Available online at www.sciencedirect.com
ScienceDirect
Environmental Research 102 (2006) 172–180
Environmental Research
www.elsevier.com/locate/envres

Levels of analysis for the study of environmental health disparities ☆

M. Soobader[a,*], C. Cubbin[b], G.C. Gee[c], A. Rosenbaum[d], J. Laurenson[d]

[a]STATWORKS, 800 Matthew Court, Suite 102, Braintree, MA 02184, USA
[b]Department of Family and Community Medicine, Center on Social Disparities in Health, University of California, San Francisco, CA, USA
[c]Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI, USA
[d]ICF International, Fairfax, VA, USA

Received 27 October 2005; received in revised form 20 April 2006; accepted 2 May 2006
Available online 15 June 2006

Abstract

Reducing racial/ethnic and socioeconomic environmental health disparities requires a comprehensive multilevel conceptual and quantitative approach that recognizes the various levels through which environmental health disparities are produced and perpetuated. We propose a conceptual framework that incorporates the micro level, contained within the local level, which in turn is contained within the macro level. We discuss the utility of multilevel techniques to examine environmental level (both physical and social) and individual-level factors to appropriately quantify and improve our understanding of environmental health disparities. We discuss the reasoning and the methodological approach behind multilevel modeling, including differentiating between individual and contextual influences on individual outcomes. Next we address the questions and principles that guide the choice of levels or geographic units in multilevel studies. Finally, we address the ways in which different data sources can be combined to produce suitable data for multilevel analyses. We provide some examples of how such data sources can be linked to create multilevel data structures, and offer suggestions to facilitate the integration of multilevel techniques in environmental health disparities research and monitoring.

Keywords: Environmental justice; Indicators; Social disparities; Race/ethnicity

1. Introduction

Executive order 12898 on Environmental Justice states that "environmental human health analyses, whenever practicable and appropriate, shall identify multiple and cumulative exposure" (White House Office of President, 1994). This core theme recognizes that an individual's health is a complex function of his/her own characteristics, factors in his/her environmental context, and interactions between the individual and his/her context. However, recent research increasingly places primary emphasis on investigating the micro and individual levels or "downstream" factors, often to the exclusion of more macro or "upstream" factors. And while the micro level is indeed important, a major limitation of focusing only on micro level processes is that the environmental context itself is removed from the line of inquiry. This has led some to argue that environmental scientists and epidemiologists have become "prisoners of the proximate" (McMichael, 1999). Inattention to the complex interactions between individuals and their environments may lead to inappropriate science, and thus incomplete and perhaps misguided interventions and policies. This trend toward a micro level research approach has been noted as a barrier towards understanding racial/ethnic disparities in health because minority populations tend to live in far more hazardous physical and social environments and have lower levels of socioeconomic position (SEP) than does the majority population (Evans and Kantrowitz, 2002; Williams et al., 2001).

Rather, reducing racial/ethnic and socioeconomic environmental health disparities requires a comprehensive multilevel research approach that recognizes the various

☆This paper was written for the "Environmental Health Disparities Workshop: Connecting Social and Environmental Factors to Measure and Track Environmental Health Disparities," under contract to EPA, Contract No. EP-W-04—049, Task Order No. 11. *Note*: No human subjects or experimental animals were used in this study.

*Corresponding author.
E-mail address: mah-j@statworks.com (M. Soobader).

0013-9351/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.envres.2006.05.001

levels through which environmental health disparities are produced and perpetuated. In endorsement of this approach, gene-environment interactions and cumulative risk assessments have also renewed interest in a more complex, multi-factorial and multilevel understanding of health disparities (US Enviromental Protection Agency, 2003). Gene-environment research suggests that genetic vulnerabilities may be expressed or amplified when certain environmental triggers occur. Cumulative risk assessment examines how multiple exposures from multiple sources over time contribute to health. Recent conceptual environmental frameworks further support this approach. For example, Schulz and Northridge (2004) suggest that fundamental factors, such as geographic topology and wealth distribution, shape intermediate contexts, such as land use and local economies, which in turn shape proximate risks, such as housing quality and/or unfair treatment, ultimately shaping individual health (Schulz and Northridge, 2004). Similarly, Morello-Frosh et al. (2002) suggest that income inequality and social capital influence the ability of local communities to affect environmental and social policy actions, thereby influencing these communities' abilities to resist environmental health stressors (such as the location of hazardous waste facilities) and subsequent health effects. Gee and Payne-Sturges (2004) suggest that macro level residential segregation leads to differential local level environmental hazards and social stressors, which in turn lead to differential individual level stressors and subsequent illness and health disparities.

Fig. 1 presents a generalized conceptual framework to inform the present discussion. In this framework, the primary causal pathway to a particular health outcome is from the macro to the local to the micro level with the secondary pathway as a feedback loop. Using lead poisoning as an example, the *micro level* refers to factors (e.g., demographic, behavioral) pertaining to the individual, such as age or pica. The *local level* refers to the immediate context that surrounds the individual, such as the concentrations of lead in the surrounding soil or in the home. The *macro level* refers to both the larger geospatial region (e.g., states, counties) that encapsulates the local level and/or the broader social context (e.g., political climate and laws/enforcement regarding lead-based paint in housing or lead in gasoline).

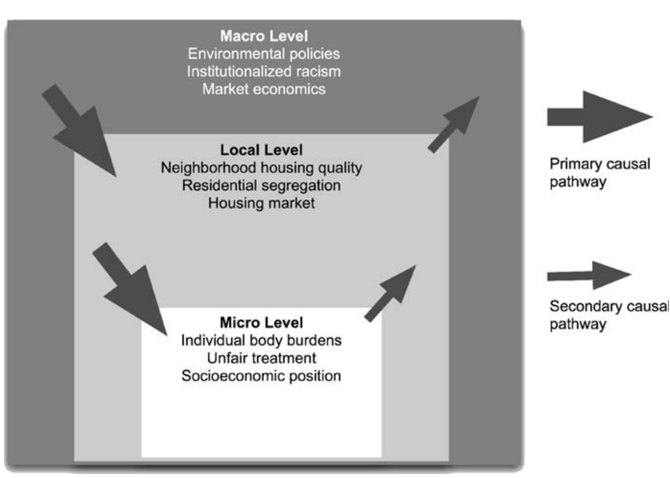

Fig. 1. Multilevel conceptual framework.

Such a conceptual framework approach underlies multilevel techniques that allow for the consideration of numerous levels *simultaneously*; that is, factors that affect health are simultaneously considered as operating at the level of the individual and the level of contexts (Subramanian et al., 2003). A simple two-level study of individuals (micro level) within neighborhoods (local level) would allow us to examine whether observed environmental health disparities were due to characteristics of individuals and/or characteristics of the neighborhood (context), as well as whether the factors at the micro level interact with the factors at the macro level.

Taking lead as our example, one could ask whether childhood lead poisoning is associated with pica after accounting for age, gender and nutritional status of the child. Such an analysis would be conducted at a single level: the individual (micro) level. However, one could expand the research question to ask whether lead in the water system (local level factor) has an independent association with childhood lead poisoning after controlling for the child's (micro level) pica, age, gender and nutritional status. The presence of a house-level association may indicate that an exposure route that does not involve pica behavior but is directly related to lead in the house's water system. The introduction of the *household* into the analysis makes this analysis multilevel. This analysis could be taken a step further by asking whether the effects at the individual and household level might act synergistically (a cross-level interaction). For example, pica may have no association with lead poisoning among children living in homes built within the last 10 years, but pica may be significantly associated with lead poisoning for children living in homes older than 25 years (i.e., before 1978). Thus, there might be a cross-level interaction between the individual (age) and her context (household). This two-level analysis could be extended to a three-level analysis, for example, by considering whether areas zoned for mixed residential-commercial use (macro level) have higher rates of childhood lead poisoning than areas zoned purely for residential use, after controlling for age of home (local level) and pica (micro level), and the cross-level interaction between pica and home.

Multilevel analysis can also aid the investigation of racial/ethnic health disparities. For example, a single individual level study might initially show that African American children have higher rates of lead poisoning than White children even after controlling for pica, age, gender and nutritional status. However, a multilevel analysis might find that these racial disparities disappear after accounting for land use characteristics. This suggests that disparities may result from or may be mediated by neighborhood organization rather than race *per se*. Thus, multilevel approaches can improve our understanding of the complex relationships between individual and contextual influences on health (Macintyre et al., 2002).

This paper presents fundamental conceptual and analytic approaches for promoting health and sustainability by using multilevel techniques to quantify and monitor socioeconomic and racial/ethnic disparities in environmental health. Our approach will be more heuristic than mathematical in order to reach a broader audience. More technical discussions can be found in numerous sources (Bingenheimer and Raudenbush, 2004; Bryk and Raudenbush, 1992; De Leeuw and Kreft, 2001; Goldstein, 1995; Hox and Kreft, 1994; Singer, 1998).

We discuss the utility of multilevel techniques to examine environmental- (both physical and social) and individual-level factors to appropriately quantify and improve our understanding of environmental health disparities, highlighting key multilevel research questions. Next we address the questions and principles that guide the choice of levels or geographic units in multilevel studies. These include the research question, the health outcome being considered, the definition of the level, and the exposure period. Finally, we address the ways in which different data sources can be combined to produce suitable data for multilevel analyses. We provide some examples of how such data sources can be linked to create multilevel data structures, and offer suggestions to facilitate the integration of multilevel techniques in environmental health disparities research and monitoring.

We conclude with a summary of the types of research questions on environmental health disparities that could be answered with multilevel modeling approaches, and with suggestions to facilitate the integration of multilevel techniques in environmental health disparities research and monitoring.

2. Multilevel models

2.1. Why multilevel models?

Multilevel models are an extension of ordinary multiple regression that explore individual and contextual parts of variation in exposure. Multilevel models provide a mathematical modeling approach to examine between-place and between-people variability (Duncan and Jones, 2000). These variations are modeled by recognizing that individuals within groups, groups within local contexts, and local contexts within macro contexts may share similar characteristics. Therefore, multilevel techniques explicitly model correlated data where the assumption of independence between observations is violated and conventional OLS techniques are not appropriate. This is in contrast to procedures that attempt to correct for the correlated structure of the data [e.g., generalized estimating equations (GEE)]. Commonly used multilevel statistical packages include MLwiN, HLM, and Mixed procedures within SAS. Other multilevel software include such products as aML, EGRET, GENSTAT, LIMDEP, LISREL, MIXREG, Mplus, R, S-Plus, SPSS, STATA, SYSTAT, and WINBUGS.

2.2. Multilevel research questions?

Modeling the correlated data structure is a salient point for environmental research that advocates the role of shared environments in contributing to exposure and health differentials between populations. Within this context multilevel research can provide a conceptual modeling approach for examining:

- A single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Exposures differentially affecting subgroups of the population and/or geographic areas, and/or producing synergistic outcomes.
- The fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.

Expanding on the case of lead exposure, below are examples of research questions that describe how the multilevel modeling approach can potentially contribute to environmental health disparities research:

- What levels are important for the study of lead exposure, and what is the relative importance of the different levels?
- Are there contextual differences in individual lead exposure levels, after taking into account individual characteristics?
- What is the contribution of neighborhood and school levels, that may not be nested within one another but overlap, to lead exposure?
- What is the average association between individual lead exposure and neighborhood quality, and does this association differ for different individuals based on their poverty status, after accounting for individual characteristics and the neighborhoods in which the individuals live?
- What is the average association between resource allocation for lead abatement programs at the state level and neighborhood quality in relation to lead exposure, and does this association differ for neighborhoods based on their quality profile, after accounting for individual characteristics and the characteristics of the neighborhoods in which the individuals live?
- While individual lead exposures may have declined over time, have neighborhood contextual disparities declined or increased, and for which population groups have the contextual disparities declined or increased?

- Which types of individuals and which types of places have changed over time with respect to lead exposure?

Such research questions can provide a comprehensive approach to monitoring environmental health disparities. These research questions are dealt with in more depth in Appendix A (available online at http://yosemite.epa.gov/ochp/ochpweb.nsf/content/Tech_2apps.htm/$File/Tech_2apps.pdf).

2.3. Identifying the relevant levels

This section addresses the questions and principles that guide the choice of levels or geographic units in multilevel studies. These include the research question being addressed; the theoretical pathways linking the micro, local, and macro levels; health outcomes and exposures under consideration; data availability; and the administrative or intervention application of the research. Social theories underlying choice of levels are discussed in detail in Appendix B (available online at http://yosemite.epa.gov/ochp/ochpweb.nsf/content/Tech_2apps.htm/$File/Tech_2apps.pdf). Quantifying the importance of different levels using a multilevel approach can provide insights into theoretical pathways linking micro–local–macro processes. Additional considerations in defining a relevant level of analysis include extant policies (e.g., the partitioning of EPA Regions), time sensitivity of the measures under consideration, and the mechanisms that may mediate the relationships between levels (Diez Roux, 2001).

Although we use conventions such as "individuals" to refer to the micro level, a multilevel approach is not necessarily restricted to the micro level defined as individuals or persons; the micro level can refer to industrial facilities or other theoretically defined units. The classic example of environmental justice examines the placement of industrial facilities in communities, with the central argument that race is a key determinant of facility location (Perlin et al., 2001). If we consider facilities as the micro level, we can consider whether racial composition at the local level influences the placement of facilities, net of other economic characteristics. Further, we could consider the influence of municipal or state policies as determinants of local racial composition and facility placement.

We briefly discuss census geography, as it is the most commonly used source to identify contextual levels. We will then discuss some issues to help guide the choice of levels.

2.4. Level specification

Table 1 provides examples of levels that have been used or could be considered in multilevel studies (Brooks-Gunn et al., 1997; Diez Roux, 2004). The most commonly used contextual levels for multilevel analyses are based on census geography which is built hierarchically; e.g., from the block level representing 85 individuals, up to the block group with an average population of 1500, to census tracts that represents about 4000 people, to county subdivisions, to counties and then states. (A more detailed description of census geographic units is provided in Appendix C, available online at http://yosemite.epa.gov/ochp/ochpweb.nsf/content/Tech_2apps.htm/$File/Tech_2apps.pdf.) Block groups and census tracts have been used to approximate neighborhoods, as their boundaries are defined to create population groups that are homogenous with regard to social and economic characteristics (Brooks-Gunn et al., 1997; Diez Roux, 2004). Although Zip codes have been used in health research, their use is not recommended. While census tracts and block groups are delineated to be homogenous units, Zip codes are defined by the US Postal Service (USPS) for efficient mail delivery and can range in size from a single building to large areas that cross state boundaries. The use of Zip codes is further complicated by extensive modifications in the past 10 years, the creation of new Zip codes, and the deletion of existing Zip codes. To overcome these discrepancies, the 2000 US Census defined Zip Code Tabulation Areas (ZCTA) mapped to census blocks to replace Zip codes. Although the efforts to create ZCTAs are welcome, there are major concerns with the use of ZCTAs in research. In particular, the current 5-digit ZCTA area may no longer correspond to the USPS 5-digit Zip code area and no relational files between Zip codes and ZCTAs are planned to be released by the census (Krieger et al., 2002).

Table 1
Examples of units at different levels

Macro levels	Local levels	Micro levels
Countries	Change in place (time)	Individual growth (time)
EPA regions	Communities	Cellular matrices
States	Neighborhoods	Blood markers
MSAs	Work places	DNA
Counties	Zoning districts	Observations
Cities	Voting districts	Individuals
State economic areas	Congressional districts	Families
Labor market areas	School districts	Homes
State legislative districts	Traffic analysis zones	Schools
	Census tracts	Industrial facilities

The Traffic Analysis Zone (TAZ), the primary unit to describe traffic data, may be a useful level for the study of environmental health disparities. A TAZ consists of one or more census blocks, block groups, or census tracts, and is defined by states and/or local transportation officials. Macro level units such as states and counties may be important to consider in policy evaluation efforts (US Department of Commerce, 1994).

Zoning districts may also be an important level for the study of environmental health disparities, as zoning is a commonly used planning tool that can influence the location of environmental hazards. Areas zoned for industrial use have been shown to be concentrated in poor and minority neighborhoods, resulting in higher environmental burdens for these neighborhoods. Zoning districts and their relationship to neighborhoods and/or census tracts should be considered important levels in environmental health disparities research (Maantay, 2001, 2002).

2.4.1. Populations not defined by conventional levels

Conventional geographic units (e.g., census tracts) are likely inadequate indicators of the "contexts" for homeless and migrant populations (e.g., migrant workers). However, some creativity in operationalizing these conventional units can be informative; e.g., Culhane et al. (1996) present an interesting methodology using census tracts for studying homeless populations. Also, most of the research using census data focuses on metropolitan areas, leaving relatively unexplored the applicability of these indicators for use with rural populations. Further, emerging immigrant enclaves may form and change between decennial census periods, leading to possible mismeasurement of sociodemographic characteristics of these areas. This problem may potentially be addressed by The American Community Survey, which will provide comparable data to that available on the US Census on an annual basis for all states, cities, counties, and metropolitan areas, and every 5 years for smaller geographic units.

2.5. Issues to be considered in level specification

2.5.1. The research question

Conceptually, the research agenda needs to be broadened to include the macro, local and micro level processes through which environmental health disparities are structured. It is well recognized that disadvantage is locally concentrated. The questions to address are: What are the higher-level precursor processes that result in the observed geographic clustering of economic disadvantage, social disadvantage, vulnerability, and environmental exposures (Brooks-Gunn et al., 1997), and how do micro level characteristics of individuals tie in with these processes? Such inquiries provide a comprehensive approach to studying environmental health disparities and are more likely to ensure that all relevant levels are considered.

2.5.2. Health outcome

The health indicator(s) under study influences the level(s) at which the analyses are done. The causal links between health and physical and social environments are typically complex and multifactorial, requiring multiple levels of inquiry. Certain health indicators may be more variable at the micro or local levels, while others may be more dependent on the macro level. Dependence of the health indicator may also vary by race/ethnicity. For example, census tract level variability in mortality was found to be six times greater for Blacks than for Whites (Subramanian et al., 2005).

2.5.3. Social definitions vs. administrative definitions vs. intervention levels

Neighborhoods defined by census tracts or other administrative units have several advantages, including convenience, replicability, good documentation, and availability for public-use. However, residents' perceptions of a neighborhood may not coincide with census tracts and other administratively-defined neighborhoods. Thus, administratively-defined neighborhoods may fail to capture or misclassify important aspects of the social, physical and built environments. Further, racial/ethnic and socioeconomic groups may vary in their perceptions of neighborhood boundaries (Williams et al., 2001). We need to also consider boundary definitions for interventions. For example, the state may be an important level for health outcomes that are influenced by state policies, while neighborhoods may be an important level for health outcomes that are influenced by city policies.

2.5.4. Modifiable area unit problem (MAUP)

With the increase in multilevel studies, the issue of the modifiable area unit problem (MAUP) has received much attention. This relates to the definition of the level(s) (area unit) into which individuals or areas are aggregated. This is an important consideration since changes in the size and number of levels (areas) can influence the observed associations significantly, reversing the direction of the observed relationships in the extreme case (Armhein, 1994; Heywood, 1998).

2.5.5. Exposure period

While the choice of levels is an important consideration in multilevel models, the effect of the social and physical environment on individuals is dependent on the exposure period within the local environment. The underlying assumption in multilevel models is that individuals in a given area share common exposures and experiences. Presumably, the local environment is a less relevant level for individuals whose work, school, or social activities are outside of the local area. Although population exposure modeling using population activity data can account for variability in exposure periods, the spatial pattern of daily life in relation to the definition of the local environment remains an area for further research (Sastry et al., 2002). In addition, some research suggests that ethnic groups may differ in terms of their geographic mobility and variance in health (Massey, 2001; Subramanian, et al., 2003).

2.5.6. Environmental exposures

Characteristics, measurement, multiple sources, relevant policies, and the research question for environmental exposure influence the choice of analytic levels. Outdoor air pollutants and ambient and drinking water quality are used as examples to provide an in-depth description of how these issues influence the choice of levels (see Appendices D and E). For example, estimation procedures directly influence the level at which data are available, with air quality monitoring data available at more aggregate levels such as counties and dispersion modeled data potentially available at the level of individual addresses. In the case of water quality, using water monitoring or modeling data with demographic data at the census tract level can be difficult. For example, in a study of problems associated with collecting drinking water quality data for community studies, the task of evaluating water quality for each census tract was complicated by the fact that single census tracts were served by more than one system (Whorton et al., 1988). This is a classic problem that multilevel modeling can potentially address to identify the relative importance of two contexts at the same level that are not contained within one another but overlap (see "Multiple Contexts at the Same Level" section in Appendix A).

3. Data linkages for multilevel models

Data requirements for multilevel models require that, at a minimum, observations have identifiers that differentiate the level of each observation. This section describes the ways in which different data sources can be combined to produce suitable data for multilevel analyses. Census data are the most commonly used source for characterizing and defining contexts, such as neighborhoods based on census tracts or block groups; therefore many national surveys are now being routinely geocoded, allowing linkage to census data to facilitate multilevel analyses. In addition, because of how data are collected (e.g., environmental exposures) or because of the sampling strategy of national surveys, geographic identifiers are available in some datasets. Table 2 shows examples of health/monitoring systems and their linkages with data sources at other levels.

3.1. Geocoding

Geocoding individual records is an important step towards creating multilevel data sources. Geocoding refers to the process by which individual addresses are assigned a corresponding set of latitude and longitude coordinates. These coordinates can then be coded to any geographic unit by determining within which geographic unit the specific coordinates are based. For example, these coordinates can be mapped to their respective census blocks, census tracts, and EPA areas.

Because geocoding requires a physical address, this approach is especially problematic for mobile populations (e.g., homeless persons, migrant workers) and requires further methodological development to appropriately contextualize these populations.

The issue of geocode accuracy is not adequately addressed in most datasets or studies. Very large discrepancies in accuracy have been noted between commercial firms, so researchers are encouraged to report and evaluate the accuracy of the geocoding methods (Krieger et al., 2001), as has been done in some studies (Cubbin and Winkleby, 2005; Pollack et al., 2005; Winkleby et al., in press).

Contextual identifiers can pose a serious confidentiality problem. Because block groups are relatively small areas—when rare health outcomes are considered and cross tabulated with other covariates—the potential to identify the respondent within the block group is high. Therefore, when the decision is made to geocode, restrictions need to be in place to protect the confidentiality of individuals. General approaches to reduce confidentiality breaches include anonomized identifications, cell suppression, aggregation, top coding, and reducing the detail for categorical variables (Cox, 1996). To facilitate the analysis of contextual identifier-linked national health surveys, the

Table 2
Sample data sources with geographic identifiers

Data	Indicators	Available levels
Environmental exposures		
Aerometric Information Retrieval System (AIRS)	Common air pollutants	County, MSA, state
National Air Toxics Assessment (NATA)	Hazardous air pollutants	Census tract, County, MSA, state
Toxic Release Inventory data (TRI)	Toxic chemical releases from industrial facilities	Individual facilities, county, state
Safe Drinking Water Information System (SDWIS)	(a) Drinking water contaminants (b) Violations of monitoring	Water system, county, MSA, state
Superfund NPL Assessment Program (SNAP)	Residence in relation to site	Site locations, county, MSA, state
Body burdens		
National Health and Nutrition Examination Survey (NHANES)	(a) Mercury in women of childbearing age (b) Lead and cotinine in children's blood	Block group, tract, county, MSA, state
Health indicators		
National Health Interview Survey (NHIS)	(a) Asthma prevalence (b) Preexisting health conditions (c) Asthma emergency room visits	Block group, tract, county, MSA, state
Surveillance, Epidemiology End Results Program (SEERS)	Cancer by type	Block group, tract, (CA only) Specific states and MSAs

National Center of Health Statistics has created Research Data Center (http://www.cdc.gov/nchs/r&d/rdc.htm).

3.2. Examples of data linkages

This section provides some examples of how different data sources can be linked to create multilevel data structures.

3.2.1. Contextual variables linked to geocoded data

The geocoding of individual records allows for these records to be linked with contextual variables. Contextual variables from census data may be linked with individual records (e.g., of the NHIS and NHANES) at the tract, county, city, MSA, and state levels. Contextual variables from The Area Resource File on health services, economic, and environmental characteristics can also be linked to these surveys at the county level (http://www.bhpr.hrsa.gov/healthworkforce/data/arf.htm). In order to investigate disparities, we recommend that health studies routinely append information on neighborhood poverty and racial composition, at a minimum. Further, we recommend that studies also append other census information, with geographic identifiers randomly recoded and confidentiality agreements in place, as has been done with the National Longitudinal Study of Adolescent Health contractual dataset (http://www.cpc.unc.edu/projects/addhealth).

3.2.2. State blood lead screening data linked to tax assessor data

Miranda et al. (2002) propose a promising approach to geocoding North Carolina blood lead screening data to the tax parcel. The tax parcel is the land entitlement unit and for residential units the tax parcel describes the housing structure (single or multifamily) and surrounding yard. Tax assessor data can potentially provide variables such as year of construction, tax value, date remodeled, and renter/owner occupied, quantifying the value and overall maintenance of the housing structure. Since age of housing can vary even within a census block, geocoding children to their actual residential unit is advantageous. Note that the quality of tax assessor data should be examined.

3.2.3. State blood lead screening data linked to census data

Krieger et al. (2003) have proposed a geocoding strategy and the use of area-based measures based on census data for monitoring SEP in the absence of socioeconomic data for state blood lead screening data. The authors advise that the best suited census tract and census block group socioeconomic measures are those: (1) most sensitive to capturing economic deprivations, (2) meaningful for comparison across regions and over time, and (3) easily interpretable with categorical cut points. The study recommends the use of census tracts, census blocks groups, and poverty-related measures based on the strongest socioeconomic gradients observed. This approach has been validated using multiple geographic levels (block group, census tract, and zip code), area-based measures (occupational class, wealth, poverty, income, education, crowding, and composite indices), and health outcomes (low birth weight, infectious disease and injury, cancer incidence, and all-cause and cause-specific mortality). The underlying assumption of this approach is that these area-based measures capture a mixture of individual- and area-level socioeconomic effects. However given that only area-level information is available, the effects of composition, context, and the interaction between composition and context cannot be differentiated.

3.2.4. NHANES linked to aerometric information retrieval system (AIRS)

Schwartz (2001) presents a multilevel analytic study using NHANES linked to AIRS data. This approach was adopted to examine the association between blood markers (fibrinogen level, platelet counts, and white blood cell count) and air pollution (PM_{10}, sulfur dioxide, and nitrogen dioxide). As this study demonstrates, associations can be made from the macro level through to the intra-individual level.

The development of appropriate multilevel data sources is an important step towards the successful application of multilevel techniques to quantify and monitor socioeconomic and racial/ethnic disparities in environmental health.

4. Conclusion

Eliminating, rather than merely reducing, racial/ethnic and socioeconomic disparities in health is a major US health policy objective. This objective, coupled with extreme residential segregation by race/ethnicity and SEP experienced by the US population, calls for the need to incorporate innovative approaches to examining risks occurring at multiple levels and over time.

Multilevel models can potentially contribute to environmental health disparities research and monitoring by providing an analytic approach for:

- Informing environmental policies and examining the impact of existing policies on local contexts and individual exposures.
- Examining a single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Examining multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Examining exposures differentially affecting subgroups of the population and/or geographic areas, and/or producing synergistic outcomes.
- Examining the fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.

- Developing theoretical models to explain environmental health disparities or to generate hypotheses.

Although numerous challenges in multilevel research remain, we call attention to emerging data and approaches for assessing the convergence of social, economic, racial/ethnic, and environmental factors operating at multiple levels simultaneously in generating and sustaining environmental health disparities.

We offer several suggestions to facilitate the integration of multilevel techniques in the study of environmental health disparities:

(a) Future studies should collect data at multiple levels or allow for the geocoding of participants (with appropriate safeguards for confidentiality). Extant data sources should, where possible, be linked with contextual data.
(b) A periodically updated "meta-contextual database" should be developed that merges environmental databases (e.g., NATA) with social data (e.g., Census). Because of the interest in health disparities, this dataset, at a minimum, should contain information on socioeconomic and racial/ethnic characteristics. The dataset should provide geographic information that will allow individual level data to be geocoded to it.
(c) Future research should evaluate which "levels" of data and analysis are most appropriate for their research questions. Published research, whether individual level, ecological level or multi-level, should justify the choice of level selected.
(d) Funding and administrative infrastructure should be made available at the federal level to facilitate the training of transdisciplinary research teams that are familiar with multilevel techniques. By their very nature, transdisciplinary teams (e.g., environmental toxicologists, psychologists, health educators, sociologists and epidemiologists) often focus on different levels of analysis. Multilevel analysis may be one approach that helps promote transdisciplinary theories as well as analysis.

In closing, multilevel techniques encompass a conceptual framework that considers individuals within their respective contexts and a set of analytic tools with which to investigate relationships between individuals and contexts. Although not without their limitations, multilevel techniques may be particularly useful for the study of environmental health disparities.

References

Armhein, C.G., 1994. Searching for the elusive aggregation effect: evidence from statistical simulations. Environ. Planning A 27 (1), 105–119.

Bingenheimer, J.B., Raudenbush, S.W., 2004. Statistical and substantive inferences in public health: issues in the application of multilevel models. Annu. Rev. Public Health 25, 53–77.

Brooks-Gunn, J., Duncan, G.J., Aber, L.J., 1997. Neighborhood Poverty: Context and Consequences for Children. Russel Sage Foundation, New York.

Bryk, A.S., Raudenbush, S.W., 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, Newbury Park, UK.

Cox, L.H., 1996. Protecting confidentiality in small population health and environmental statistics. Stat. Med. 15 (17–18), 1895–1905.

Cubbin, C., Winkleby, M.A., 2005. A multilevel analysis examining the influence of neighborhood-level deprivation on health knowledge, behavior changes, and risk of coronary heart disease: findings from four cities in northern California. Am. J. Epidemiol. 162, 559–568.

Culhane, D.P., Lee, C., Wachter, S.M., 1996. Where the Homeless Come from: A Study of the Prior Address Distribution of Homeless Families Admitted to Homeless Shelters in New York City and Philadelphia. Fannie Mae Foundation, Washington, DC.

De Leeuw, J., Kreft, I., 2001. Software for multilevel analysis. In: Leyland, A.H., Goldstein, H. (Eds.), Multilevel Modelling of Health Statistics. Wiley, New York.

Diez Roux, A.V., 2001. Investigating neighborhood and area effects on health. Am. J. Public Health 91 (11), 1783–1789.

Diez Roux, A.V., 2004. Estimating neighborhood health effects: the challenges of causal inference in a complex world. Soc. Sci. Med. 58 (10), 1953–1960.

Duncan, C., Jones, K., 2000. Using multilevel models to model heterogeneity: potentials and pitfalls. Geographical Anal. 32 (4).

Evans, G.W., Kantrowitz, E., 2002. Socioeconomic status and health: the potential role of environmental risk exposure. Annu. Rev. Public Health 23, 303–331.

Gee, G.C., Payne-Sturges, 2004. Environmental health disparities: a framework integrating psychosocial and environmental concepts. Environ. Health Perspect. 112 (17), 1645–1653.

Goldstein, H., 1995. Multilevel Statistical Models. Arnold, London.

Heywood, 1998. Introduction to Geographical Information Systems. Addison Wesley Longman, New York.

Hox, J., Kreft, I., 1994. Multilevel analysis methods. Sociolog. Methods Res. (22), 283–299.

Krieger, N., Waterman, P., Lemieux, K., Zierler, S., Hogan, J.W., 2001. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. Am. J. Public Health 91 (7), 1114–1116.

Krieger, N., Waterman, P., Chen, J.T., Soobader, M.J., Subramanian, S.V., Carson, R., 2002. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas—the Public Health Disparities Geocoding Project. Am. J. Public Health 92 (7), 1100–1102.

Krieger, N., Chen, J.T., Waterman, P.D., Soobader, M.J., Subramanian, S.V., Carson, R., 2003. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J. Epidemiol. Community Health 57 (3), 186–199.

Maantay, J., 2001. Zoning, equity, and public health. Am. J. Public Health 91 (7), 1033–1041.

Maantay, J., 2002. Zoning law, health, and environmental justice: what's the connection? J. Law Med. Ethics 30 (4), 572–593.

Macintyre, S., Ellaway, A., Cummins, S., 2002. Place effects on health: how can we conceptualise, operationalise and measure them? Soc. Sci. Med. 55 (1), 125–139.

Massey, D., 2001. Residential segregation and neighborhood conditions in U.S. metropolitan areas. In: Smelser, N.J., Wilson, W.J., Mitchell, F. (Eds.), America Becoming: Racial Trends and Their Consequences. National Academy Press, Washington, DC.

McMichael, A.J., 1999. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. Am. J. Epidemiol. 149 (10), 887–897.

Miranda, M.L., Dolinoy, D.C., Overstreet, M.A., 2002. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. Environ. Health Perspect. 110 (9), 947–953.

Morello-Frosch, R., Pastor Jr., M., Porras, C., Sadd, J., 2002. Environmental justice and regional inequality in southern California: implications for future research. Environ. Health Perspect. 110 (Suppl. 2), 149–154.

Perlin, S.A., Wong, D., Sexton, K., 2001. Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. J. Air Waste Manage. Assoc. 51 (3), 406–421.

Pollack, C., Cubbin, C., Ahn, D., Winkleby, M.A., 2005. Neighborhood deprivation and alcohol consumption: does the availability of alcohol play a role. Int. J. Epidemiol. 34, 772–780.

Sastry, N., Pebley, A.R., Zonta, M., 2002. Neighborhood Definitions and the Spatial Dimension of Daily Life in Los Angeles. Working Paper Series 03-02 DRU-2400/8. California Center for Population Research, Los Angeles.

Schulz, A., Northridge, M.E., 2004. Social determinants of health: implications for environmental health promotion. Health Educ. Behav. 31 (4), 455–471.

Schwartz, J., 2001. Air pollution and blood markers of cardiovascular risk. Environ. Health Perspect. 109 (Suppl. 3), 405–409.

Singer, J.D., 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J. Educ. Behav. Stat. 24 (4), 323–355.

Subramanian, S.V., Jones, K., Duncan, C., 2003. Multilevel methods for public health research. In: Kawachi, I., Berkman, L. (Eds.), Neighborhoods and Health. Oxford University Press, New York, NY, pp. 65–111.

Subramanian, S.V., Chen, J.T., Rehkopf, D.H., Waterman, P.D., Krieger, N., 2005. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among Black populations in Massachusetts. Am. J. Public Health 95 (2), 260–265.

US Department of Commerce, 1994. Geographic Areas Reference Manual. Economics and Statistics Administration, US Census Bureau.

US Enviromental Protection Agency, 2003. Framework for cumulative risk assessment. EPA/600/P-02/001F. Office of Research and Development, Washington, DC.

White House Office of President, 1994. Federal actions to address environmental justice in minority populations and low-income populations. In: Executive Order 12898.

Whorton, M.D., Morgan, R.W., Wong, O., Larson, S., Gordon, N., 1988. Problems associated with collecting drinking water quality data for community studies: a case example, Fresno County, California. Am. J. Public Health 78 (1), 47–51.

Williams, B.L., Florez, Y., Pettygrove, S., 2001. Inter- and intra-ethnic variation in water intake, contact, and source estimates among Tucson residents: implications for exposure analysis. J. Expo Anal. Environ. Epidemiol. 11 (6), 510–521.

Winkleby, M.A., Cubbin, C., Ahn, D. High mortality among low SES women and men living in high SES neighborhoods. Am. J. Public Health, in press.