Incorporating Unequal Exposure to Environmental Hazards in Sample Design: Capturing children's proximity to polluting facilities

Alissa Cordner, Seth Spielman, Susan Short, Rachel Goldberg, John Logan, Stephen Buka

The impact of the environment on health is of central concern to population researchers, public health practitioners, policy makers and the public. The environmental justice literature demonstrates that racial-ethnic minorities and the poor are significantly more likely than whites or the affluent to live near environmental hazards (Arora & Cason 1999; Bryant and Mohai 1992; Bullard 1990; Downey 2006; Mohai, Pellow and Roberts 2009; Morello-Frosch, Pastor & Sadd 2002). Recent work argues that research should focus on the ways race and class interact to produce dynamic configurations of community characteristics and polluting facility characteristics that impact community levels of environmental risk in non-linear ways (Grant et al. 2010). These patterns of environmental inequalities are assumed to be connected to a variety of negative outcomes for residents, including stress, health problems, and decreased quality of life. However, to date the substantive *impacts* of unequal exposure are less fully documented than are the *patterns* of unequal exposure (Brulle and Pellow 2006, Downey 2006, Northridge et al. 2003). In short, while the literature in environmental justice makes a strong case for the need to connect exposure to environmental pollution and health outcomes, it is plagued by problems of endogeneity and selection bias, often due to the lack of suitable data.

The overlaps between race, affluence, and exposure to industrial activity, and the potential for interaction between them, complicate efforts to isolate the health effects of exposure to industrial activity. This paper details our development of a sampling method that attempts to bridge environmental (spatial) and population-based sampling techniques to ensure variability along both exposure to polluting activities and socio-demographic axes. Our sampling design relies upon a multi-dimensional measure of proximity to industrial activity that utilizes multiple sources of data. Our goal, in the context of local implementation of the of the National Children's Study, is to establish a sample of young children who experience variation in social and environmental contexts. Ultimately, the study will collect biological and environmental (indoor and outdoor) samples, as well as survey data, at regular intervals over a 20 year period. The design reported here will enhance efforts to identify links between environmental exposures and health in both the short and long term.

***Environmental Health and Justice* –**
Much of the quantitative environmental justice literature is plagued by methodological problems. While large national data sources report the location and quantities of certain chemical releases, small polluters are often omitted from these data sources. Some studies only rely on one measures of pollution from these large national databases, while others assume equality between facilities by using only the location of facilities and ignoring the risk they pose to the population. However, it is important to account for both the impact of large industrial facilities and the cumulative effects of the many small industrial facilities that are not captured in large national databases (e.g., Evans & Kantrowitz 2002, Morello-Frosch et al. 2011).

***Uniting Geographic- and Population-based Sampling* –**
The objectives of traditional population surveys are to identify and collect information from a representative sample of the target population. A sample is evaluated by its fidelity to the characteristics of a population of interest (Levy and Lemeshow 1999). However, a representative sample of a population is not necessarily a representative sample of the environments experienced by individuals. Unlike a population that consists of discrete individuals, the physical environment is a continuously varying landscape. At any point in a forest one can measure soil moisture or temperature. In collecting samples of the physical environment the goal is to accurately represent the spatial variation of some phenomenon of

interest (Haining 1990). Just as it is impractical to survey an entire population, it is often impossible to completely measure the environment. When surveying the environment sample points are often determined in advance, and data collection consists of visiting these predetermined locations (Delmelle 2011). Additionally, non-response is never a problem when working with dirt and air, but it is when working with people. These differences make the integration of spatial and survey sampling techniques difficult.

Within the context of probability-based sampling designs there has been limited progress integrating spatial and survey approaches to sampling (Lee et al. 2006). Occasionally, the two approaches to come together in interesting ways. Siri and colleagues (2008) in a study of urban malaria used a spatial sampling approach to select households. They selected sample locations as a set of coordinates and then used a Global Positioning System (GPS) to visit these locations and recruit into the study the household nearest the target location. Kumar (2007) also used a GPS to identify locations in a study of the health effects of air quality in Dehli. Kumar reports that in over 1500 of 2000 preselected sample locations households were successfully recruited into the study. However, one wonders if the non-responding households were located randomly in space or clustered in particular types of environments.

To study the impact of the environment on health, variation is required on both the social and environmental level. Moreover, we should expect that the interaction of person and place is associated with particular health outcomes (Grant et al. 2010, Morello-Frosch et al 2011). For example, children living in the same neighborhood can be expected to have different blood lead levels, depending in part on social characteristics. While all housing stock may contain lead paint, lead-safe practices may be adopted differently by families with different characteristics. Accordingly, it is preferable to investigate such interactions through a research design that includes for example not only the more typical high pollution/low SES and low pollution/high SES neighborhoods, but also the less common low pollution/low SES and high pollution/high SES neighborhoods.

Methods: Developing the Polluter Proximity Score –
In an influential 2002 article, Faber and Krieg develop a cumulative measure of exposure to numerous pollutants, this measure was constructed at the community-scale. This measure integrated multiple sources of data and assigned a weight to different types of facilities based upon the risk they posed to the population. To develop a measure of exposure to pollutants in Bristol County, we adopt a similar approach. However, Faber and Krieg's measure is sensitive to somewhat arbitrary municipal boundaries. Facilities, even those near the border of communities, can only be located in a single place. Thus, we describe the development and application of a spatial measure of proximity to polluting industrial facilities. This allows us to think of children's potential exposure to hazardous pollution as a function of their proximity to facilities whose hazard profiles and density can vary.

We apply the Faber and Krieg approach to sampling design by incorporating spatial data on six types of facilities: Superfund sites, solid waste facilities (broken down into incinerators or non-incinerators), Toxic Release Inventory (TRI) Reporting Facilities, MA DEP Major Facilities, and MA DEP Tier Classified 21E Oil or Hazardous Material Release Sites. These facility locations are displayed in Map 1. We make this measure spatially explicit, so that the impact of facilities is allowed to cross municipal borders.

We created a grid of 100m cells, for each cell we calculated a proximity score that measures the proximity of that cell to six different sources of industrial pollution. Sources of pollution were weighted following Faber and Krieg according to their risk (Table 1). The resulting map is shown in Map 2. We examine the overlap between this proximity score and SES measures, focusing on locations that do and do not conform to the expected relationship between SES and pollution. The finest resolution available for our socio-economic data was the census tract. The tract-level scores were divided into quartiles for thematic mapping purposes. This census tract breakdown is shown in Map 3. We examine how moving from 100m cells to census tracts changes the picture of pollution at different geographic levels.

Discussion: Applying the Polluter Proximity Score –
One of the challenges at the interface between survey and spatial sampling is that it is difficult, through the selection of coordinates, to ensure the maintenance of a probability sample where probability of selecting a particular type of household is controlled through research design. Lee (2006) proposes highly detailed electronic maps of residential parcels for sampling. While this approach is promising, detailed spatial socio-economic data is seldom available at the household level. We use the tract-level proximity scores and measures of completed education, race, homeownership, poverty, nativity, and language to create strata that varied across socio-economic and environmental axes. Out of 14 total strata for Bristol County, three were specifically chosen based on the polluter proximity measure: high proximity to polluters, high SES; low proximity to polluters, low SES; and high proximity to polluters, low SES. To ensure that there was an equal probability of selection across strata, these strata were sub-divided into geologically homogenous sampling units called segments. One segment is to be randomly selected from each stratum by a third party.

Other National Children's Study sites have developed measures of environmental hazard or social and environmental stressors (e.g., Downs et al., 2010). Our measure is unique in its multi-dimensional measure of polluter proximity, and its integration of spatial approaches.

This case study makes two important methodological contributions to survey research on health and environment: it integrates two disciplinary models of sampling, and it uses a more robust measure of environmental exposure than previous similar study designs. In place-based longitudinal studies, the impact of the environment on health is most measurable when looking at people who move out of the area, but this is expensive, and 'movers' differ from 'stayers' in important ways that make it difficult to tease out confounders. Incorporating environmental variability into the study design is a way for researchers to make progress on this endogeneity problem.

A sampling frame that captures variation in both environmental exposures and SES also helps researchers move beyond the assumption that poor people always live in the most polluted neighborhoods and rich people always live in the least polluted neighborhoods. Although polluter proximity does correlate in a largely linear fashion with SES and race (Krieg and Faber 2004), this assumption obscures more complex relationships between SES and environmental hazards (Grant et al. 2010). Stratifying on both environmental and population characteristics potentially allows researchers to investigate the independent contributions of environmental hazards and SES status. Incorporating these measures into sampling design will also allow researchers to examine more closely the association between proximity to polluting facilities and exposure to harmful chemicals.

The final paper will expand on these themes and provide greater detail on the methodology employed.

References:

Arora, Seema and Timothy Cason. 1999. "Do Community Characteristics Influence Environmental Outcomes? Evidence from the Toxics Release Inventory." *Southern Economic Journal*. 65:691-716.

Robert Brulle and David Pellow. 2006. "Environmental Justice: Human health and environmental inequalities." *Annual Review of Public Health*. 27: 103-24.

Bryant, Bunyan, and Paul Mohai. 1992. *Race and the incidence of environmental hazards: A time for discourse*. Boulder: Westview Press.

Bullard, Robert. 1990. *Dumping in Dixie: Race, Class, and Environmental Quality*. Boulder: Westview Press.

Downey, Liam. 2006. "Environmental Racial Inequality in Detroit." *Social Forces*. 85: 771-796.

Downs, Timothy J., Y. Ogneva-Himmelberger, O. Aupont, Y. Wang, A. Raj, P. Zimmerman, R. Goble, O. Taylor, L. Churchill, C. Lemay, T. McLaughlin, and Marianne Felice. 2010. "Vulnerability-Based Spatial Sampling Stratification for the National Children's Study, Worcester County, Massachusetts: Capturing Health Relevant Environmental and Sociodemographic Variability." *Environmental Health Perspectives*. 118(19):1318-1325.

Evans, Gary, and Elyse Kantrowitz. 2002. "Socioeconomic status and health: The potential role of environmental risk exposure." *Annual review of public health*. 23:303-331.

Faber, Daniel, and Eric Krieg. 2002. "Unequal exposure to ecological hazards: Environmental injustices in the Commonwealth of Massachusetts." *Environmental Health Perspectives*. 110(2):277-288.

Grant, Don, Mary Nell Trautner, Liam Downey and Lisa Thiebaud. 2010. "Bringing the polluters back in: Environmental Inequality and the Organization of chemical production." *American Sociological Review*. 75 (4):479-504.

Haining, R. 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press. 410pp.

Krieg, Eric, and Daniel Faber. 2004. "Not so black and white: environmental justice and cumulative impact assessments." *Environmental Impact Assessment Review*. 24: 667-694.

Kumar, N. 2007. "Spatial Sampling for a Demographic and Health Survey." *Population Research and Policy Review*. 26(5-6): 581-99.

Lee C, Moudon A, and Courbois J. 2006. "Built Environment and Behavior: Spatial sampling using parcel data." *Annals of Epidemiology*. 16(5): 387-394.

Levy, P. S. and Lemeshow, S. 1999. *Sampling of Populations: Methods and Applications*. New York: Wiley-Interscience.

Mohai, Paul, David Pellow, and J. Timmons Roberts. 2009. Environmental Justice. *Annual Review of Environment and Resources*. Vol. 34.

Montaquila, Jill, J. Michael Bricka, and Lester R. Curtin. 2010. "Statistical and practical issues in the design of a national probability sample of births for the Vanguard Study of the National Children's Study." *Statistics in Medicine*. 29:1368-1376.

Morello-Frosch, Rachel, Miriam Zuk, Michael Jerrett, Bhavna Shamasunder and Amy Kyle. 2011. "Understanding the cumulative impacts of inequalities in environmental health: implications for policy." *Health Affairs*. 30(5): 879-887.

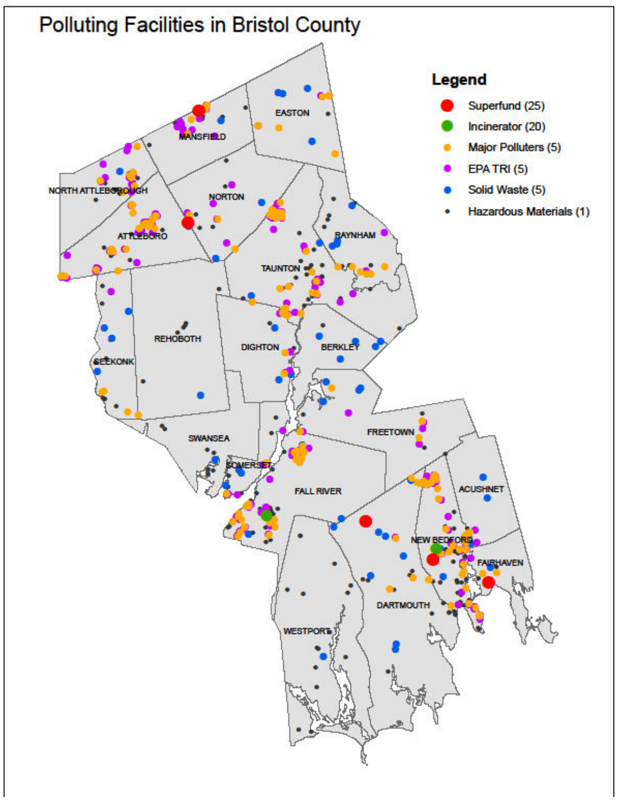
Morello-Frosch, R., M. Pastor, James Sadd. 2002. "Integrating Environmental Justice and the Precautionary Principle in Research and Policy Making: The Case of Ambient Air Toxics Exposures and Health Risks among Schoolchildren in Los Angeles." *The Annals of the American Academy of Political and Social Science*. 584: 47-68.

Northridge, Mary, Gabriel Stover, Joyce Rosenthal, and Donna Sherard. 2003. "Environmental equity and health: Understanding complexity and moving forward. *American Journal of Public Health*. 93(2): 209-215.

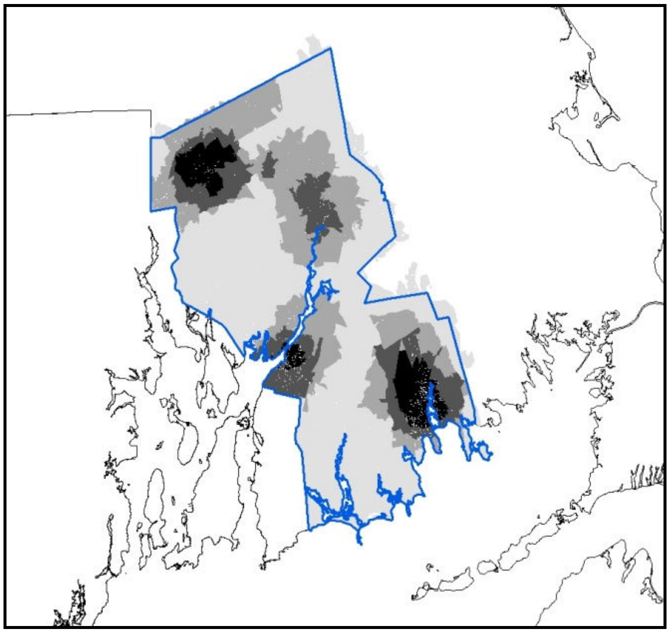
Siri, J. and Lindblade, K. and Rosen, D. and Onyango, B. and Vulule, J. and Slutsker, L. and Wilson, M. 2008. "A census-weighted, spatially-stratified household sampling strategy for urban malaria epidemiology." *Malaria journal*. 7(1).

Table 1: Type of hazardous facility or site	Points
Superfund	25
Solid waste – incinerator	20
Solid waste – not incinerator	5
Toxic Release Inventory (TRI) Reporting Facilities	5
MA DEP Major Facility	5
MA DEP Tier Classified 21E Oil or Hazardous Material Release Sites	1

Map 1: Location of polluting facilities, Bristol Co, Ma



Map 2: Density gradient of polluter proximity index in quartiles, Bristol Co, MA



Map 3: Polluter proximity by census tracts, Bristol Co., MA