



## APPENDIX AVAILABLE ON THE HEI WEB SITE

### Research Report 169

#### **Effects of Short-Term Exposure to Air Pollution on Hospital Admissions of Young Children for Acute Lower Respiratory Infections in Ho Chi Minh City, Vietnam**

**HEI Collaborative Working Group on Air Pollution, Poverty, and Health in Ho Chi Minh City**

#### **Appendix F. SOP for Derivation of Daily Pollutant Data**

Note: Appendices Available on the Web may appear in a different order than in the original Investigators' Report, and some remnants of their original names may be apparent. HEI has not changed the content of these documents, only the letter identifier.

Appendix F was originally Appendix B

---

Correspondence may be addressed to Dr. Long Ngo, Harvard Medical School, Department of Medicine, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Room 130, Boston, MA 02215; e-mail: [lngo@bidmc.harvard.edu](mailto:lngo@bidmc.harvard.edu).

The study by the HEI Collaborative Working Group on Air Pollution, Poverty, and Health in Ho Chi Minh City was supported with funds from HEI's Public Health and Air Pollution in Asia (PAPA) Program, the Poverty Reduction Cooperation Fund of the Asian Development Bank (Technical Assistance 4714-VIE), and in-kind support from the government of Vietnam. The contents of this document have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties and no endorsement by them should be inferred.

This document was reviewed by the HEI Health Review Committee but did not undergo the HEI scientific editing and production process.

© 2012 Health Effects Institute, 101 Federal Street, Suite 500, Boston, MA 02110-1817

## **Appendix B: SOP for Derivation of Daily Pollutant Data**

HEALTH EFFECTS INSTITUTE  
ASIAN DEVELOPMENT BANK  
HO CHI MINH CITY'S DEPARTMENT OF  
HEALTH

---

The Effects of Short-Term Exposure on Hospital Admissions for Acute Lower Respiratory Incidence in Young Children in Ho Chi Minh City

**Standard Operating Procedure  
Derivation of Daily Pollutant Data**

Effective Date: April 8, 2010

**Reviewed and Approved by:** Long Ngo and Sumi Mehta

## 1.0 Purpose

This standard operating procedure (SOP) describes the process which will be used to extract and convert data from the hourly frequency format to the 24-hour format for the exposure pollutants NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, and PM<sub>10</sub>.

## 2.0 Personnel

- 1) Project investigators and our senior statisticians will provide technical guidance
- 2) Staff from HEPA (Dam) will be responsible for extracting and deriving the hourly data.
- 3) Staff from Quality Assurance (QA) unit will be conducting testing and validation of the procedure and final 24-hour daily data.
- 4) Study coordinator will provide general support to all phases of this procedure (Mr. Dan Vu).

## 3.0 Procedure

1) 5 minute data: These data will be used to cross-check against HEPA-derived hourly data for quality assurance purposes.

HEPA staff will extract the 5-minute data for NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, and PM<sub>10</sub> from January 01, 2003 to December 31, 2005 for each of the four non-traffic stations which has all or some of these pollutants: ZOO, QT, TSH, and D2.

5 minute data

- Extract five minute data from EnviMan Reporter software by year and station for all pollutants (NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>).
- Use tab delimited option to extract data to text files

The format for DATE is mmddyyyy

The format for TIME is 00:00:00

The structure of the 5-minute dataset will be as follows:

The name of the data set will have the format of SITE\_YEAR

The raw 5-minute data sets will have the following variable names:

D2 station: DATE, TIME, NO2, SO2, O3, PM10

ZOO station: DATE, TIME, NO2, SO2, O3, PM10

QT station: DATE, TIME, NO2, SO2, O3, PM10

TS station: DATE, TIME, NO2, SO2, O3, PM10

2) Hourly data screening:

- Hourly values for each monitoring site will be manually reviewed to flag recurrent values. All strings of four or more repeated values, indicating a problem with the monitoring system, will be deleted.

- Site-specific daily time-series will be reviewed to identify other potential data quality concerns.

Sections of the data series which are extremely sparse and/or considered implausible will be removed.

3) Daily average pollutant values will be created for each monitoring station by taking the mean of 24 hourly values for PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub>, and by generating maximum 8 hour moving averages for O<sub>3</sub>. A 75% completeness criterion was applied to all hourly data. No additional constraints to the data, i.e. no thresholds, will be applied. Tabulate the number of missing days for daily data distribution for these overall estimates for each pollutant by year for each stations.

4) From the daily data, take the mean (average) of the pollutants across stations. Data from all available stations after step 3) above should be used. Tabulate the number of missing days for daily data distribution for these overall estimates for each pollutant by year.

5) The final daily estimate dataset will be named POLLUTANT\_ALRI and has the following variables:

Monitoring Site	Pollutant			
	PM10	NO2	SO2	O3
District 2	Jan 2003 –	Jan 2003 –	Jan 2003 -	Jan 2003 –
	Dec 2005	Dec 2005	Apr 2005	Dec 2005
QT	Jan 2003 –		Jan 2003 -	Jan 2003 –
	Dec 2005		Apr 2005	Dec 2005
TSH	Jan 2003 –			Jan 2003 –
	Dec 2005			Dec 2005
Zoo	Jan 2003 –	Jan 2003 –		Jan 2003 –
	Dec 2005	Dec 2005		Dec 2005

6)

DATE : Real date in the format of MMDDYYYY

DAY: The sequence of day from 1 to 1096 (3 years of daily data)

MONTH: Real month from 1 to 12

YEAR: Value of 2003, 2004, 2005

WKDAY: Weekday with 1: Monday, to 7: Sunday

SEASON: 1 for rainy season (May – October), and 0 for dry season (November – April)

HOLIDAY: 1 for holidays, 0 for non-holidays

D2\_NO2: Daily estimate of NO2 for site D2

D2\_O3: Daily maximum 8 hourly average estimate of O3 for site D2

D2\_PM10: Daily estimate of PM<sub>10</sub> for site D2

D2\_SO2: Daily estimate of SO2 for site D2

QT\_NO2: Daily estimate of NO2 for site QT (daily estimate is from median of hourly data)

QT\_O3: Daily maximum 8 hourly average estimate of O3 for site QT

QT\_PM10: Daily estimate of PM<sub>10</sub> for site QT

QT\_SO2: Daily estimate of SO2 for site QT

TSH\_O3: Daily maximum 8 hourly average estimate of O3 for site TSH

TSH\_PM10: Daily estimate of PM<sub>10</sub> for site TSH

ZOO\_NO2: Daily estimate of NO2 for site ZOO

ZOO\_O3: Daily maximum 8 hourly average estimate of O3 for site ZOO

ZOO\_PM10: Daily estimate of PM<sub>10</sub> for site ZOO

NO2: Daily estimate of NO2 from taking the mean of sites with NO2

O3: Daily maximum 8 hourly average estimate of O3 from taking the mean of sites with O3

PM10: Daily estimate of PM<sub>10</sub> from taking the mean of sites with PM<sub>10</sub>

SO2: Daily estimate of SO2 from taking the mean of sites with SO2

For the final analysis, a series of lag variables was added to this dataset. For all of the variables above, we created lags 1 to 10, and mean of lag 1 to 6. We focused on the mean of lags 1 to 6 for the pollutant variables in the final analysis.

The data will be maintained in EXCEL, SAS, and R format. Changes to the source data will be made to the hourly data and the entire procedure above will be re-run to derive the final dataset POLLUTANT\_ALRI.

#### **4.0 Testing and Validation by Quality Assurance (QA)**

The QA unit will have access to all 5-minute, hourly, and daily data.