 (missing data).
- 5 random respondents with proxies set to replacement code -9987 (closest airport farther than 22.2-mile search radius).
- 5 random respondents each from 6 different UTM zones (30 respondents total).
- 5 random respondents with no UTM zone assignment (interchangeable with -9990 above).
- 20 random respondents with non-missing values (values do not equal -9987, -9990, or -9992), which provided the opportunity to verify mixtures of time intervals with missing and non-missing proxies.

A check also was performed to ensure that proxies were correct for a single respondent located in a water body, which required “snapping” (moving) the respondent location to the closest point on land. All post-processing quality control checks were successful.

4. Usage Note

Access to and use of the FAA noise data described herein are subject to a negotiated *Memorandum of Understanding* between the FAA and UNC that requires:

(1) Use of the FAA noise data only by Add Health restricted-use contract holders working in the UNC Secure Research Workspace under an approved Add Health restricted-use contract
(2) Submission to FAA of a short description of the intended use of the FAA noise data
(3) Review and approval by FAA of the intended use of the data and restricted-use contract holder before granting access to the FAA noise data
(4) Submission to FAA of any request to access and report any airport-specific / identifiable noise data
(5) Submission of all abstracts, posters, manuscripts, other publications, and presentations involving FAA noise data to FAA for review and approval (at least six weeks before submission for publication)
(6) Inclusion of the following acknowledgement in all publications using FAA noise data: *“This research was funded by the U.S. Federal Aviation Administration Office of Environment and Energy through ASCENT, the FAA Center of Excellence for Alternative Jet Fuels and the Environment, project 003 through FAA Award Number 13-C-AJFE-BU under the supervision of Adam Scholten. Any opinions, findings, conclusions or recommendations expressed in this this material are those of the authors and do not necessarily reflect the views of the FAA.”*

The submissions described in (2 and 4-5) must accompany all applications for access to these restricted use data through the Add Health Data Portal, https://data.cpc.unc.edu/projects/2/view. Related questions should be directed to Add Health Contracts, addhealth_contracts@unc.edu.

5. Data Files

5.1 Structure

The Aircraft Noise Measures are provided in six files encompassing metrics based on daytime (LAEQD), nighttime (LAEQN), and day-night (DNL) noise contours over large geographic areas surrounding 90 major U.S. airports and proxies based on count of, distance to, and enplanements at approximately 900 additional airports. The files are provided in a multiple-records-per-respondent long format, each comprised of four variables. The files contain one record for each of the 20,745 Add Health Wave I sample members (as identified by a masked respondent identifier, AID) at every time period during their follow-up (as identified by the dates from and date to, RMENEX###DFR and RMENEX###DTO), and the corresponding aircraft noise measures and proxies (RMENEX###001), where ###= EQD, EQN, DNL, CNT, DIS, or ENP. Please consult the accompanying codebook for additional details.

5.2 Contents

The Aircraft Noise Measures data files include the variables below, which are described in the corresponding codebook documentation that also contains frequencies.

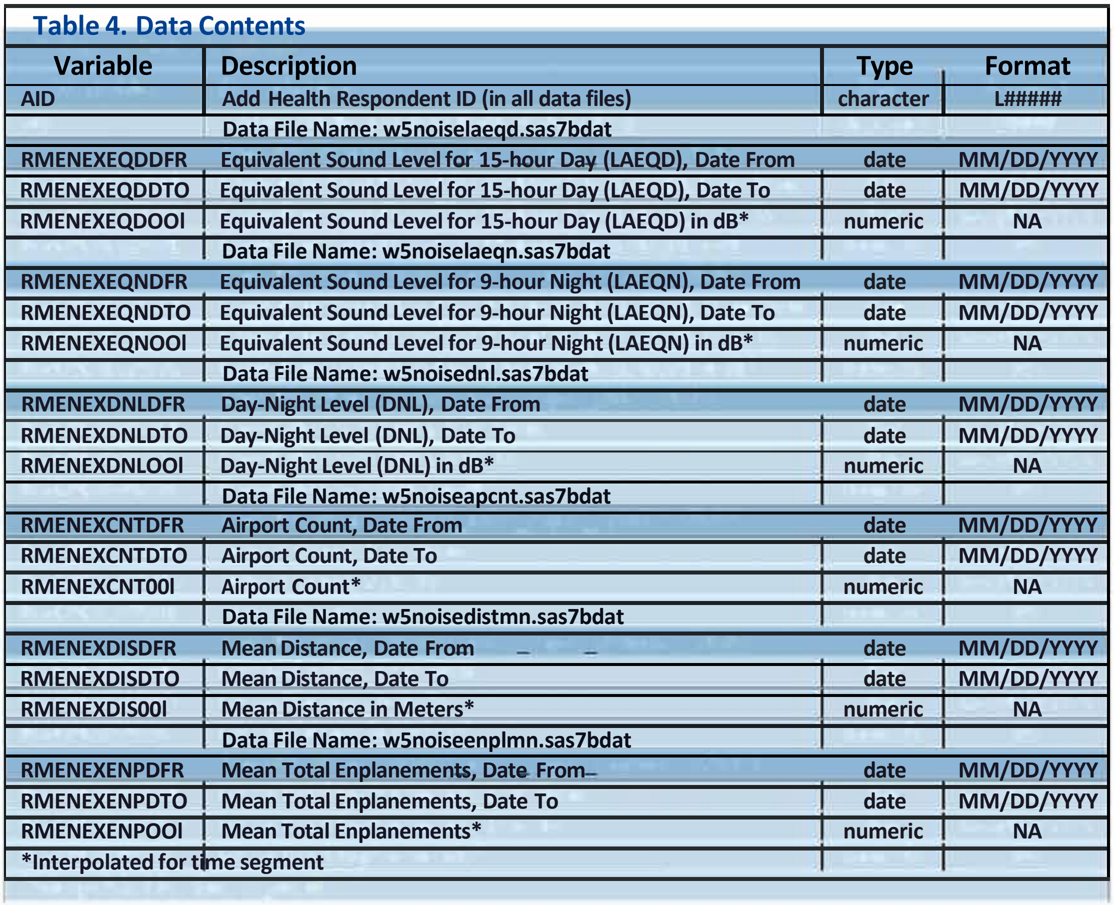

Table 4. Data Contents

Variable	Description	Type	Format
AID	Add Health Respondent ID (in all data files)	character	L#####
	Data File Name: w5noiselaeqd.sas7bdat		
RMENEXEQDDFR	Equivalent Sound Level for 15-hour Day (LAEQD), Date From	date	MM/DD/YYYY
RMENEXEQDDTO	Equivalent Sound Level for 15-hour Day (LAEQD), Date To	date	MM/DD/YYYY
RMENEXEQD001	Equivalent Sound Level for 15-hour Day (LAEQD) in dB*	numeric	NA
	Data File Name: w5noiselaeqn.sas7bdat		
RMENEXEQNDFR	Equivalent Sound Level for 9-hour Night (LAEQN), Date From	date	MM/DD/YYYY
RMENEXEQNDTO	Equivalent Sound Level for 9-hour Night (LAEQN), Date To	date	MM/DD/YYYY
RMENEXEQN001	Equivalent Sound Level for 9-hour Night (LAEQN) in dB*	numeric	NA
	Data File Name: w5noisednl.sas7bdat		
RMENEXDNLDFR	Day-Night Level (DNL), Date From	date	MM/DD/YYYY
RMENEXDNLDTO	Day-Night Level (DNL), Date To	date	MM/DD/YYYY
RMENEXDNL001	Day-Night Level (DNL) in dB*	numeric	NA
	Data File Name: w5noiseapcnt.sas7bdat		
RMENEXCNTDFR	Airport Count, Date From	date	MM/DD/YYYY
RMENEXCNTDTO	Airport Count, Date To	date	MM/DD/YYYY
RMENEXCNT001	Airport Count*	numeric	NA
	Data File Name: w5noisedistmn.sas7bdat		
RMENEXDISDFR	Mean Distance, Date From	date	MM/DD/YYYY
RMENEXDISDTO	Mean Distance, Date To	date	MM/DD/YYYY
RMENEXDIS001	Mean Distance in Meters*	numeric	NA
	Data File Name: w5noiseenplmn.sas7bdat		
RMENEXENPDFR	Mean Total Enplanements, Date From	date	MM/DD/YYYY
RMENEXENPDTO	Mean Total Enplanements, Date To	date	MM/DD/YYYY
RMENEXENP001	Mean Total Enplanements*	numeric	NA

*Interpolated for time segment

6. References

1. Nguyen DD, Levy JI, Kim C, Lane KJ, Simon MC, Hart JE, Whitsel EA, VoPham T, Malwitz A, Peters JL. Characterizing temporal trends in populations exposed to aircraft noise around U.S. airports: 1995-2015. J Expo Sci Environ Epidemiol 2023.

2. Simon MC, Hart JE, Levy JI, VoPham T, Malwitz A, Nguyen D, Bozigar M, Cupples LA, James P, Laden F, Peters JL. Sociodemographic Patterns of Exposure to Civil Aircraft Noise in the United States. Environ Health Perspect 2022;130(2):27009.

3. Peters JL, Zevitas CD, Redline S, Hastings A, Sizov N, Hart JE, Levy JI, Roof CJ, Wellenius GA. Aviation Noise and Cardiovascular Health in the United States: a Review of the Evidence and Recommendations for Research Direction. Curr Epidemiol Rep 2018;5(2):140-152.

4. Wright BA, Peters ER, Ettinger U, Kuipers E, Kumari V. Moderators of noise-induced cognitive change in healthy adults. Noise Health. 2016 May-Jun;18(82):117-32. doi: 10.4103/1463-1741.181995.PMID: 27157685

5. Bozigar M, Huang T, Redline S, Hart JE, Grady ST, Nguyen DD, James P, Nicholas B, Levy JI, Laden F, Peters JL. Associations between Aircraft Noise Exposure and Self-Reported Sleep Duration and Quality in the United States-Based Prospective Nurses' Health Study Cohort. Environ Health Perspect 2023;131(4):47010.

6. Wright B, Peters E, Ettinger U, Kuipers E, Kumari V. Understanding noise stress-induced cognitive impairment in healthy adults and its implications for schizophrenia. Noise Health. 2014 May-Jun;16(70):166-76. doi: 10.4103/1463-1741.134917.PMID: 24953882

Appendix A
Geoprocessing of LAEQD, LAEQN, and DNL Contours

1. Identified the Universal Transverse Mercator (UTM) zone for each of the 90 major airports.

   This was accomplished by obtaining the centroid (geographic center point) of each airport's year 2000 45 dB LAEQD, LAEQN, and DNL contour and conducting a point-in-polygon spatial join to UTM zone polygons from ESRI D&M.

2. Reprojected each of the 45-75 dB LAEQD, LAEQN, and DNL contours for the 90 major airports from geographic coordinates (longitude and latitude in decimal degrees) to UTM coordinates (eastings and northings in meters) for each of the five study years (1995, 2000, 2005, 2010, and 2015) to enable more accurate estimation and distance measurement.

3. Merged the UTM-projected 45-75 dB LAEQD, LAEQN, and DNL contours for each airport into a collection of mutually exclusive rings for each of the five study years to facilitate spatial joins of individual LAEQD, LAEQN, and DNL levels to geocoded Add Health respondent addresses.

4. Reprojected geocoded Add Health respondent addresses from geographic to UTM coordinates and estimated LAEQD, LAEQN, and DNL for Add Health respondents by UTM zone and year of source data (1995, 2000, 2005, 2010, and 2015).

5. Merged individual outputs from UTM zones into single files for each measure (LAEQD, LAEQN, and DNL) and year of source data.

   This step created five year-specific files for each measure containing the following fields:

AID:	Add Health Respondent Location ID
WAVE:	Add Health Wave
ZONE:	UTM zone
APCODE:	Airport Identification Code
DB:	Noise Measure in Decibels

Notes:

- SAS version 9.4 was used to leverage the year-specific measures generated using spatial R to interpolate measures for all available respondent locations and dates between 1993 and 2015.
- These year-specific, wave-referenced files were converted to date-based files by joining them to the master Add Health coordinates file for respondent full address histories from 1994 to 2019.

Appendix B
Geoprocessing of Airport Count, Mean Distance, and Mean Total Enplanements

1. Reprojected ESRI D&M and NTAD airport locations from geographic coordinates (longitude and latitude in decimal degrees) to UTM coordinates (eastings and northings in meters) for closest available source data calendar years (1994, 2000, 2003, 2008, and 2015) to facilitate more accurate distance measurements for proxy assignments.

2. Isolated the largest 45 dB DNL contour in UTM coordinates to identify the search radius to use for calculating airport count, mean distance, and mean total enplanements in relation to Add Health respondent locations (approximately 22.2 miles).

   To be conservative in identifying the search radius for other airports, the bounding box for each study airport's merged DNL contours for each study year (see step 3 of geoprocessing tasks for DNL data set) was calculated and the maximum dimension from those measures was extracted. The maximum dimension returned from the 450 bounding boxes (90 study airports times five study years) was a difference in X coordinates (longitude) of 71,446 meters. Dividing this in half produced a search radius of 35,723 meters (about 22.2 miles).

   Note: The bounding box for a geographic entity such as an airport contour is a rectangle of the smallest size possible that still encompasses the full entity. It is defined by two coordinate pairs, one identifying the minimum X and Y coordinates (lower-left corner of the rectangle) and one identifying the maximum X and Y coordinates (upper-right corner of the rectangle).

3. Added in military airports to supplement ESRI D&M and NTAD.

   A review of ESRI D&M showed that coverage of military airports was not complete or consistent from study year to study year. One reason for this is that military airports are considered restricted facilities with no obligation to report enplanements. To fill in gaps in the geographic and temporal coverage of military airports, an August 2016 snapshot of information on military airstrips available from the Digital Aeronautical Flight Information File (DAFIF), a product of the National Geospatial-Intelligence Agency (NGA, www.nga.mil), was used. Although the DAFIF data snapshot represented a single point in time, it provided a reliable source of information on military airfields that had been in existence for a long period of time. The DAFIF was withdrawn from public distribution on 1 Oct 2005 (see Federal Register Volume 69, Number 222, pages 67546-67547), but made available to the aircraft noise study through an agreement with the FAA.

   An important fact to note about the DAFIF data on military airfields is that they did not provide information on total enplanements. For that reason, DAFIF military airfields were used to calculate counts and mean distances with respect to respondent locations, but not mean total enplanements. In the case of military airfields from DAFIF, total enplanements was set to missing rather than zero to avoid contributing a spurious value to the calculation of mean total enplanements.

   One gap in the DAFIF data provided by the FAA was a lack of airfields in Alaska. To fill in that gap, Alaska military airfields were extracted from a Department of Defense (DoD) point-based shapefile providing the names and locations of “Military Installations, Ranges, and Training Areas” (MIRTA). The metadata for the MIRTA data, which was updated 18 Jan 2017, are as follows:

> This dataset, released by DoD, contains geographic information for major installations, ranges, and training areas in the United States and its territories. This release integrates site information about DoD installations, training ranges, and land assets in a format which can be immediately put to work in commercial geospatial information systems. Homeland Security/Homeland Defense, law enforcement, and readiness planners will benefit from immediate access to DoD site location data during emergencies. Land use planning and renewable energy planning will also benefit from use of this data. Users are advised that the point and boundary location datasets are intended for planning purposes only, and do not represent the legal or surveyed land parcel boundaries.
> Source: https://catalog.data.gov/dataset/military-installations-ranges-and-training-areas, accessed 16 May 2018.

After downloading the MIRTA shapefile, Google Earth was used to view imagery and confirm locations. International Civil Aviation Organization (ICAO) codes in the MIRTA shapefile were then used in combination with airport searches at www.airnav.com/airports to extract geographic coordinates, because the coordinates available on airnav.com, a site for civilian pilots, were found to be much more accurate than those from the MIRTA file in terms of pinpointing runways. The MIRTA airfield for the Oliktok Long-Range Radar Site was deleted, because it was closed in 1995.

As a supplement to the military airports from the DAFIF data set, total enplanements for the MIRTA airfields were set to missing for calculation of mean total enplanements.

4. Removed overlap between military airports and those already in ESRI D&M and NTAD, and implemented filtering of airports based on total enplanements.

   There was some overlap between the military airports available from the DAFIF and MIRTA sources developed by the DoD with those available from ESRI D&M and NTAD, so that overlap was removed after restricting the ESRI D&M and NTAD airports to those with 250 or more total enplanements. In the event of an overlap, the ESRI D&M or NTAD version of the military airport was retained instead of that from the supplemental sources. This filled in some of the missing values for total enplanements, as there were some military airports in the ESRI D&M and NTAD files for which total enplanements had been reported.

5. Extracted counts of, distances to, and total enplanements by source data year and UTM zone for individual airports within approximately 22.2 miles of respondent locations.

   Identifying airports within the specified search radius around respondent locations by UTM zone allowed the generation of three outputs, two primary and one derived. The primary outputs were (1) a file containing unique airport IDs and total enplanement values for each UTM zone and calendar year of interest and (2) a matrix containing the distances in meters between all respondents and airports for a UTM zone and calendar year of interest. The second primary outputs, the distance matrices, were merged with respondent address date ranges to derive a file containing date-specific distances to individual airports by UTM zone and calendar year of interest with respect to total enplanements.

6. Merged outputs from individual UTM zones into single files for each source data year.

   This step created five year-specific files containing the following fields:

AID:	Add Health Respondent Location ID
WAVE:	Add Health Wave
LON_WGS84:	Longitude in Decimal Degrees, WGS84 Datum
LAT_WGS84:	Latitude in Decimal Degrees, WGS84 Datum
ZONE:	UTM zone
APCODE:	Airport Identification Code
DISTM:	Distance in meters
TOT_ENP:	Total enplanements

   If there were no airports within a UTM zone for a particular respondent, the output file contained an NA value for each APCODE within the respondent's search radius. The TOT_ENP field was also set to NA, whereas the DISTM value was set to zero. These values for missing airports allowed them to be distinguished from military airports from DoD sources, which had no total enplanements values. For these military airports, the output file contained the APCODE for those within each respondent's search radius as well as a distance value in meters, but the total enplanements value was set to zero.

Note: SAS version 9.4 was used to leverage the year-specific proxies generated using spatial R to interpolate proxies for all available respondent locations and dates between 1993 and 2015.