HEI

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 E. SOP for Construction of the Clinical Database from Children's Hospitals 1 and 2

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 E was originally Appendix A

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*.

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

Appendix A: SOP for Construction of the Clinical Database from Children's Hospitals 1 and 2

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
Construction of the Clinical Database from Children Hospitals 1 and 2

Update Date: October 05, 2006
Review Date: November 01, 2007

Reviewed and Approved by: Long Ngo, Sumi Mehta, Dzung Do, Thuan Thach

1.0 Purpose

This standard operating procedure (SOP) describes the process which will be used to construct the clinical database from the children hospital 1 (CH1), and hospital 2 (CH2).

2.0 Personnel

1) Project investigators and our senior statisticians will provide technical guidance
2) Staff from the two hospitals will be responsible for collecting the data
3) Staff from Quality Assurance (QA) unit will be conducting testing and validation of the final database.
4) Study coordinator will provide general support to all phases of this procedure.

3.1 Procedure for Admissions Data

1) The clinical database will consist of records that meet all of the following conditions:

   a) Admission date from January 01, 2003 to December 31, 2005
   b) Age at admission date is less than 5 years
   c) Residence of Ho Chi Minh City on admission date who did not live in one of the five rural districts:
Huyện Củ Chi, Huyện Hóc Môn, Huyện Bình Chánh, Huyện Nhà Bè, Huyện Cần Giờ
   d) **Admission diagnosis** includes primary diagnosis code ICD10 of
        J10 to J18, and J20 to J22 and / **or**
        Discharge diagnosis includes primary diagnosis code ICD10 of
        J10 to J18, and J20 to J22
   e) If a child was admitted more than once during this time, all records of this child will be retrieved

Name the clinical database CH1_CLINICAL for CH1, and CH2_CLINICAL for CH2

2) Each record will have the following variables:

   HOSPID: The unique hospital ID assigned to the child upon admission
   CASEID: The unique Case Record Number
   FIRSTNAME: The child's first name
   LASTNAME : The child's last name
   DOB: The child's date of birth, formatted as dd/mm/yyyy (e.g. 04/12/2004)
   GENDER: The child's gender (code 1 for Male, 2 for Female)
   CITY: City (should be only for HCMC)
   DISTRICT: Code for District (this is an important variable used for district-level poverty definition) – see appendix A for the code list which is from HCMC Bureau of Statistics)
   WARD: Code for Ward
   ADMITDATE: Admission Date, formatted as dd/mm/yyyy
   DISCHARGEDATE: Discharge Date, formatted as dd/mm/yyyy
   ICD10ADMIT: ICD10 code for primary admission diagnosis
   ICD10DISC: ICD10 code for primary discharge diagnosis

3) Create the following additional datasets from 1 above
        a) Exclude repeated visits occurring **within the same 14 day period** (as per RESPIRE)
        b) Create three additional hospital admission databases for pneumonia and influenza only (J10-J18) (called it PNEUINFLU), J21 bronchioliti (called it BRONCHI) and other ALRI (J20-J22) (OTHERALRI)

3.2 FINANCIAL DATA

CH1: Payment status information is available in the clinical database for CH1, the financial database also has more details such as the amount the patient owed, how much the patient has paid, and the reason for nonpayment or reduced payments. The financial database from CH1 also has insurance information which indicates the type of insurance (veterans, free insurance due to being poor) the patient has in order to be eligible for nonpayment or reduced payment. After June 2005 all children have been eligible for free health care, so the effect modification analysis will be limited to June 2005.

Therefore for CH1, a database named CH1_PAYMENTSTATUS will be constructed from the financial database. CH1_PAYMENTSTATUS will have the following variables:

   HOSPID: The unique hospital ID assigned to the child upon admission
   CASEID: The unique Case Record Number
   FIRSTNAME: The child's first name
   LASTNAME: The child's last name
   AMOUNTOWED: The total amount owed to the hospital
   AMOUNTPAID: The total amount paid to the hospital
   PAYMENTSTATUS: Code for payment status (e.g. 1: fully paid, 2: reduced fee, 3: ran away, 4: free, 5: insurance)

CH2: The financial database exists; however, there is no way to link the financial database records with the records in the clinical database. The financial database records have first and last name, age of the child, and information on amount owed and paid. However, there is no consistency in how the first, last name, and age were recorded. Using these variables to link with the clinical database without thoroughly checking would like cause the link to be unreliable. To establish linkage to the clinical database CH2_CLINICAL (created above in step 1 and 2), chart review and manual checks need to be done to construct the financial database with the primary key ID consistently recorded for each child. Using the first, last name, age and possible other information in the financial records (there are other variables in the financial records such as addresses and insurance card IDs that were not entered into the financial database), the ID can be assigned to each child in the financial record. Financial records for CH2 will have to be extracted manually from January to August 2003. Therefore, the financial database for CH2 will be constructed with the name CH2_PAYMENTSTATUS that has the following variables:

   HOSPID: The unique hospital ID assigned to the child upon admission
   FIRSTNAME: The child's first name
   LASTNAME: The child's last name
   AMOUNTOWED: The total amount owed to the hospital
   AMOUNTPAID: The total amount paid to the hospital
   PAYMENTSTATUS: Code for payment status (e.g. 1: fully paid, 2: reduced fee, 3: ran away, 4: free, 5: insurance)

5) Link clinical and financial data for CH1 by merging CH1_CLINICAL with CH1_PAYMENTSTATUS. Form the new dataset and call this CH1_ALRI.

6) Link clinical and financial data for CH2 by merging CH2_CLINICAL with CH2_PAYMENTSTATUS. For the new dataset and call this CH2_ALRI.

7) Concatenate CH1_ALRI and CH2_ALRI, and call this new dataset CH_ALRI. This is the final dataset for the data from both pediatric hospitals.

8) For the final analysis, a decision based on a sequence of earlier analyses was made to focus the final analysis a subset of the data created above. The J code was specified to include only the following code: pneumonia (13, 14, 15, 16, 18), and acute bronchiolistis (21). The final clinical data base has the name of CH_ALRI_PNEUMO_BRONCHI which as a total of 15,717 observations (5249 from CH1, and 10468 from CH2).

4.0 Testing and Validation by Quality Assurance (QA)

The QA unit will have access to all datasets created above (CH1_CLINICAL, CH1_PAYMENTSTATUS, CH1_ALRI, CH2_CLINICAL, CH2_PAYMENTSTATUS, CH2_ALRI, CH_ALRI). They will perform the following specific tasks to test and validate these datasets:

CH1 staff will serve as QA staff for CH2 and CH1 staff will serve as QA staff for CH2.

1) The QA staff will check the naming convention and format for each variable in all datasets to make sure that the variables were named as indicated in this SOP, and the format (date, time, code) is strictly followed.

2) The QA staff will list the frequency distribution of all variables in CH1_CLINICAL, and CH2_CLINICAL, including the percentage of missing data for each variable.

3) The QA staff will randomly select 5% of the records from CH1_CLINICAL and CH2_CLINICAL and go back to clinical charts to verify that the information recorded in the database matches the information recorded in clinical charts.

4) The QA staff will list the frequency distribution of all variables in CH1_PAYMENTSTATUS, and CH2_PAYMENTSTATUS, including the percentage of missing data for each variable.

5) Using the same records from the 5% selected in step 3, QA staff will go back to the paper financial records to verify that the information recorded in the database matches the information recorded in CH1_PAYMENTSTATUS, and CH2_PAYMENTSTATUS.

6) Check these 5% records in CH1_ALRI and CH2_ALRI to make sure that the linkage worked properly (i.e. the ID in CH1_ALRI has information of the same ID from CH1_CLINICAL, and CH1_PAYMENTSTATUS, and likewise for CH2).

7) Check the total number of records in CH_ALRI to be the sum of the records from CH1_ALRI and CH2_ALRI.

8) Extract ID, FIRSTNAME, LASTNAME, DOB of all records from CH_ALRI with nonpayment or reduced payment, and check with the DOLISA database to see what percentage of these IDs are included in the DOLISA database.

9) SES Validation: Choose 50 'non-poor' cases per year from each hospital (300 total). Take these cases to DOLISA to see how many of these are recorded as poor by DOLISA.

10) Internal data consistency: the database will be checked for internal consistency with regard to the following variables: age versus date of birth, diagnosis at admission versus diagnosis at discharge, diagnosis and dates of admission for multiple admissions, text diagnosis versus ICD10 code.

11) The QA staff will write a report detailing the results of the QA work from step 1 to step 8 above. The study investigators and team will determine if these data from CH1 and CH2 are approved for statistical analysis.

5.0 APPENDIX

District Code	District Name
1	District 1
3	District 2
5	District 3
7	District 4
9	District 5
11	District 6
13	District 7
15	District 8
17	District 9
19	District 10
21	District 11
23	District 12
25	District Gò Vấp
27	District Tân Bình
28	District Tân Phú
29	District Bình Thạnh
31	District Phú Nhuận
33	District Thủ Đức
34	District Bình Tân
35	Huyện Củ Chi
37	Huyện Hóc Môn
39	Huyện Bình Chánh
41	Huyện Nhà Bè
43	Huyện Cần Giờ