Chapter 1

Introduction

1.1 Background

Air pollution in the outdoor ambient environment ranks among the leading risk factors contributing to both the global and regional burden of disease in South Asia (Lim et al., 2012; WHO, 2004). Over the last two decades, numerous toxicological, controlled human exposure, and epidemiological studies have examined the health effects associated with short- and long-term exposure to ambient air pollution (in the form of exposures to particulate matter (PM\textsubscript{10} and PM\textsubscript{2.5}) and gases such as nitrogen/sulfur oxides and ozone) and have served as the basis for pollutant specific air quality guidelines (AQGs) provided by the WHO (WHO, 2006).

Several systematic reviews provide an elegant consolidation of the evidence on adverse health effects of ambient air pollution for estimation of the attributable burden of disease. Health effects examined in relation to ambient air pollution have included cerebrovascular and ischemic heart disease (Katsouyanni et al., 1997; Samet, Dominici, Curriero, et al., 2000), COPD (H. R. Anderson et al., 2011), asthma (Schikowski et al., 2014), acute respiratory infections (Mehta et al., 2013), low birth-weight (Sapkota et al., 2012) and cancers (Loomis et al., 2013). In the most recent burden of disease assessment performed as part of the Global Burden of Disease (GBD) 2010 project (Lim et al., 2012), nearly 627,000 premature deaths and 17.8 million disability adjusted life-years (DALYs) were attributable annually in India to ambient air pollution (AAP), in the form of fine particulate matter.

Time-series studies of the effects of short-term exposure on morbidity and mortality
from cardiovascular or respiratory diseases have provided some of the most consistent evidence of serious adverse health effects of air pollution for regulatory policies in North America (Samet, Dominici, Curriero, et al., 2000; Samet, Dominici, Zeger, et al., 2000; Samet, Zeger, et al., 2000), Europe (Katsouyanni et al., 1997) and Asia (HEI, 2010). This has been supplemented by cohort studies such as the MESA (M. A. Cohen et al., 2009), Nurses Health (Puett et al., 2009) and ACS (Pope et al., 2002) studies in the US and more recently, the SAPALDIA (Zemp et al., 1999) and ESCAPE (Adam et al., 2015; Pedersen et al., 2013) studies in Europe, to estimate the effects of long-term exposure on annual average rates of mortality from chronic cardiovascular or respiratory diseases and impacts on life expectancy.

While the rapidly developing countries of Asia including India, bear bulk of the burden of ambient air pollution exposures, the representation of studies from high-exposure settings in the global ambient air pollution epidemiological literature has been modest (HEI, 2010). A comprehensive review of literature of Asian studies in 13 countries conducted by the Health Effects Institute, USA in 2010, identified some 43 studies (published between 1987 and 2007) reporting health effects in relation to ambient air pollution in India (HEI, 2010). The studies (mostly concentrated in the cities of Delhi and Mumbai) report results primarily from cross-sectional assessments of respiratory morbidity, estimating relative prevalence of respiratory symptoms and lung function in relation to inter-zonal differences in air quality within cities.

More recently, results from a coordinated set of time-series studies examining the association of natural all-cause mortality with PM$_{10}$ exposures in the cities of Chennai, Delhi and Ludhiana have been published (Balakrishnan et al., 2011; R. Kumar et al., 2010; Rajarathnam et al., 2011). The concentration-response functions developed through these studies suggest a generally similar risk of mortality associated with PM exposure (ranging from 0.15% to 0.4% increase in risk per 10-µg/m$^3$ increase in PM$_{10}$ concentrations) when compared with the multi-city studies conducted in China, South Korea, Japan, Europe, and North America. These initial studies in India (performed with data over a three year time period between 2002 and 2004) pointed out to the need for additional validation of the methods, using data over extended time periods as well as the need to develop model refinements to address unique features of exposure and health data sets available through relevant Governmental agencies in India.

Using a dataset on PM10, meteorological parameters and mortality extending over a 7 year period (between 2002-2008) the present thesis was conceptualized to both expand the initial results reported from Chennai as well as develop statistical methods to address limitations resulting from exposure misclassification and exposure missingness in semi-parametric models that examine such associations using time-series analyses. Such methods would afford the opportunity for application with other city data-sets in India, for the (much needed) expansion of the evidence base on short-term health effects of air pollution.

1.2 Rationale for study methods

The time-series approach investigates associations between short-term changes in pollution levels (such as PM\(_{10}\)) with short-term changes in acute health outcomes (mortality, respiratory hospital admissions due to cardiovascular and respiratory illness, emergency room visits asthma attacks and chronic disease), and it has been applied widely to investigate the health effects in developed countries on account of availability (and accessibility) of electronic data sets on air quality, mortality and morbidity. Although cohort studies provide a distinct advantage over time-series studies in establishing the strength of evidence for an observed association, the relative ease of execution and reliance on routinely collected data has made it attractive for application across geographically diverse settings (such as across multiple cities and/or countries).

Time-series studies however have only recently become feasible in India (Balakrishnan et al., 2011) with state pollution control boards and directorates of public health engaged in collection and archival of ambient air quality and health data. There is a need to standardize these methods further, to make this routinely feasible. Doing so would enable routine conduct of ecological studies directed at air pollution and health in India, till such time that feasibilities for cohort studies are more widely established.

While time-series studies have played a pivotal role in the setting of regulatory standards for air pollution (Greenbaum et al., 2001) in many developed countries, the nature and characteristics of time series data make risk estimation challenging, requiring complex statistical methods sufficiently sensitive to detect effects that can be small relative to the combined effect of other time-varying covariates such as weather and seasonal fluctuations in health outcomes due to epidemics (Schwartz et al., 1996; Katsouyanni et al., 1996;
Samet et al., 1997). In the last decade, many advances have been made in analytical approaches used in time series analyses for examining the association between air pollution and health. Standard regression methods used initially have been almost fully replaced by semi-parametric approaches (Speckman, 1988; Hastie & Tibshirani, 1990; Green & Silverman, 1994) such as Generalized linear models (GLM) with regression splines (McCullagh & Nelder, 1989), Generalized additive models (GAM) with non-parametric splines (Hastie & Tibshirani, 1990) and GAM with penalized splines (Marx & Eilers, 1998).

However, most methodological refinements in time series studies conducted in Europe and North America have been directed at addressing confounding bias from other time varying covariates. Limited efforts have been focused on addressing exposure misclassification arising from using data from a limited number of air pollution monitors that do not adequately capture population level gradients in exposures. While using data from single monitors or using simple city level averages from multiple monitors has been validated in developed country studies because of a high degree of correlation across monitors, this may not be very reliable in developing country cities. Urban zones in Indian cities can experience steep spatial gradients in air pollutant concentrations resulting from heterogeneous land-use profiles with consequent implications for exposures of population sub-groups. There was thus a need to develop models that used conventional city level averages from air pollution monitors together with alternative exposure series to arrive at an optimal model to explain the association with health effects.

Finally, with air quality monitors in India recording data on approximately 100 days per year the problems of serial co-relation in outcome measures (that are recorded every day) arising from missing exposure data can be significant. Advanced modeling approaches to overcome this limitation, such as the use of autoregressive Poisson regression models to address missing covariates have not been tried with air pollution and health data sets in India.

Based on the above rationale, the present thesis focused on developing semi-parametric methods that used a 7 year (2002-2008) data set on all-natural cause mortality, air quality and meteorological covariates in Chennai city to develop exposure-response relationships for PM$_{10}$ and cardio-respiratory mortality while duly addressing issues of inter-monitor variability, exposure data missing-ness and serial correlation. In addition, using a 3 year sub-set, exposure response relationships were developed for cardio-respiratory morbidity.
that included inpatient hospital admissions due to cardiovascular and respiratory illness. Simulation studies were also included to validate alternative models developed using these methods.

The primary outputs from the thesis were aimed at providing a reliable method for time-series analyses using routinely collected air pollution and health data in Indian cities. This would in turn, allow an expanded national base of information from such studies in the future, consolidating the evidence for air quality actions and policy in the country.