

National Longitudinal Study of Adolescent to Adult Health

Wave IV Ambient Air Pollutants: Individual Pollutant, Daily Particulate Matter, and Toxic Gas Estimates Codebook



July 2018

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded 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. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>).

Ambient Air Pollutant Estimates Codebook

INTRODUCTION

Add Health's Ambient Air Pollutants database consists of three separate individual by temporal-level files:

- aapip4: Individual Pollutant estimates
- aappm4: Daily Particulate Matter (PM) estimates
- aaptx4: Daily Toxic Gas (TOX) estimates

Modeled ambient concentrations are to be used as surrogates for personal exposure to those ambient air pollutants, and the modeled concentrations have been assigned to each Add Health participant for this purpose. Generally, for each study participant, personal daily exposures to ambient air pollutants were estimated from the day in which biometric data were obtained and over the preceding year, leading to 366 daily exposure estimates for each Add Health study participant in Wave IV. Air pollutants included in the exposure metrics are:

Particulate Matter (PM)	Gases
PM _{2.5} , PM ₁₀ , SO ₄ , NO ₃ , NH ₄ , OC (organic carbon), EC (elemental carbon), Fe, Al, Si, Ti, Ca, Mg, K, Mn, Na, Cl	O ₃ , NO ₂ , HNO ₃ , HONO, H ₂ O ₂ , CO, SO ₂ <u>Year 2007 only gases:</u> acrolein, acetaldehyde, benzene, butadiene, ethanol, formaldehyde, naphthalene

These air pollutants were modeled in daily Community Multiscale Air Quality (CMAQ) model runs, which were performed by the U.S. EPA over a 12 km x 12 km grid covering the contiguous U.S. for the years 2006-2008. CMAQ is a 3-D photochemical model that simulates ground level concentrations of several pollutants such as O₃, PM species (including NH₄⁺, Na⁺, Cl⁻, EC, OC, and selected trace metals), and several air toxics. A recent version of the CMAQ model (v5.0.1)¹, was used to calculate the hourly distribution of air pollutants over the contiguous U.S. Comparisons between CMAQ output and monitoring data from the U.S. EPA's Air Quality System (AQS) network for PM_{2.5} and O₃ have found the 12 km x 12 km grid to provide a reasonable estimate of concentration².

For seven prevalent PM species (EC, OC, NH₄, NO₂, NO₃, O₃, SO₄), a state-of-the-science approach was applied to estimate personal daily exposures to ambient air pollutants, based on both modeled and measured ambient air quality data. A downscaling data fusion approach has been applied to blend the U.S. EPA's AQS monitoring data and the CMAQ modeled air quality data. The U.S. EPA's AQS data were obtained from more than 2000 air pollutant monitors nationwide, and the monitoring network has been strictly quality assured and quality controlled. The data were directly downloaded from the U.S. EPA's website. Quality assured simulation output meeting those criteria is available from the U.S. EPA. These measures were only available for year 2007 and, therefore, these data contained fewer observations than either of the other two files. Consequently, not all Add Health participants inhabiting areas measured will have an estimate at all 366 time points, depending on when the Wave IV in-home interview was completed.

The downscaling approach took advantage of both the measured (more accurate) and the modeled (more spatial and temporal coverage) air quality data and lead to more spatially resolved estimates of species values at locations where air quality data are unavailable to measure exposure³. Specifically, the downscaling algorithm regressed AQS data on the CMAQ model estimates with spatially-varying regression coefficients. In locations where AQS data were not available, the CMAQ model estimates were then used to predict the actual concentrations, based on the downscaling regression coefficients. This approach can also correct the bias in CMAQ estimates⁴⁻⁹. Downscaled exposures were assigned at the centroid of each zip-code tabulation area (ZCTA) for the date of CVD risk factor measurement and the Wave IV interview date plus the preceding 365 days, and the spatial and temporal identifiers were subsequently removed to protect participant confidentiality.

VARIABLE NAMING CONVENTIONS

With the exception of AID, all variables in this contextual database adhere to the following nomenclature:

1st and 2nd characters – Refer to the Ambient Air Pollutants database (AP).

3rd and 4th characters – Identify the type of data comprising the files. "PM," "TX," and "IP" denote Daily PM, Daily TOX, Individual Pollutants data, respectively.

5th and 6th characters – Report that these pollution variables are attached to respondent locations at Wave IV of Add Health, 2008 (08).

7th and 8th characters – The final two unique digits serve to distinguish the variables comprising these data files. The aapip4 file differs slightly from the aappm4 and aaptx4 files. Each individual pollutant has three measures the CMAQ value, mean, and standard deviation, denoted with 1, 2, and 3, respectively.

LEVEL OF ANALYSIS

- aapip4: Zip code tabulation area centroid (nearest to respondent)
- aappm4: Grid cell (12km by 12km) centroid (nearest to respondent)
- aaptx4: Grid cell (12km by 12km) centroid (nearest to respondent)

AIR POLLUTANTS

PMIJ	Sum of PM in the Aitken (i) and accumulation (j) Modes (i.e. PM _{ij}) (PM _{2.5})
ANH4IJ	Nitrate
O3	Ozone
H2O2	Hydrogen Peroxide
AALJ	Aluminum
ACAK	Coarse Mode Ca ²⁺ (Calcium Ion)
AMNJ	Manganese
PM10	Coarse, Ground-Level Particulates
AOCIJ	Fine Organic Carbon
NO2	Nitrogen Dioxide
CO	Carbon Monoxide
ASIJ	Fine Particulate Silicon
AMGJ	Fine Particulate Magnesium
ANAIJ	Sodium
ASO4IJ	Sulfate
AECIJ	Elemental Carbon
HNO3	Nitric Acid
SO2	Sulfur Dioxide
ATIJ	Fine Mode Titanium
AKJ	Fine Mode Potassium
ANAK	Coarse Mode Sodium Ion (Na ⁺)
ANO3IJ	Nitrate (ANO3I[1]+ANO3J[1])
TNO3	Total Nitrate (Tno3 = No3+Hno3)
HONO	Nitrous Acid
AFEJ	Fine Mode Iron
ACAJ	Fine Mode Calcium
AKK	Coarse Mode Potassium Ion (K ⁺)
ACLIJ	Fine Particulate Chloride

REFERENCES

- (1) D. Byun, K. L. Schere, "Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system," *Applied Mechanics Reviews*, vol. 59, no. 1-6, pp. 51-77, 2006.
- (2) J. Warren, M. Fuentes, A. Herring, et al., "Spatial-Temporal Modeling of the Association between Air Pollution Exposure and Preterm Birth: Identifying Critical Windows of Exposure," *Biometrics*, vol. 68, no. 4, pp. 1157-1167, 2012.
- (3) M. Fuentes, A. E. Raftery, "Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models," *Biometrics*, vol. 61, no. 1, pp. 36-45, 2005.

- (4) G. Chen, J. Y. Li, Q. Ying, et al., "Evaluation of observation-fused regional air quality model results for population air pollution exposure estimation," *Science of the Total Environment*, vol. 485, pp. 563-574, 2014.
- (5) Y. Choi, "The impact of satellite-adjusted NO_x emissions on simulated NO_x and O₃ discrepancies in the urban and outflow areas of the Pacific and Lower Middle US," *Atmospheric Chemistry and Physics*, vol. 14, no. 2, pp. 675-690, 2014.
- (6) J. L. Crooks, H. Ozkaynak, "Simultaneous statistical bias correction of multiple PM_{2.5} species from a regional photochemical grid model," *Atmospheric Environment*, vol. 95, pp. 126-141, 2014.
- (7) K. Foley, P. Dolwick, C. Hogrefe, et al., "Dynamic Evaluation of CMAQ Part II: Evaluation of relative response factor metrics for ozone attainment demonstrations," *Atmospheric Environment*, vol. 103, pp. 188-195, 2015.
- (8) S. L. Napelenok, H. Simon, P. V. Bhawe, et al., "Diagnostic Air Quality Model Evaluation of Source-Specific Primary and Secondary Fine Particulate Carbon," *Environmental Science & Technology*, vol. 48, no. 1, pp. 464-473, 2014.
- (9) B. J. Reich, H. H. Chang, K. M. Foley, "A Spectral Method for Spatial Downscaling," *Biometrics*, vol. 70, no. 4, pp. 932-942, 2014.

Wave IV: Ambient Air Pollutants, Individual Pollutant Estimates Codebook

Number of observations: 5,685,078

Respondent ID	AID	char 8
5,685,078		range 10000000 to 99999999

Number of Days Before Wave 4 Interview	APIP081	num 3
5,685,078		range 0 to 365

EC: CMAQ	APIP0821	num 10
5,685,078		range 0.0012 to 31.1338

EC: Post Mean	APIP0822	num 11
5,685,078		range 0.0043 to 11.0349

EC: Post Standard Deviation	APIP0823	num 11
5,685,078		range 0.0111 to 9.1783

NH4: CMAQ	APIP0831	num 11
5,685,078		range 0.0000 to 16.2201

NH4: Post Mean	APIP0832	num 12
5,685,078		range -0.0465 to 12.8067

NH4: Post Standard Deviation	APIP0833	num 11
5,685,078		range 0.0160 to 6.2901

NO2: CMAQ	APIP0841	num 9
5,685,078		range 0.0234 to 117.9190

NO2: Post Mean	APIP0842	num 11
5,685,078		range 0.5310 to 68.1553

NO2: Post Standard Deviation	APIP0843	num 11
5,685,078		range 0.7395 to 13.1921

NO3: CMAQ	APIP0851	num 11
5,685,078		range 0.0000 to 27.1402

NO3: Post Mean	APIP0852	num 12
----------------	-----------------	--------

5,685,078		range -0.0040 to 127.0370	
NO3: Post Standard Deviation		APIP0853	num 11
5,685,078		range 0.0002 to 45.6736	
O3: CMAQ		APIP0861	num 8
5,685,078		range 0.1056 to 109.9520	
O3: Post Mean		APIP0862	num 11
5,685,078		range 1.5242 to 70.5622	
O3: Post Standard Deviation		APIP0863	num 11
5,685,078		range 1.6418 to 12.5246	
OC: CMAQ		APIP0871	num 10
5,685,078		range 0.0049 to 161.3150	
OC: Post Mean		APIP0872	num 11
5,685,078		range 0.0674 to 41.8605	
OC: Post Standard Deviation		APIP0873	num 11
5,685,078		range 0.0668 to 45.5272	
SO4: CMAQ		APIP0881	num 10
5,685,078		range 0.0085 to 40.7621	
SO4: Post Mean		APIP0882	num 11
5,685,078		range 0.0284 to 27.3101	
SO4: Post Standard Deviation		APIP0883	num 11
5,685,078		range 0.0341 to 32.5210	

Wave IV: Ambient Air Pollutants, Daily Particulate Matter Estimates Codebook

Number of observations: 5,685,078

Respondent ID	AID	char 8
5,685,078		range 10000000 to 99999999

Number of Days Before Wave 4 Interview	APPM0801	num 3
5,685,078		range 0 to 365

Daily PM: PMIJ	APPM0802	num 9
5,685,078		range 0.0435 to 420.6620

Daily PM: ANH4IJ	APPM0803	num 11
5,685,078		range 0.0000 to 16.2201

Daily PM: O3	APPM0804	num 8
5,685,078		range 0.1056 to 119.4610

Daily PM: H2O2	APPM0805	num 11
5,685,078		range 0.0000 to 14.3457

Daily PM: AALJ	APPM0806	num 11
5,685,078		range 0.0000 to 6.5680

Daily PM: ACAK	APPM0807	num 11
5,685,078		range 0.0000 to 29.1566

Daily PM: AMNJ	APPM0808	num 11
5,685,078		range 0.0000 to 0.1326

Daily PM: PM10	APPM0809	num 9
5,685,078		range 0.0465 to 522.5900

Daily PM: AOCIJ	APPM0810	num 10
5,685,078		range 0.0053 to 193.3600

Daily PM: NO2	APPM0811	num 9
5,685,078		range 0.0426 to 117.9190

Daily PM: CO	APPM0812	num 7
--------------	-----------------	-------

5,685,078		range 39.5657 to 4556.40
-----------	--	--------------------------

Daily PM: ASIJ		APPM0813	num 11
5,685,078		range 0.0000 to 22.4010	

Daily PM: AMGJ		APPM0814	num 11
5,685,078		range 0.0000 to 0.5774	

Daily PM: ANAIJ		APPM0815	num 11
5,685,078		range 0.0000 to 4.5902	

Daily PM: ASO4IJ		APPM0816	num 9
5,685,078		range 0.0124 to 40.7621	

Daily PM: AECIJ		APPM0817	num 10
5,685,078		range 0.0030 to 38.1845	

Daily PM: HNO3		APPM0818	num 11
5,685,078		range 0.0000 to 13.1103	

Daily PM: SO2		APPM0819	num 11
5,685,078		range 0.0000 to 167.5920	

Daily PM: ATIJ		APPM0820	num 11
5,685,078		range 0.0000 to 2.1362	

Daily PM: AKJ		APPM0821	num 11
5,685,078		range 0.0002 to 11.6559	

Daily PM: ANAK		APPM0822	num 11
5,685,078		range 0.0000 to 32.5331	

Daily PM: ANO3IJ		APPM0823	num 11
5,685,078		range 0.0000 to 27.1402	

Daily PM: TNO3		APPM0824	num 10
5,685,078		range 0.0027 to 39.3705	

Daily PM: HONO		APPM0825	num 11
5,685,078		range 0.0002 to 5.4373	

Daily PM: AFEJ		APPM0826	num 11
5,685,078		range 0.0000 to 3.8709	

Daily PM: ACAJ		APPM0827	num 11
5,685,078		range 0.0000 to 10.9200	

Daily PM: AKK		APPM0828	num 11
5,685,078		range 0.0000 to 8.4268	

Daily PM: ACLIJ		APPM0829	num 11
5,685,078		range 0.0000 to 12.4682	

Wave IV: Ambient Air Pollutants, Toxic Gas Estimates Codebook

Number of observations: 3,386,355

Respondent ID	AID	char 8
3,386,355		range 10000000 to 99999999

Number of Days Before Wave 5 Interview	APT0801	num 3
3,386,355		range 0 to 365

Daily TOX: ACROLEIN	APT0802	num 11
3,386,355		range 0.0000 to 2.3342

Daily TOX: ALD2	APT0803	num 9
3,386,355		range 0.0414 to 22.8027

Daily TOX: BENZENE	APT0804	num 10
3,386,355		range 0.0075 to 15.6243

Daily TOX: BUTADIENE13	APT0805	num 11
3,386,355		range 0.0000 to 2.1388

Daily TOX: ETOH	APT0806	num 10
3,386,355		range 0.0046 to 81.8674

Daily TOX: FORM	APT0807	num 9
3,386,355		range 0.0132 to 39.4403

Daily TOX: NAPHTHALENE	APT0808	num 11
3,386,355		range 0.0000 to 1.0410