APPENDIX AVAILABLE ON REQUEST

Research Report 157, Public Health and Air Pollution in Asia (PAPA): Coordinated Studies of Short-Term Exposure to Air Pollution and Daily Mortality in Two Indian Cities

Part 1. Short-Term Effects of Air Pollution on Mortality: Results from a Time-Series Analysis in Chennai, India

Kalpana Balakrishnan et al.

Part 2. Time-Series Study on Air Pollution and Mortality in Delhi

Uma Rajarathnam et al.

Appendix B. Common Protocol for Time-Series Studies of Daily Mortality in Indian Cities

---

Correspondence for Part 1 may be addressed to Dr. Kalpana Balakrishnan, Professor and Head, Department of Environmental Health Engineering, Sri Ramachandra University, Porur, Chennai 600 116, India. *kalpanasrmc@vsnl.com*. Correspondence for Part 2 may be addressed to Dr. Uma Rajarathnam, Enzen Global Solutions, 90, Madiwala, Hosur Road, Bangalore 560 068, India. *uma.r@enzenglobal.com*.

The PAPA Program was initiated by the Health Effects Institute in part to support the Clean Air Initiative for Asian Cities (CAI-Asia), a partnership of the Asian Development Bank and the World Bank to inform regional decisions about improving air quality in Asia. Additional funding was obtained from the William and Flora Hewlett Foundation. 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.

January 30, 2007

PROTOCOL FOR TIME-SERIES STUDIES OF DAILY MORTALITY IN INDIAN CITIES

I. RATIONALE

Metropolitan cities across the world have a multitude of environmental concerns that affect human health. Burgeoning populations, worsening air quality, polluted rivers, congested roads, and hazardous occupational environments - all represent problems faced by environmental managers of all countries. Addressing these issues in developing country settings is all the more challenging as these compete with each other for limited resources allocated for environmental improvement.

In India although the available national statistics on environmental quality clearly identify the need for addressing environmental concerns within development plans, environmental management plans often have had to rely on indirect assessments of the impact of local environmental degradation. The State and Central Pollution Control Boards are vested with a large share of the responsibility to collect information on environmental quality along with other agencies such as the National Environmental Engineering Research Institute. The collection of this data is largely driven by the need to document compliance to the adopted national standard without explicit requirements for modeling health impacts on the exposed population.

Environmental epidemiological studies in India have largely been focused on gathering cross sectional prevalence information on environmental health endpoints. Very few studies have attempted generating exposure – response functions to enable health impact assessments. The critical challenge in conducting such studies is usually the lack of good quality exposure data. Even when available they are not in linkage with health data sets on the same population. The differences between developed countries and India in terms of multiplicity of exposures, differences in base-line health status and the sheer differences in exposure levels (exposures for several criteria pollutants are often more than an order of magnitude higher than encountered in developed country settings) contributes to uncertainties in extrapolating concentration-response information across the countries and estimating risks quantitatively. Risk perception and risk communication mechanisms continue to be handicapped either due to the lack of locally derived relationships that reduce acceptability or due to lack of technical understanding on methodologies that limits transferability across settings, resulting in a gross lack of initiative for environmental management. There is thus a need to generate a critical mass of local estimates so that they may be leveraged with estimates from within the region and globally to enable implementation of interventions without undue protraction.

Recent comparative risk assessment exercises carried out by the World Health Organization to estimate the burden of disease attributable to select risk factors indicate that environmental factors may contribute to nearly 30% of the total disease and disability burden in India with children bearing a large share of this burden. The same exercise also suggests the majority of the burden attributable to outdoor- air pollution comes from developing countries. Further, recent meta-analyses (Cohen AJ, Anderson

HR, Ostro B, et al. 2004[1]; PAPA Review) also suggest that proportional increases in daily mortality per 10μg/m^3 increase in PM_{10} are similar among North America, Western Europe, and developing countries. However, the worldwide data have not been appropriately analyzed to determine whether there are real differences in the magnitude of the effects of short-term exposure across regions. Also, there are relatively few meta-analysis studies in Asia. Most studies are not geographically representative, and have taken inconsistent approaches to the definition of health outcomes and data analyses that complicate comparisons with each other and with the broader literature.

The proposed project represents a pilot effort, whereby the acute effects of outdoor air pollutants on mortality would be assessed through retrospective time -series analyses in three cities in India, viz. Chennai, Delhi and Ludhiana with a view to generate baseline exposure-response information.

The time-series studies of daily mortality in 4 other Asian countries are already underway within the PAPA program. Integration of the Indian studies with these on-going Asian studies especially if designed from the start to be comparable, would enhance region-specific combined analyses, providing more definitive estimates of local effects for decision makers. Efforts to bring the world's data together for such analyses are underway with funding from HEI and the EC in the APHENA project. These efforts would also be strengthened by the additional variability in air pollution, climate and population characteristics that Asian studies could contribute.

II. SPECIFIC OBJECTIVES

The specific objectives of the Indian studies are to:

1. Describe the distribution of 24-hour averages of criteria air pollutant levels across all air quality monitoring networks within the three cities over a 3 year period from 2002 to 2004.
2. Retrieve and organize data on all cause mortality excluding reported non-natural deaths over the same period.
3. Develop a protocol for the conduct of a time-series analyses using the exposure and mortality datasets for individual studies.
4. Develop a protocol for the design and analysis of data from multiple Indian cities
5. Develop a management framework to conduct the coordinated analysis.
6. Conduct a coordinated analyses using the exposure and health datasets from the three cities.
7. Contribute to the pool of Asian studies under the PAPA program for meta-analyses.
8. Report the results of the coordinated analyses in an HEI final report and papers in the broader peer-reviewed literature
9. Build capacities in time–series analyses for the study teams and improve abilities for teams to transfer these skills for additional studies in the future within India and the region.
10. Contribute to the development of improved routine air quality monitoring programs and recording of deaths.

[1] Cohen AJ, Anderson HR, Ostro B, Pandey KD, Krzyzanowski M, Kuenzli N, Gutschmidt K, Pope CA, Romieu I, Samet JM, Smith KR. 2004. Mortality impacts of urban air pollution. In: Comparative Quantification of Health Risks: Global and Regional Burden of Disease Due to Selected Major Risk Factors (Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds), vol 2. World Health Organization, Geneva, Switzerland.

III. METHODOLOGY

This section describes the overall approach to data collection, analyses and reporting for all three participating centers in India. Wherever needed, city specific issues/methods are also presented. Finally, the framework for team interactions and co-ordination is also described.

A. BACKGROUND INFORMATION ON PARTICIPATING CENTERS

i. Participating research centers

Selection of cities has been governed by interests expressed by existing investigators through responses to RFIQs issued by HEI. Three cities Chennai, Delhi and Ludhiana were selected to conduct the time –series analyses. The studies are being conducted by The Sri Ramachandra Medical College & Research Institute (SRMC &RI), The Energy Research Institute (TERI) and The Post Graduate Institute for Medical Education and Research (PGIMER), respectively.

ii. Description of individual studies

- **Chennai*:***

Study design will center on integrating and analyzing data from two principal **retrospective data** gathering components viz. collection of daily average environmental levels of criteria air pollutants(PM 10, SO_2 and NOx) and collection of daily mortality data (city wide) over the period 2002 to 2004. Air quality data will be obtained from the Chennai city air quality monitoring network stations currently operating at eight locations by the Tamil Nadu State Pollution Control Board. Each station monitors criteria air pollutants (PM 10, NOx and SOX) on 100 to 140 days in a year. Mortality data will be collected from the Chennai City Corporation that maintains an electronic database of deaths registered across ten zonal offices within the city. With a population of around 6 million, on an average about 60 to 100 deaths are recorded everyday in Chennai. Data on meteorological parameters will be collected from the Indian Meteorological Office at Chennai. Additional details on data collection and cleaning are provided in the next sections. The location of air sampling stations and the geographical boundaries for the data collection are shown in Annexure 3.

The Chennai study will be executed by Dr. Kalpana Balakrishnan (PI) from SRMC, Dr. Bhaswati Ganguli (Lead Bio-statistician) from Calcutta University and Mr. Sankar (Project Co-ordinator) from SRMC along with a team of research fellows. Dr. VN Rayudu from the Tamil Nadu State Pollution Control Board and Dr. Harry Caussy from the SEARO office of the WHO will serve as external co-investigators.

- **Delhi:**

The study uses retrospective time series data for the period of 2002 – 2004 on air quality and registered data on deaths in Delhi to study the changes in daily death counts due to change in air quality levels. The methodology involves: (i) Collection of 24 hour average levels of criteria air pollutants namely PM_{10}, SO_2 and NO_2 for the study period between 2002 -2004, across the 11 locations being monitored by the Central Pollution

Control Board (CPCB), National Environmental Engineering Research Institute (NEERI) and The Energy and Resources Institute (TERI). (ii) Visits to the monitoring sites and collection of information about the monitoring site to evaluate the site selection criteria and to determine the suitable sampling location(s) that could represent the population exposure in the city of Delhi. (iii) Review of monitoring protocols, frequency, QA/QC procedures for data harmonization. (iv) Collection of meteorological data for the study period (v) Collection of death records from the Registrar of Births and Deaths, Municipal Corporation of Delhi and New Delhi Municipal Corporation and (vi) statistical analysis of the data to study the change in daily deaths (total death excluding death due to accidents) with daily air quality levels after controlling for weather parameters by adopting common protocol with city specific modifications.

Detailed description on study area, air quality monitoring locations are provided in Annexure 3.

The study team includes Dr. R Uma (PI), Dr. Subramanya Nairy (Bio-statistician), Ms. Meena Seghal, Mr. Santhosh Raghavan from TERI and Dr. R C Patnaik from the Municipal Corporation of Delhi, Dr Ramesh Kumar from New Delhi Municipal Council (NDMC), Dr. S K Chhabra from Vallabai Patel Chest Institute and Dr. Kilnani from All India Institute of Medical Sciences. Data collection was done with the help of field assistants.

- **Ludhiana:**

Study design will center on analyzing data from two retrospective data gathering components viz. collection of daily average environmental levels of criteria air pollutants (RSPM, SO_2 and NOx) and collection of daily mortality data over a period 2002 to 2004. The Punjab Pollution Control Board collects daily air quality data (on working days only) from 4 centers in the city (on an average 90 to 140 days in a year). The data from each station is compiled electronically at the city office of State Pollution Control Board. The electronic air quality data provided by Punjab Pollution Control Board for the period January 2002 to December 2004 will be used. Parameters include 8 hour and 24- hour averages of RSPM and 4 hour and 24-hour averages for SO_2 and NOx. Standard procedures for data collection are adopted as per Bureau of Indian Standards (BIS). The electronic data set will be compared with the field data registers. For the calculation of 24-hour average concentration of RSPM, NO_x and SO_2, there must be at least 75% of the total monitoring duration available on that particular day. With a population of around 1.5 million, on an average about 9 to 50 deaths are recorded per day in Ludhiana. Data on meteorological parameters will be obtained from Regional Meteorological Centre, New Delhi for the Ludhiana City. The location of air sampling stations and the geographical boundaries for the data collection are shown in Annexure 3.

The Ludhiana study will be executed by Dr. Rajesh Kumar (PI) with Dr. J.S. Thakur, Dr. S.P.S. Bhatia, Dr. G.P.I. Singh, Dr. H.K. Parwana, Dr. M.L. Garg, Dr. S.K. Jindal (Co-Investigators) and Dr. Suresh Sharma (Lead Statistician).

B. DATA COLLECTION FRAMEWORK

I. Health outcomes

- Based upon the assessment of the quality of mortality records available across the three study cities, only relative rates of **all cause mortality adjusted for deaths reported as due to non-natural causes (i.e. accident, trauma or suicide) will be collected for analyses. Information on the cause of death in the mortality records is highly incomplete and *hence it may be impossible to exclude non – natural deaths altogether*.** Also, since ICD codes for the cause of deaths is not available in any of the cities, investigators will use the assistance of a trained medical professional to flag non- natural deaths based on the listed descriptive cause of death (wherever available).
- Following this, relative rates of daily mortality stratified by sex and age, for the categories of all ages, 0-4, 5-44, 45-64 and 65+ will be estimated. It has been noted that newborn ages are often erroneously expressed in the death records and hence the category 0-4 will be attempted wherever possible.
- Deaths are registered according to the place of death. Deaths of those residing within but dying outside the city limits are not likely to be included in the city records under normal circumstances. Initially all deaths occurring within the city geographic boundaries will be estimated. Following this, exclusion of deaths of those residing outside the city but dying within will be estimated wherever a permanent address is available.

i. Assessment of quality of health outcome data

- Using mortality datasets that contain individual-level information, each city will conduct descriptive analyses to obtain the frequency distributions and/or univariate distributions for all categorical variables (e.g., sex) as well as continuous variables (e.g., age). Investigator in each city will carefully check these distributions for the miscoded, missing, and out of range data. Errors, questions, and/or concerns regarding specific data points will be discussed, validated, answered, and corrected in each city.
- Specifically, duplications and discrepancies in age will be identified by cross-verifications against names, sex, recorded cause of death, day and month of death. A table of all errors so identified, will be compiled for each city.

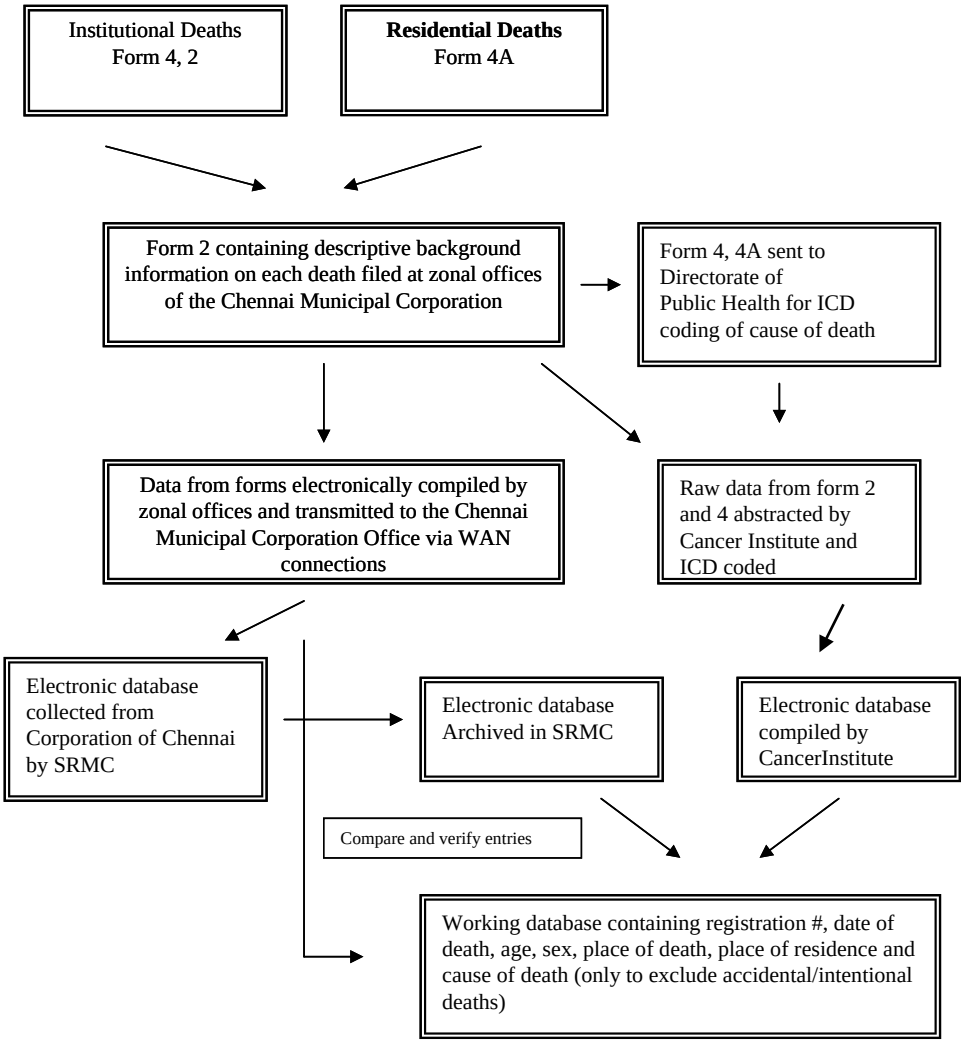
ii. City-specific considerations

- **Chennai**

Chennai team has access to two datasets on mortality- the primary that includes all reported deaths provided by the Chennai Corporation and a modified dataset provided by the National Cancer Institute that records all non-infant deaths that occurred among residents of Chennai city. The Corporation of Chennai is responsible for compiling and maintaining mortality data and vital statistics for the city of Chennai. Birth and Death registrars in the individual Zonal Offices (1 to 10) under the Corporation of Chennai collect the mortality data on forms. This information is compiled electronically at the Zonal Offices. A computerized wide area network reporting system links information between the Zonal Offices and the EDP cell at the Corporation of Chennai. We will use

this electronic database as furnished by the Corporation of Chennai as the primary database. The National Cancer Institute has retrieved information available in the death records (of all non-infant deaths residing within Chennai Corporation limits) for the purposes of creating a cancer registry within the city. The two electronic databases will be linked to track discrepancies in a few fields. Any discrepancies will be verified against individual records available at the individual zonal offices. Based on a pilot retrieval exercise for a period of three months from the individual zonal offices and discussions with the HEI QA audit team during their site visits, only the following fields of information were identified as useful in cross-verification. They include date of death, place of death, place of residence, age and gender. Cause of death information will be retrieved but its use will be limited to exclusion of accidental / intentional deaths. Depending upon the completeness of the place of death and cause of death fields, cause specific and zone specific rates of daily mortality will be attempted for the second stage analyses. A schematic showing the health data collection steps in Chennai is shown in figure 1 below.

Figure 1. Mortality data collection steps used by Chennai centre



- **Delhi**

In Delhi, three Local Bodies viz. Municipal Corporation of Delhi (MCD), New Delhi Municipal Council (NDMC) and Delhi Cantonment Board are carrying out the registration of births & deaths. The Directorate of Economics & Statistics, Govt. of National Capital Territory of Delhi (NCT of Delhi), being the office of the Chief Registrar, Births & Deaths, co-ordinates the work related to the registration of Births & Deaths.

The registration of deaths in Delhi is done by three local bodies viz MCD (188 centres), NDMC (10 centers) and Delhi Cantonment Board (1 centre). About 85 000 deaths per year are registered in Delhi, in which about 67 000 deaths per year were recorded in MCD and about 17 000 deaths per year were recorded in NDMC area. Cantonment area is under the Indian Army and death records from cantonment area, which is less than 2 % of total deaths, are not available. In Delhi, about 55% of reported deaths are institutional deaths (reported from the hospital) and about 45% are reported by households.
At present the scheme of Medical Certification of Cause of Death (MCCD) in Delhi has been confined to the institutional deaths occurred in the hospitals/nursing homes and other private institutions. The domiciliary deaths have not been brought under the ambit of MCCD due to various reasons such as lack of awareness among the citizens and also due to non-availability of adequate infrastructure in medical facilities. **Figures 2 and 3 depict the death registration and data collection system in Delhi.**

Available data from NDMC for the entire study period is not in the electronic form. Hard copy of the death registration form has 2 parts (Personal information for legal part; information for statistical analysis). For the present study, information on the statistical part of the death certificate was made available. Field assistants will be employed to screen the statistical information recorded in the death certificate and entered into the "VITAL" database (created in visual fox pro soft ware) maintained by NDMC. After data entry, records will be transferred into MS Excel database and brought to TERI.

Since 2004, MCD has introduced the online death registration system. In which the all hospital deaths could be registered online and domiciliary deaths will be entered in to the online system through respective MCD local centers. Complied mortality data for the year 2004 in electronic form will be collected from MCD. The data will be crosschecked from the online database available for public.

Figure 2. Mortality data flow for NDMC region

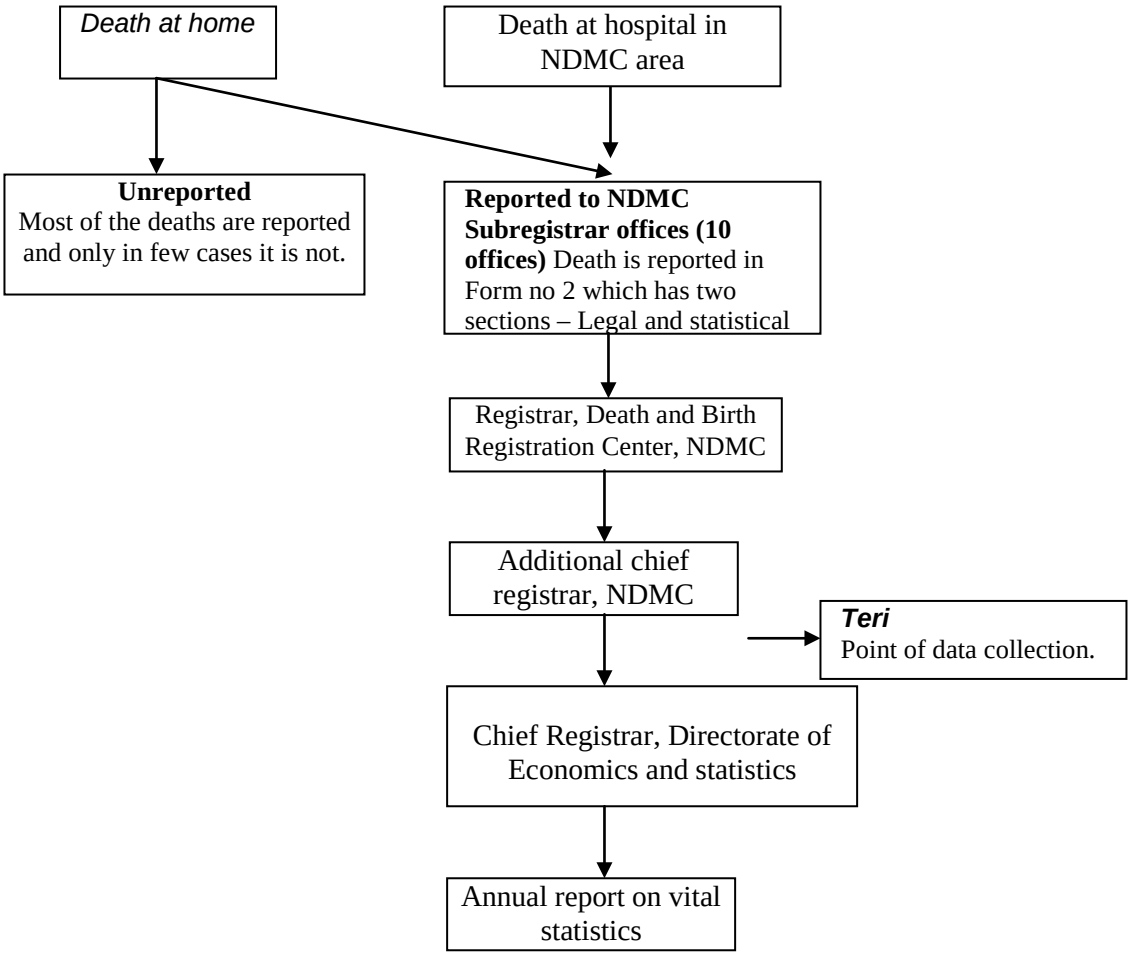
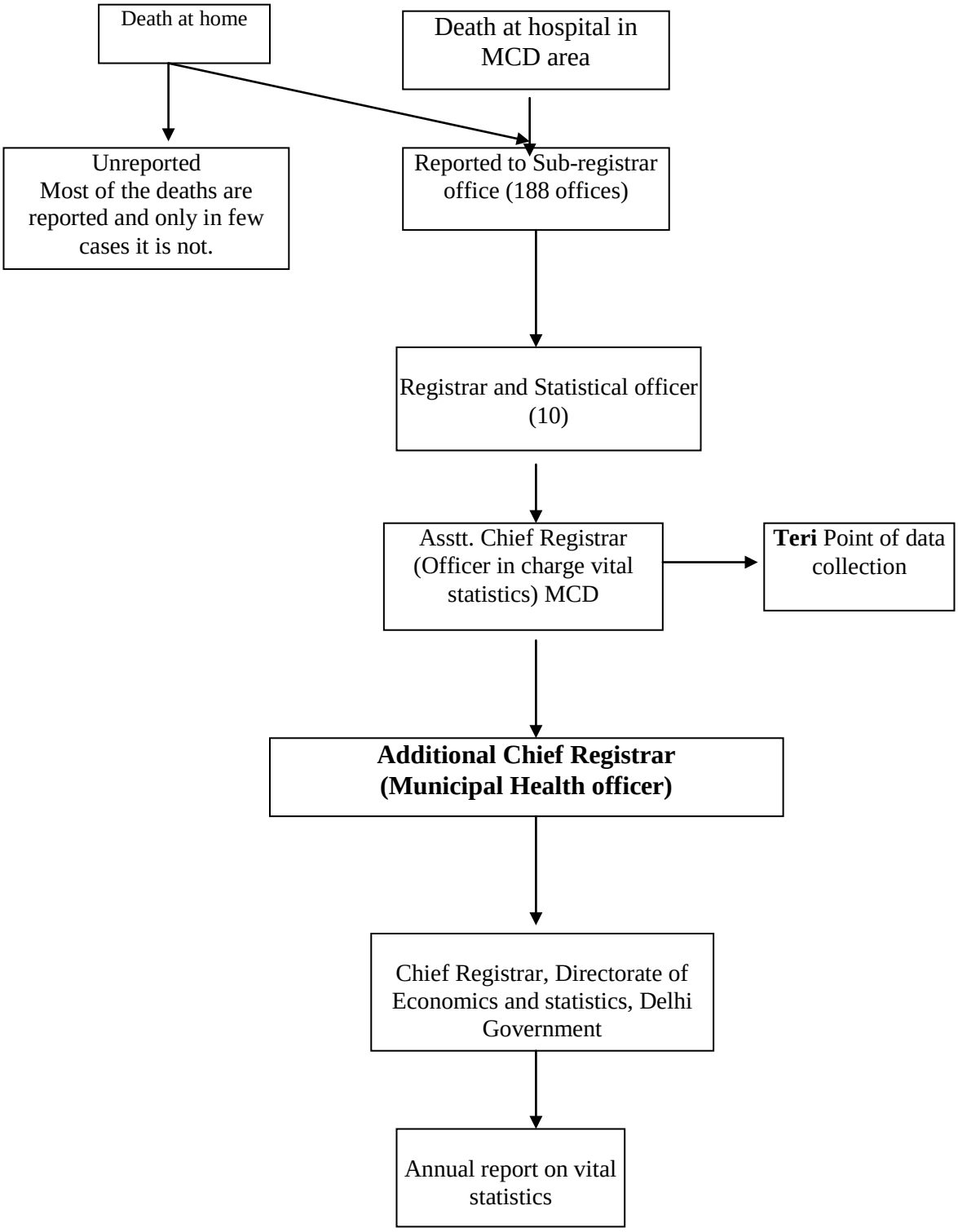


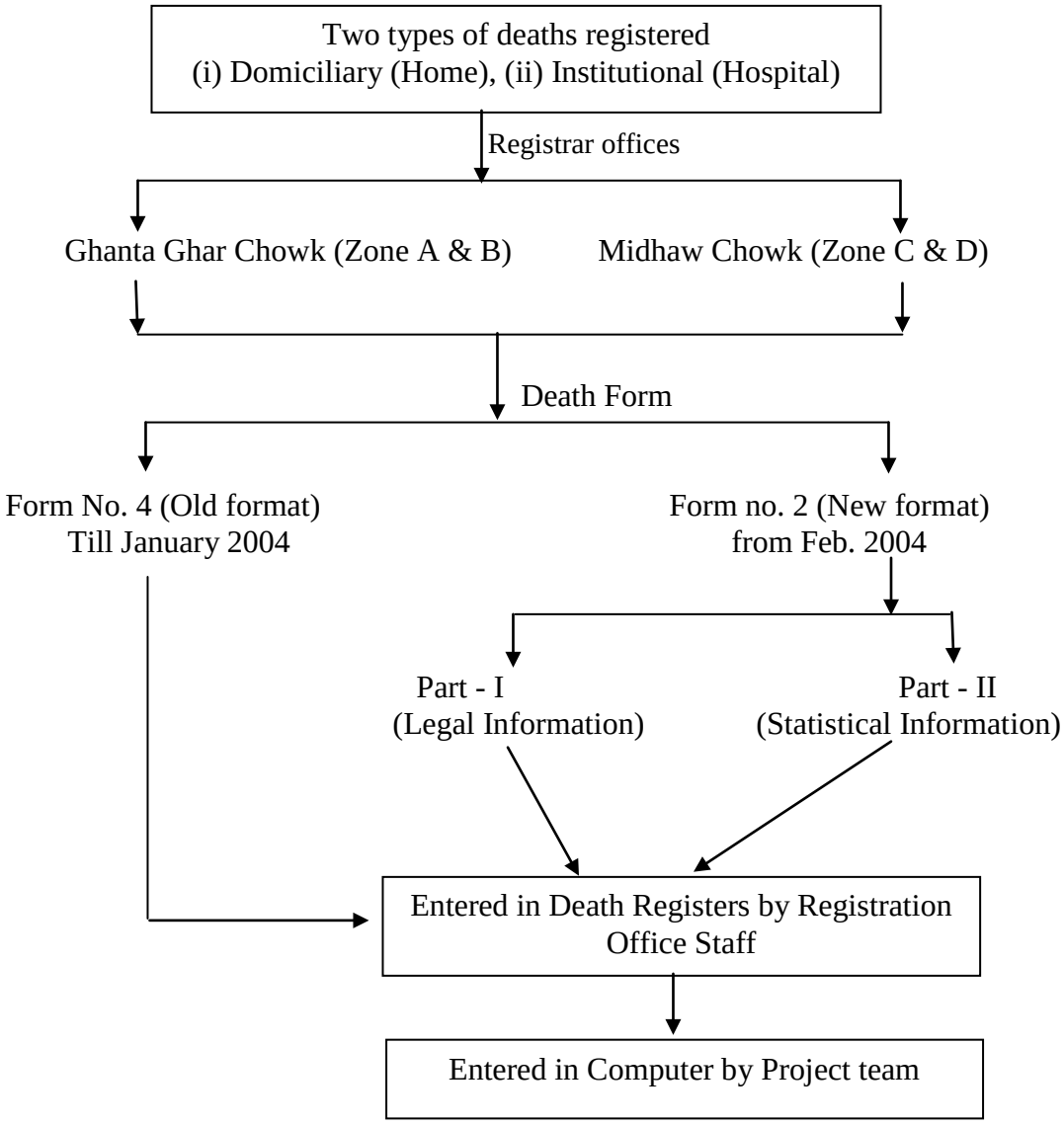
Figure 3. Mortality data flow for MCD region



- **Ludhiana**

Ludhiana City is divided into four Zones A, B, C and D for Birth and Death registration. There are two centers established by the Registrar's office, which collect mortality data for the city, i.e., Ghantaghar Chowk for zone A and B, and Midhaw Chowk for zone C and D. Municipal Corporation collects mortality data from all hospitals/ nursing homes on forms. Those who die at home are reported by family members to the registration office by filling a form. The data from death forms is written by registration clerk in death register. The data available in the death registers in registrar's office will be entered in the computer by project team. Beside identification number, name of the deceased, father/husband name, age, gender, marital status, cause of death, date of death, place of death and address will be included for entry in computer. Every year about 9500 deaths are registered. A schematic diagram showing the mortality data collection steps in Ludhiana is shown below:

Figure 4. Mortality Data Recording in Ludhiana



II. Air Pollution Data

All three Indian cities have a network of monitoring stations operated by the state or central pollution control boards. The initial objectives of collection of air pollutant data would be to describe the distribution of pollutant levels across monitors, assess the degree of completeness of data available from each monitor, assess degrees of correlation and differences in levels recorded across monitors and construct a set of exposure time series. While data on both PM 10 and gaseous pollutants will be used to generate descriptives for distribution of pollutant levels across monitors, it is anticipated that the resolution of the gaseous pollutant measurement methods (i.e. wet chemical methods) may not permit their inclusion in creating alternative exposure time-series data sets. Data on PM 10 will be used to guide monitor selection and gaseous pollutant data will mirror PM 10 data sets in terms of selection of single vs. combinations of monitors. The core model will thus be a single pollutant model centered on PM_{10} with subsequent rounds of analyses including gaseous pollutants

The monitoring period will cover the period between **January 2002 and December 2004.** Pollutants and their respective averaging times to be used are specified below.

Pollutant	**Averaging time**
Sulphur dioxide (SO_2)	24-hr average of multiple 4 or 8 hour measurements
Nitrogen dioxide (NO_x)	24-hr average of multiple 4 or 8 hour measurements
Respirable Suspended Particulate Matter* (RSPM) (median aerodynamic diameter of ~10 μm)	24-hr average of multiple 8 hour measurements

* For the purposes of the Indian studies PM_{10} refers to RSPM measured as per protocols described in Annexure 1.

i. Site selection criteria

Air quality stations in India are classified as industrial, commercial or residential sites based on the land use classification of the zone in which the monitors are located. The following criteria will be used to select eligible stations.

- The sites must be located within the greater metropolitan city boundary as defined by the concerned municipal agency in each city.
- Monitors within city zones designated as industrial hotspots by the Indian Ministry of Environment and Forests will be excluded. Because of mixed land-use profiles in all cities, monitors in zones classified as commercial, industrial and residential by the city municipality will be considered provided they meet the criteria listed below.
- The sites should be large enough to ensure the availability of space for monitoring, and should be located in flat space and elevated between one and 14 feet above ground level.
- Curbside (or roadside) stations will not be included.
- The sites should be located 5 m upwind from building exhausts, vents, chimneys etc. and at least 2 m from walls.

ii. Measurement methods

All stations in all cities use protocols stipulated by the CPCB for monitoring. The details of the methods used are provided in Annexure 1.

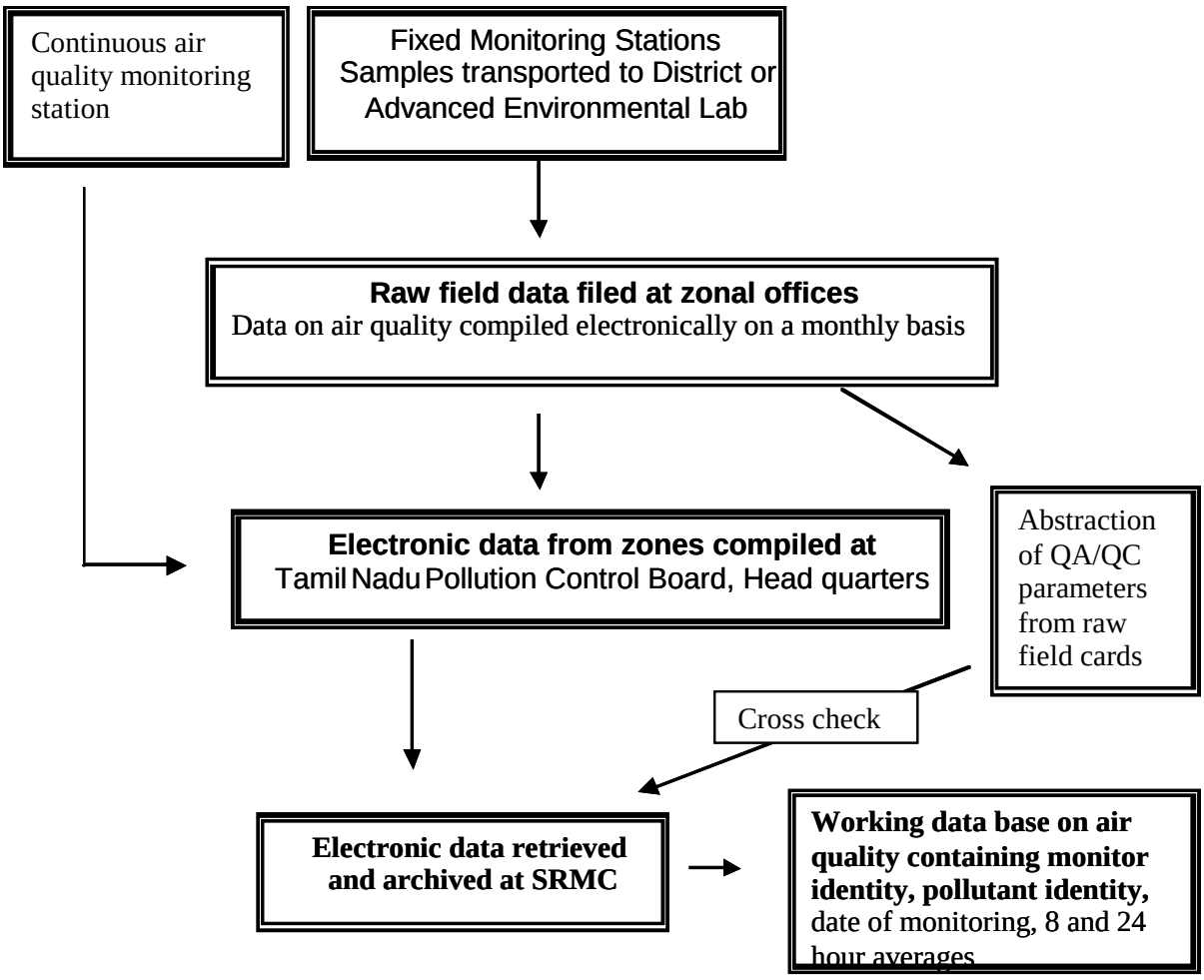
iii. QA/QC for environmental data

All stations follow QA/QC protocols of the CPCB as furnished in Annexure 2. City specific protocols for data cleaning are described below

- **Chennai:**

The Tamil Nadu Pollution Control Board collects daily air quality data (on weekdays only) at 8 centers in the greater Chennai area as well as operates a continuous air quality station co-located with one of these stations. The data from each station is compiled electronically at the State Pollution Control Board. We will use the electronic air quality data provided for the period January 2002 to December 2004. Parameters include 8 hour and 24- hour averages of PM_{10}, SO_2 and NOx. Raw data on QA/QC parameters will be abstracted manually from the field data cards maintained by the individual monitoring station. The electronic data set will be compared with the field data cards for selection of valid data points as per protocols recommended by the Central Pollution Control Board of India. For the calculation of 24-hour average concentration of RSPM it is required to have at all three 8-hour values on that particular day. For NO_x and SO_2 at least 75% of the total monitoring duration has to be available on that particular day. In addition, the continuous air quality monitoring station (co-located with one of the stations (Kathivakkam) in the greater Chennai area) records electronically and the same will be obtained from the State Pollution Control Board. Parameters recorded include half-hourly readings and eight/ 24- hour averages for all three pollutants. The continuous monitoring station deploys a β gauge for PM_{10} measurements and chemi-luminescence methods for gases while the other stations use an indigenous cyclone based gravimetric method for PM and wet chemical methods for gases. Data from the continuous monitoring station will serve as a broad indicator of comparability of methods. A schematic for collection of air pollution data in Chennai is shown below.

Figure 5. Air pollution data collection steps used by Chennai centre



- **Delhi**

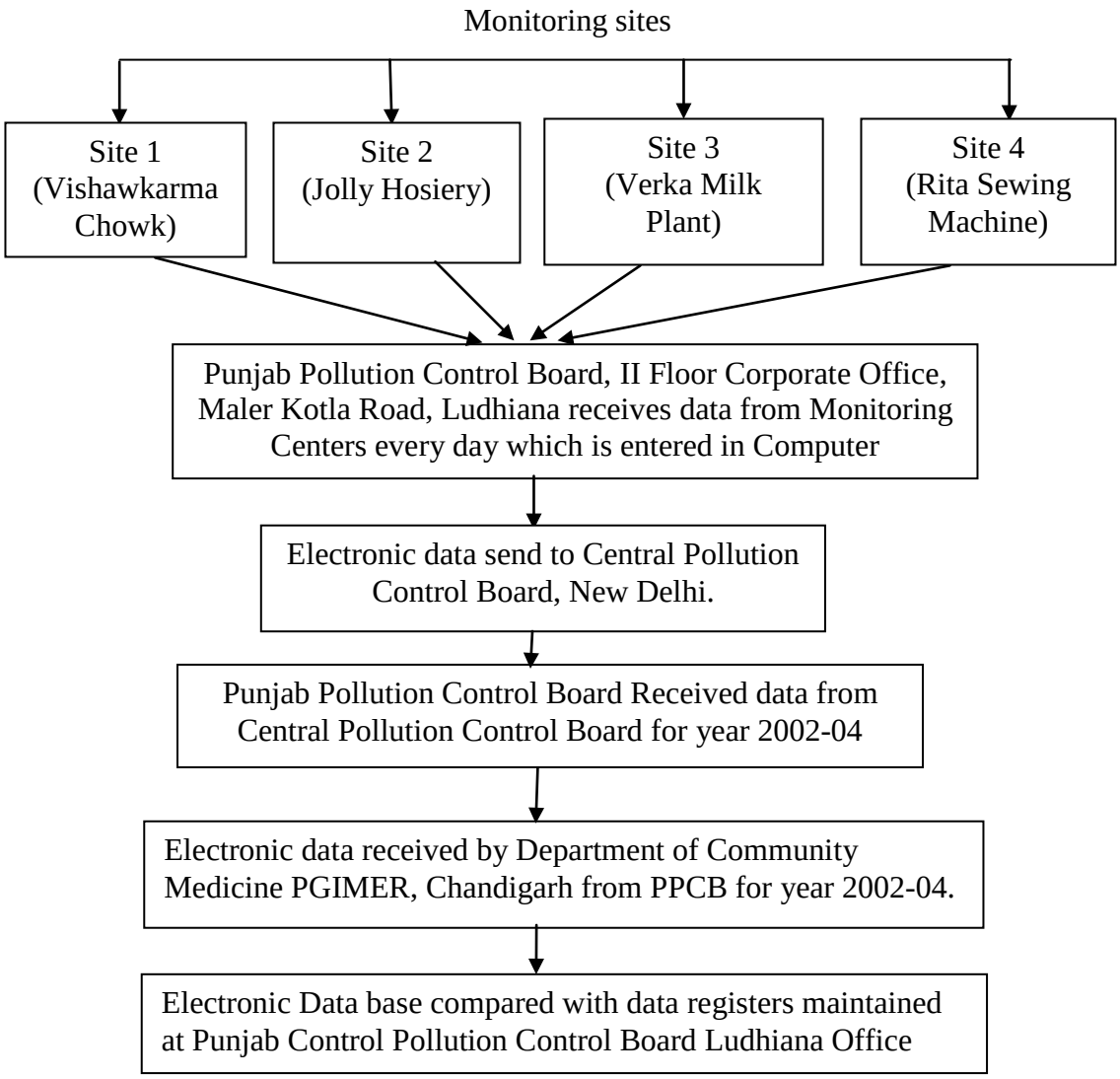
In Delhi, air quality is being monitored at 11 locations; the Central Pollution Control Board (CPCB) directly operates seven sites; National Environmental Engineering Research Institute (NEERI) operates three monitoring sites on behalf of the CPCB as part of the network under the National Air Quality Monitoring Program (NAMP) and one site is operated by The Energy and Resources Institute (TERI). These monitoring sites (shown in Annexure 3) typically attempt to cover two major type of land use: residential and industrial. The site locations, monitoring agency, reporting organization are also given in Annexure 3.

In all these locations, sampling is done as three 8 hour sampling in a day and the values are averaged and reported as 24-hour average value. For the days with monitoring period less than 16 hours/day are not considered for calculating the daily average and not reported. Data for 10 stations obtained from CPCB are 24-hour average values; those were entered in the MS- EXCEL file and used for analysis. Similarly 24 hour average values from IHC monitoring location was used for analysis. However for data validation, raw data (8 hourly values) from IHC are available; whereas 8 hourly average values from CPCB monitoring locations are not available.

- **Ludhiana**

Ludhiana has a network of four monitoring stations operated by the Punjab State Pollution Control Board. Air pollution data can be segregated for four Zones of the City. Air quality monitoring is done for RSPM, NOx and SO_2. Two of these stations operate together three days in a week and another two operate on another set of three days in a week. Measurements are not done on national holidays and Sundays. High volume air sampler is used for air quality monitoring. Air quality data will be obtained in electronic form for 3 years, i.e., 2002, 2003 and 2004. Gravimetric method is used for RSPM and chemical method is used for NOx and SO_2. Standard procedures for data collection are adopted as per Bureau of Indian Standards (BIS). Calibration of equipment and standardization of procedures is done regularly. The electronic data set will be compared with the data maintained in registers at each monitoring station.

Figure 6. Air- Pollution Data Collection Steps, Ludhiana City



iv. Construction of alternative exposure series

In Chennai and Ludhiana each monitoring station is monitored for ~104 days in a year with no monitoring on any site on weekends, holidays and some weekdays. In Delhi, two stations operate year round while 5 stations operate sporadically. Thus there will be missing values in the exposure series for a proportion (in some cases a significant portion) of the study period. In the primary analysis in stage 1, only the actual collected data (based on city specific availability) will be used, and missing data will not be filled in. Single or multiple monitors will be included as per protocols specific for each city (outlined below). In the second stage, (For Chennai and Ludhiana) missing values will be imputed using ANOVA, distribution fitting and penalized spline fitting methods, as may be appropriate for each city (outlined below).

- **Chennai**

In Chennai, of the 8 monitors operated routinely, one monitor (Kilpauk) fails to meet the site selection criteria and hence will be excluded. Three of the monitors are located within an industrial hot spot and hence will not be included in the core model. Of the four remaining, two monitors had quality related problems because of frequent switching of monitor location. Hence, data from the two remaining monitors (Anna Nagar and Vallalar Nagar) will be used for generating the core model. The data from other monitors will be used to generate general descriptives as well as in sensitivity analyses. In the second stage of analyses, simulation exercises will be performed using to generate an imputed series using penalized spline fitting methods. The following exposure series are thus planned for Chennai (series to be used in sensitivity analyses are described later in the data analyses section).

1. Anna Nagar monitor (Series 1)
2. Vallalar Nagar monitor (Series 2)
3. Combination of Anna Nagar and Vallalar Nagar using dummy variables (Series 3)
4. Imputed series obtained by penalized spline fitting on data from all seven monitors (Series 4)

- **Delhi**

Among the 11 monitoring station being monitored regularly, one location (ITO) fails to meet the site selection criteria as it is located in the busy traffic intersection and hence will be excluded. IHC data is available for round the year, however the annual average of RSPM values are on the higher side in comparison with average of other monitoring stations. For other monitoring stations, data is available for about 30% of the days in a year. Considering the correlation coefficient of PM_{10} at IHC monitoring site with other sites (which is in the range of 0.39 to 0.66), use of average of all monitoring stations (excluding IHC values) in the core model seems to be more appropriate. The data from other monitoring stations will be used in sensitivity analysis. In Delhi, inverse of visibility data shows good correlation with PM_{10} values, another set of PM_{10} values were estimated from visibility records. Estimated PM_{10} values will be used in sensitivity analyses.

Proposed exposure series for core model - All station average (excluding ITO values)

- **Ludhiana**

Of four monitoring stations in Ludhiana city viz. Vishawkarma chowk (site 1) and Jolly hosiery (site 2) are commercial, Verka milk plant (site 3) is residential and Rita sewing machine (site 4) is industrial site. The measurement is done three times a week at each site. Air quality is measured at two sites (site1 & site 2) on Monday, Wednesday, and Friday, and at others two sites (site 3 & site 4) on Tuesday, Thursday and Saturday. Each site monitors on an average 110 days in a year. Thus there are many missing days. One of the monitors (Jolly hosiery, site 2) was discontinued in September 2003. Therefore this monitor will not be included in the analysis. For exposure series construction each monitor will be used separately. For sensitivity analysis missing values for each monitor will be generated by following methods

(a) Averaging the available data from a monitor for days adjacent to the missing day.
(b) Predicting missing values by ANOVA for each monitor.

Then the city wide exposure will be computed by averaging the recorded and imputed values from each monitor.

v. PAPA/ISOC request for basic monitoring information

In order to facilitate harmonization and comparison of the information relevant to the exposure assessments in the 3 cities, background information on measurement protocols, number and location of monitors and QA/QC information is described in the guidance manual furnished by the Central Pollution Control Board. Important information from the manual is summarized in Annexure 2.

III. Data on other co-variates

Data on time-varying potential confounders are necessary for the analyses. While in other PAPA studies, meteorological, social, and medical factors have been included, only meteorological parameters will be included for the Indian cities. Lack of routine data on medical factors such as influenza epidemics or social factors such as migration make it difficult to include others in the analyses.

i. Meteorological covariates

In Chennai the maximum and minimum readings over two twelve- hour periods are recorded at two meteorological stations (located at the airport and city center respectively). The daily maximum, minimum and average can be estimated from these datasets. Parameters recorded include temperature, dew point, relative humidity, and visibility.

In Ludhiana three readings are taken for following parameters of meteorological data at 8:30, 11:30 and 17:30 hrs.

- Temperature – Daily three times (in° C)
- Dew Point – Daily three times (in° C)
- Relative humidity – Daily three times (in %)
- Visibility – Thirteen observations per day at interval of one hour period (in meters). Measurements are taken at fixed distances starting from 0 to 100m with

10 meter gap, then 100m to 2000m with 100 m gap and then 2000m to 6000m with 500m gap.

In Delhi meteorological data for the following parameters are recorded at meteorological office, Safdarjung, New Delhi for the study period (2002 – 2004).

- Temperature – Daily maximum and minimum temperature (in° C)
- Rainfall – Daily total rainfall (in mm)
- Relative humidity – Daily minimum and maximum humidity (in %)
- Visibility – Eight observations per day at interval of 3 hour period (in meters)

Meteorological data is recorded in registers maintained by the Indian Meteorological Department in all cities. This data will be abstracted manually from these registers and data entry verified with double data entry for 10% of the observations.

Wind-speed and direction:

Meteorological stations record wind speed and direction across all three cities. These may used to qualitatively describe the measurement variation across monitors. Detailed analyses of windrows to guide monitor selection may not be feasible in any of the cities, although investigators will make an effort to gather this information. In addition dust storm days noted in the field record of IHC location will be recorded in Delhi.

ii. Calendar variables

The following calendar variables will be examined in the analysis
Special events
Season
Day of study (chronological time)
Dummy* variables for:
Official public holidays
Festivals (include dates)
(*Only in Delhi, as no monitoring is done on holidays in Chennai and Ludhiana).

C. DESIGN OF ANALYSIS

i. Basic analysis

For the core model, all three study centers will use the same regression model. Specifically, the procedure will involve the following:

1. Generalized Additive Model (GAM) with penalized spline smoothers in R.
2. Quasi-poison functions with mortality due to all natural causes as the dependent variable.
3. Penalized spline smoothers for time and meteorological parameters.
4. Chennai and Ludhiana will use exposure at single-day lags of 0 to1 day, although problems due to missing weekend data are acknowledged.
5. For Delhi, lag of 0-3days and a cumulative two-day average of lags 0 and 1 will be considered.
6. Delhi team will include auto regression function in the core model as well as in sensitivity analysis

Using this core model, each center will conduct sensitivity analysis on their own data sets (as detailed below).

ii. Sensitivity Analysis

- **Chennai**

Four alternative exposure series will be created as described in the previous section. While series 1 and series 2 represent single monitors, Anna Nagar and Vallalar Nagar without any filling of missing values respectively, series 3 and series 4 are described below

a. Dummy variable technique to combining monitors (Series 3)

Dummy variables will be assigned as follows

$$I_{AN} = \begin{cases} 1 & \text{if only Anagar available,} \\ 0 & \text{otherwise.} \end{cases}$$

$$I_{VN} = \begin{cases} 1 & \text{if only Vnagar available,} \\ 0 & \text{otherwise.} \end{cases}$$

$$I_{AV} = \begin{cases} 1 & \text{if both Anagar and Vnagar available,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} AnagarOnly &= (PMofAnagar) \times I_{AN} \\ VnagarOnly &= (PMofVnagar) \times I_{VN} \\ AnagarCom &= (PMofAnagar) \times I_{AV} \\ VnagarCom &= (PMofVnagar) \times I_{AV} \end{aligned}$$

The model will be defined as

$$\begin{aligned} log(E(Mortality)) &= \alpha_0 + \alpha_1 I_{AN} + \alpha_2 I_{VN} + \alpha_3 I_{AV} \\ &+ \beta_1 AnagarOnly + \beta_2 VnagarOnly \\ &+ \beta_3 AnagarCom + \beta_4 VnagarCom \\ &+ s(time) + s(temp) + s(rh) \end{aligned}$$

The Wald test will be applied to test for differences between these two best monitors. The same approach will also be extended to all seven monitors for the purposes of describing the monitor effects.

b. Generation of an imputed series using penalized spline fitting methods

Natural spline fitting of available PM will be done using specified degrees of freedom calculated using penalised splines or as specified in PAPA first wave study protocols. Missing PM will be predicted using the fitted spline model. For the simulation study, all

seven monitors will be considered. Three different variances for the three categories (i.e. industrial, commercial and residential monitors) will be used to generate a simulated PM series for the 7 monitors from a lognormal distribution taking into account the day effects and monitor effects. Mortality counts will be estimated using a GAM poisson model taking meteorological parameters and time as covariates together with the simulated PM series. PM values will be deliberately excluded according to the data missing-ness structure of Chennai data. Missing values will then be imputed using spline fitting. Bootstrapping standard error correction will be used in imputation and estimation of PM coefficients. The process will be repeated ~200 times to yield the average coefficient and Mean Square Error (MSE).

c. Multiple pollutant modeling

Since SO_2 data are reported as below detection limits on most monitored days multi-pollutant models will exclude SO_2 in Chennai and include only NOx. The same lag for pair of co-pollutants (PM_{10} with NO_x) in the best model developed for all natural causes will be used.

d. Dose-response curves

Smoothing function of each pollutant with 3-4 dfs using natural spline will be fitted for model of all natural causes of death. Y-axis will be the residual after fitting of non-pollutant variables.

e. Stratification

Pollutant data will be stratified on the basis of two seasons hot and warm as the temperature differential between seasons is rather small in Chennai. Monsoons are rather brief and precipitation during other seasons is quite uncommon and hence will not be examined.
Sensitivity analyses for pollution outliers will be performed without prescribing any arbitrary cut-offs.
Testing for auto-correlation may be limited by the lack of availability of data on consecutive days from the same site.

f. Cross validation of results

Each team will validate the estimates derived from models of other teams.

Delhi

Exposure series of individual monitoring locations (10 locations) will be used in the sensitivity analysis.

Exposure series for sensitivity analyses

IHC – series 1
Sarojini Nagar – series 2
Town hall – series 3
Mayapuri – series 4

Nizamudin – series 5
Srifort – series 6
Janakpuri – series 7
Shazadabagh – series 8
Shadara – series
Ashok vihar – series 10
Estimated PM_{10} from visibility – series 11

In addition exclusion of very high levels of RSPM (above 400 $\mu g/m^3$ generally recorded on dust storm days) will be considered in the sensitivity analysis.

Ludhiana

1. Missing data will be computed by following methods.

a. Averaging the available data from a monitor for days adjacent to missing day.

b. Analysis of variance (ANOVA) technique will be tried to predict RSPM.

c. If there is good correlation between visibility data and RSPM, missing value of RSPM will be predicted from visibility data.

2. Multiple pollutant modeling:
Since SO_2 and NOx are reported well below their normal limits, therefore, we will exclude SO_2 and NOx from sensitivity analysis. Moreover, value of SO_2 and NOx do not have much variation.
3. Mortality cause analysis
In Ludhiana cause of death data is not recorded properly and the analysis will be based on natural deaths, by excluding accident death.
4. Age- Specific analysis
Since there are more deaths over 65 years of age, therefore natural spline and penalized supline fitted will be carried out for this group.
5. High levels of RSPM
The exclusion of very high level of RSPM (above 450 mg/m^3) will be considered in the sensitivity analysis.
6.Each team will validate the estimates derived from models of other team.

D. PROJECT COORDINATION AND INTERACTION AMONG INVESTIGATORS

i. Team meetings

The three teams of investigators will meet with each other to share progress and trouble shoot data problems periodically. Tentatively meetings are planned once every 3- 4 months. One member from each site is designated as the first point of contact. These meetings will also be used for additional capacity building within teams on data cleaning/ analyses etc. as well as scope out subsequent project activities/ISOC interactions etc.

The individual teams will be responsible for periodic reports to be submitted to HEI. Reports of teams will be shared to guide progress. A joint report may considered

following the completion of the individual projects to facilitate in –country communication.

ii. Draft timeline for all three cities

Chennai:
Data collection: December 2006 (first phase to be completed December 2005)
Initial Data Audit: December 2005
Completion of data cleaning: December, 2006
City level analyses (Basic analysis): September 2006
Sensitivity analysis: January 2007
Draft report: April 2007
Draft Final HEI report: June 2007
Co-ordinated analysis: September 2007

Delhi

Air quality and meteorological data collection – December 2005
Initial data audit - December 2005
Completion of data cleaning: January 2007
City level analyses (Basic analysis): October 2006
Sensitivity analysis: February 2007
Draft Final report: April 2007
Final HEI report: June 2007
Co-ordinated analysis: September 2007

Ludhiana
Initial Data Audit: February, 2007
Completion of data cleaning: March, 2007
City level analysis (Basic analysis): June, 2007
Sensitivity analysis: August, 2007
Draft Final Report: October, 2007
Final HEI Report: December, 2007

iii. Technical oversight

- The International Scientific Oversight Committee (ISOC) acting on behalf of HEI, will oversee the conduct of the coordinated analyses via a combination of regular progress reports, periodic site visits, conference calls, and participation in HEI Annual Conferences.
- The ISOC and HEI staff will be available to provide support and technical advice to the investigators as needed upon request. Once the analyses have been completed a final report will be published by HEI after review by the HEI Review Committee. The Review Committee will also prepare a Commentary on the report that will be published with it.

Annexure 1

Air quality monitoring techniques

Particulars	**Pollutants**			
	SPM	**RSPM**	**SO_2**	**NO_2**
Equipment	High Volume Sampler (HVS)	Respirable Dust Sampler (RDS)	Gaseous sampling unit attached to HVS/RDS	Gaseous sampling unit attached to HVS/RDS
Flow measuring device	Pressure drop across orifice in the hopper	Pressure drop across orifice in the hopper	Rotameters /Orifice Nanometers	Rotameters /Orifice Nanometers
Flow rate	0.8-1.3 m^3/min	0.8-1.3 m^3/min	1 Litre per minute	1 Litre per minute
Sampling period	8 hourly (round the clock)	8 hourly (round the clock)	4 hourly (round the clock)	4 hourly (round the clock)
Sampling frequency*	Twice a week	Twice a week	Twice a week	Twice a week
Analytical method	Gravimetric	Gravimetric	West & Gaeke	Jacobs & Hochheiser
Min. detection limit	1μg/m3	1μg/m3	0.04 μg/ml	0.03 μg/ml
Min. reporting value	10 μg/m3	10 μg/m3	6 μg/m3	3 μg/m3
Absorption wavelength (max)	NA	NA	560 nm	550 nm

Annexure 2

QA/QC procedures being adopted by CPCB in ambient air quality monitoring network

The CPCB collects data from the network of monitoring sites in Delhi and ensures quality through analytical procedures. In 1997, the CPCB established a calibration laboratory with the assistance from the German Technical Cooperation Agency. The laboratory was used to ensure uniformity in the analytical procedure and quality of data. The laboratory was also used to train State Board's officials for quality control and quality assurance.

The programme employed a "static injection system" which carried out a "ring test" for gaseous pollutants such as SO_2 and NO_2. In 1999, four rounds of inter-laboratory exercises were undertaken involving different State Pollution Control Boards (SPCBs)/ Pollution Control Committees (PCCs), including the CPCB.

CPCB initiatives to Ensure Data Quality

Recognizing the need to improve the quality of data, CPCB has initiated several measures in recent years. These are:

(i) Periodic inspection to the monitoring site and laboratory.

(ii) Organizing training programme on ambient air quality monitoring for field and laboratory staff of various monitoring agencies.

(iii) Analytical quality control exercises using Ring Test Facility are organized regularly every year at CPCB in which various monitoring agencies including TERI and NEERI have participated.

(iv) Review meetings are held at Zonal Offices of CPCB in which problems and deficiencies with the data are discussed with the monitoring agencies.

(v) Training programs on software for data management have been held for monitoring agencies.

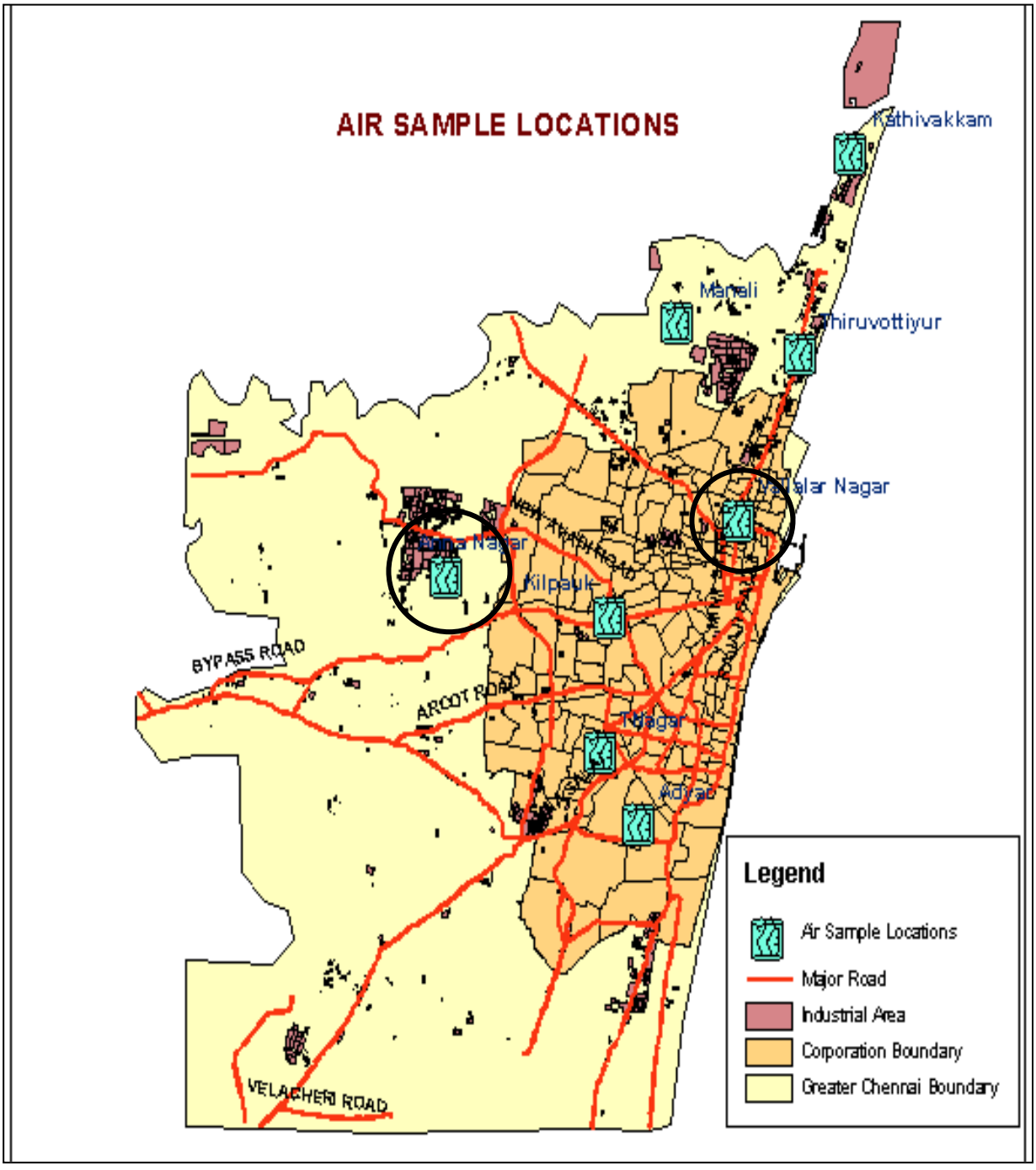
(vi) For the days with monitoring period less than 16 hours/day are not considered for calculating the daily average.

(vii) Systematic check for data entry for data validation.

Annexure 3

Basic monitoring information

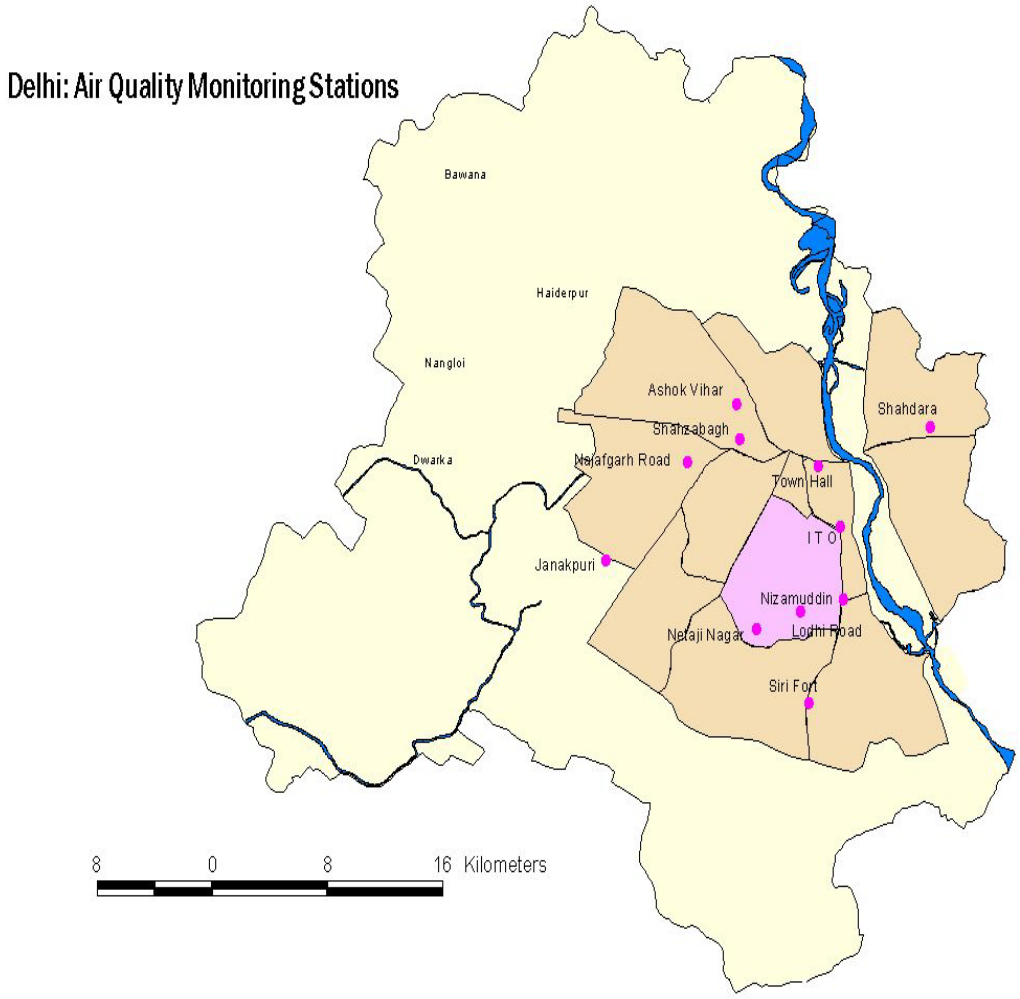
Chennai



	GPS	Height of the sampler	Site classification
Manali	13° 09.951 N, 080° 15.502 E	17 Feet	Industrial (Suburban)
Thiruvottiyur	13° 09.374 N, 080° 18.122 E	19 Feet	Industrial (Suburban)
Kathivakkam	13° 13.083 N, 080° 19.218 E	12 Feet	Industrial (Suburban)
T Nagar	13° 02.074 N, 080° 13.814 E	25 Feet	Commercial (City)
Vallalar nagar *	13° 06.345 N, 080° 16.809 E	10 Feet	Commercial (City)
Annanagar *	13° 05.319 N, 080° 10.515 E	35 Feet	Residential/ Vehicular (City)
Adayar	13° 00.744 N, 080° 14.665 E	25 Feet	Residential (City)

Manali, Thiruvottiyur and Kathivakkam are within an industrial hot spot outside the Chennai metropolitan area and hence were excluded from the core model. T Nagar and Adyar have had problems with frequent shifting of monitor location and hence are also not used in the core model. Vallalar Nagar and Annanagar (highlighted in circles) are used as the representative stations with the others being used only in the sensitivity analyses.

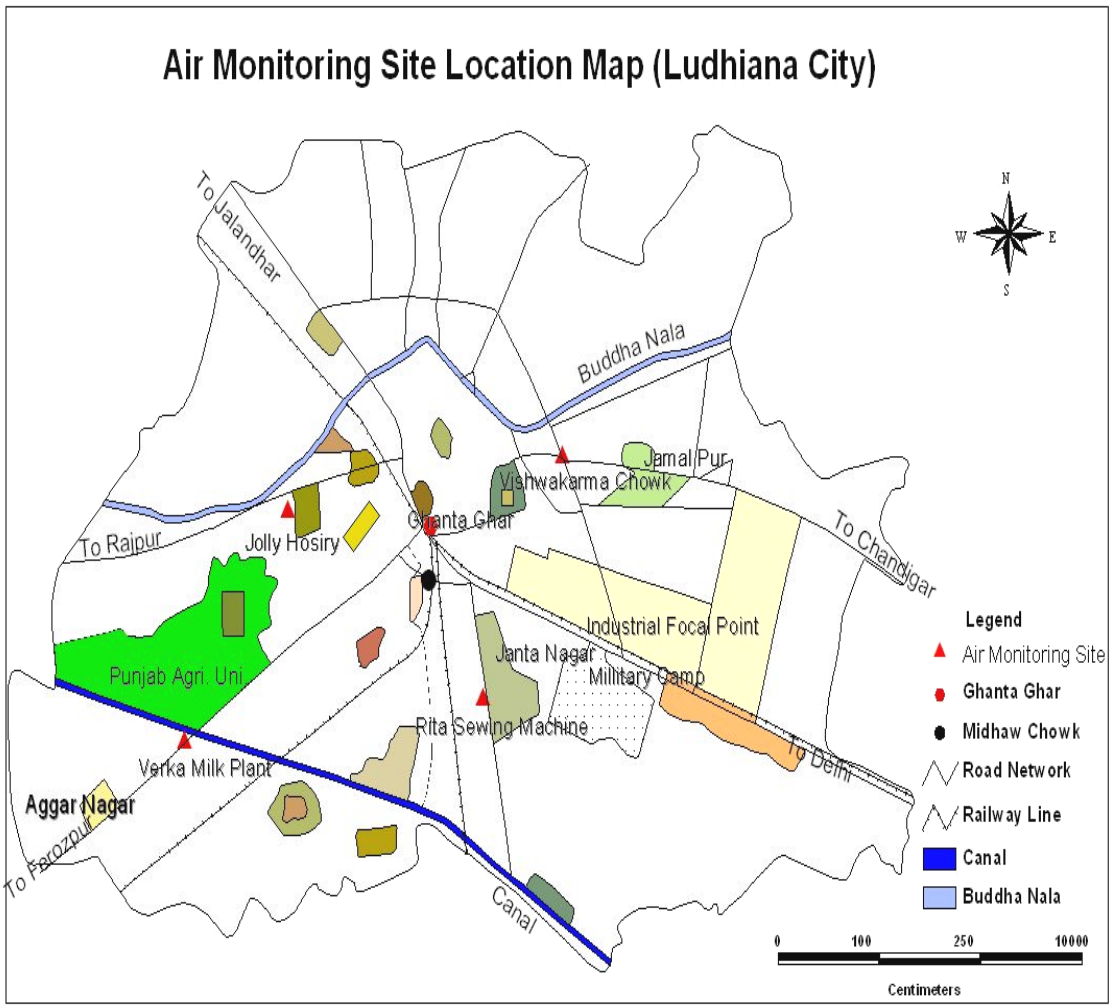
Air quality monitoring locations in Delhi



Details of air quality monitoring locations in Delhi

Site Name	Site Address	Latitude	Longitude	Site elevation (m, above MSL)	Inlet description, placement and height above ground (m)	Classification	Emissions information in vicinity of site (~2 km)	Available GIS data for the site/ region
Town Hall	Ayurvedic Dispensary building in the Town Hall	28^0 39.472 N	$077^0$13.614 E		6	Residential (Commercial)	Not Available	None
Sarojini Nagar	NDMC, Navyug School, Sarojini Nagar, New Delhi-23	28^0 34.481 N	$077^0$11.531 E	223	9	Residential	Not Available	None
Mayapuri		$28^0$37.648 N	$077^0$07.498 E	212	9	Industrial	Not Available	None
Ashok Vihar	NDPL, 11 KV Sub station, Near Lakshmi Bai College, Ashok Vihar – Phase III, Delhi	$28^0$41.071 N	$077^0$10.949 E	206	6	Residential	Not Available	None
Shahadra	BSES office, GT Road, Shahadra, Near Telephone Exchange	$28^0$40.584 N	$077^0$18.565 E	197	9.4	Industrial	Not Available	None
Janakpuri	BSES Rajdhani Power Ltd, C-Block, Janakpuri, Delhi	$28^0$37.110N	$077^0$04.614 E	205	9	Residential	Not Available	None
Siri Fort	BSES Rajdhani Sub Station, Near Shapur Jat, Siri Fort, New Delhi	$28^0$32.976 N	$077^0$12.728 E	215	5.5	Residential	Not Available	None
Nizamuddin	BSES Rajdhani, 33 KVA, DVB electric sub-station, Delhi –Mathura National Highway, Nizamuddin, New Delhi.	28^0 35.420 N	$077^0$13.720 E		15	Residential	Not Available	None
Shazada Bagh	NDPL 33 KVA electric sub-station, Shahzada Bagh, Delhi	$28^0$40.266 N	$077^0$10.432 E		12	Industrial	Not Available	None
ITO	Near Income Tax Office building	$28^0$37.707 N	$077^0$14.467 E	201	6.5	Traffic intersection	Not Available	None
IHC	TERI, IHC complex, Lodhi Road, New Delhi	$28^0$35.427 N	$077^0$13.543 E		3	Residential		None

Ludhiana



Monitoring Sites	GPS	Height of the sampler	Site classification
Vishawkarma Chowk (Site 1)	Latitude: 30^0 54' N Longitude: 75^0 48' E	14 Feet	Commercial
Jolly Hosiery (Site 2)	Latitude: 30^0 54' N Longitude: 75^0 48' E	18 Feet	Commercial
Verka Milk Plant (Site 3)	Latitude: 30^0 54' N Longitude: 75^0 48' E	11 Feet	Residential
Rita Sewing Machine (Site 4)	Latitude:30^0 54' N Longitude: 75^0 48' E	25 Feet	Industrial

Measurement Protocol Information

Vallalar Nagar	Vallalar Nagar	Vallalar Nagar
NO2	RSPM	SPM
CAAQM-Vallalar nagar	CAAQM-Vallalar nagar	CAAQM-Vallalar nagar
Top of First floor, Traffic intersection, Vallalar Nagar Bus Terminus, Vallalar nagar (Mint)	Top of First floor, Traffic intersection, Vallalar Nagar Bus Terminus, Vallalar nagar (Mint)	Top of First floor, Traffic intersection, Vallalar Nagar Bus Terminus, Vallalar nagar (Mint)
10 Feet	10 Feet	10 Feet
3E, 5.5W, 7.5N, 3.5S	3E, 5.5W, 7.5N, 3.5S	3E, 5.5W, 7.5N, 3.5S
Urban	Urban	Urban
13° 06.345 N, 080° 16.809 E	13° 06.345 N, 080° 16.809 E	13° 06.345 N, 080° 16.809 E
Traffic intersection	Traffic intersection	Traffic intersection
Vehicle emission	Vehicle emission	Vehicle emission
No	No	No
Yes	Yes	Yes
In the wist side Advertisement Board of 20 feet in height is hinderence	In the wist side Advertisement Board of 20 feet in height is hinderence	In the wist side Advertisement Board of 20 feet in height is hinderence
Indian standard method for measurement of NOx	Cyclonic flow technique PM10	Cyclonic flow technique PM10
Envirotech	Envirotech	Envirotech
Sodium hydroxide absorption solution	EPM 2000; 20.3*25.4 cm; Glass microfibre filters	Polythene cover
Ambient temperature	Ambient temperature	Ambient temperature
µg/m3	µg/m3	µg/m3
>=9	>=5	>=5
9	5	5
Not adjustable	Not adjustable	Not adjustable
No	No	No
NA	NA	NA
NA	NA	NA
3 samples /day; 8 hours/sample	3 samples /day; 8 hours/sample	3 samples /day; 8 hours/sample
8 hours	8 hours	8 hours
8hours each starting from 6.00am	8hours each starting from 6.00am	8hours each starting from 6.00am
24 hours	24 hours	24 hours
Noted accordingly	Noted accordingly	Noted accordingly
Noted accordingly	Noted accordingly	Noted accordingly
Instrument failure/Power failure/Not representative sample?	Instrument failure/Power failure/Not representative sample?	Instrument failure/Power failure/Not representative sample?

No	No	No
Yes	Yes	Yes
Yes	Yes	Yes
96ays/year	96ays/year	96ays/year
2 days	2 days	2 days
7-9 days	7-9 days	7-9 days
No	No	No
NA	NA	NA
No	No	No
Yes	Yes	Yes