Exposure Assessment in Environmental Epidemiology: Application of GIS Technology

John R Nuckols, PhD (1),* Mary H Ward, PhD (2), Stephanie J Weigel, PhD (1)
(1) Department of Environmental Health, Colorado State University, Fort Collins, CO; (2) Occupational Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD

Abstract

Environmental epidemiology evaluates associations between environmental exposures and health outcomes, with the purpose of further understanding the etiology of disease. An important component of such studies is exposure assessment. In many studies, exposure of participants over a relatively long period of time or large geographic region must be reconstructed. Such studies could be improved using technology based on geographic information systems (GIS). The purpose of this paper is to discuss the strengths and caveats of this use of GIS. For example, one strength is the ability to store, process, and analyze exposure data and other information about the study participants with spatial precision. This capability allows the researcher to access information that cannot always be ascertained in a traditional epidemiological study design. A good example of this is a study in which indirect exposure to environmental chemicals is being assessed for persons living in a highly integrated residential and agricultural or industrial landscape. It is unlikely that exposure to contaminants in such an environment could be accurately classified using traditional epidemiological methods such as survey questionnaires. This is because most people living in such landscapes have no knowledge of the chemicals being used and discharged into their environment. Using GIS-based technology, a researcher could locate sources of target compounds and calculate an exposure metric for each participant. Examples of such applications of GIS technology are presented in this paper. The caveats for applying a GIS in an exposure assessment do not differ substantially from other application areas for this technology. The user must be aware of cartographic issues, including scale and resolution. The accuracy of the data, the uncertainties in the analytical process, and the interpretation of the results remain important considerations in all GIS applications. This paper will illustrate the capabilities of a GIS for use in exposure assessment by applying it to an environmental exposure assessment for agricultural chemicals. Recommendations concerning the future of this technology in environmental health sciences will also be discussed.

Keywords: exposure assessment, epidemiology, risk, environmental health, remote sensing

Introduction

Environmental epidemiology evaluates the association between environmental exposures and health outcomes with the purpose of further understanding the etiology of

* John R Nuckols, Department of Environmental Health, 147 Environmental Health Bldg., Colorado State University, Fort Collins, CO 80526-1676 USA; (p) 970-491-7295; (f) 970-491-2940; E-mail: jnuckols@cvmbs.colostate.edu

disease. The term "environmental" implies a spatial component of the exposure metric used in the epidemiological study. In fact, a shortcoming of some environmental epidemiological studies has been that they do not locate subjects in the context of their true environment, which can often bias the exposure assessment.

This deficiency is demonstrated in the early history of health assessments of hazardous waste sites in the United States, conducted after the enactment of federal regulations to control environmental pollution. A National Research Council review of epidemiological studies of such sites concluded that misclassification or poor exposure metrics was a principal source of error (1). A review of the epidemiological studies indicated that, in most cases, exposure was defined as living within a specified distance of a hazardous waste site with little regard for fate and transport mechanisms of the study's target contaminants (2). None of the studies used computer modeling in their exposure assessment.

The advent the use of geographic information systems (GIS) in public health applications has greatly enhanced the capability to examine associations between environmental agents and disease. The purpose of this paper is to describe how GIS can be used in exposure assessment, as well as the strengths and caveats in applying GIS technology in this context.

Methods in Study Design

A GIS is, by definition, a database in which the information is spatially registered. However, a GIS not only maintains spatial registration, but displays the information in a mapped context. This is a major departure from the realm of numerical tables used in traditional epidemiology. A simple example of the power of maps can be demonstrated by visualizing a table composed of a list of travel destinations in the state of Colorado and their respective locations identified by latitude and longitude. Can you imagine trying to plan a vacation based on this type of information? This is the format of data typically used in the planning and implementation of exposure assessment for environmental epidemiological studies. As a result, the exposure assessment "plan" for most epidemiological studies is derived without the benefit of understanding the spatial context of the environment being studied and the ramifications of these data on the outcome of the study.

A good example of these issues can be found in epidemiological studies of agricultural workers. Most such studies are restricted to workers who use agricultural chemicals in their profession. The exposure metrics used in these studies are typically derived from a set of questions asked of the applicator concerning the type, frequency, and duration of chemical use. Many intensive agricultural regions of the country, however, are composed of a highly integrated landscape of agricultural and residential land use. In most cases, the inhabitants of these residences do not work directly in agricultural production. They may also be composed of more vulnerable populations such as children, women and men of child-bearing age, or elderly people. As such, they may be a more valuable population to study than agricultural workers if we want to get a true sense of the association between exposure to agricultural chemicals and certain disease outcomes. A traditional interview-based approach to studying this larger population is most likely doomed to failure. It is highly unlikely that individuals would have any knowledge of the types of pesticide used on the fields next to their residences or details about their use.

We have recently demonstrated the utility of a GIS in identifying populations possibly exposed to pesticides from agriculture (3). In a feasibility study, we demonstrated that satellite imagery could be used to reconstruct historical crop maps, and that crop type could be used as a surrogate for pesticide exposure. We used historical Farm Service Agency records as a source of ground reference data to classify a late summer 1984 satellite image into crop species in a three-county area in south central Nebraska. Residences from a population-based case-control study of non-Hodgkin's lymphoma were mapped using a GIS. Twenty-two percent (22%) of the residences were within 500 meters of one of the four major crops, an intermediate distance for the range of drift effects from pesticides applied in agriculture (4,5). Using information from pesticide surveys, we identified the crop pesticides that were used most frequently on those crops. This feasibility study demonstrated that a GIS coupled with remote sensing data and historical records on crop location can be used to create historical crop maps. It also showed that probable exposure to crop pesticides near a residence can be estimated when information about crop-specific pesticide use is available.

Exposure, in the purest sense of the word, is the dose of a target substance that reaches the individual being studied. Because measurement or reconstruction of dose is virtually impossible, most environmental epidemiological studies use a surrogate measure of exposure. A useful surrogate of exposure is a variable that is correlated with the true exposure of interest. For example, in the study described previously, the exposure measure is the crop area in proximity to an individual's residence. It is assumed, based on information from other studies, that this variable correlates with exposure to pesticides commonly used on the crops and thus is a useful surrogate for exposure. The surrogate exposure measure could be improved if the type and amount of pesticide actually applied to the crop fields was known. Further improvements could be made in the classification of exposure by taking into account factors such as the application method and usual wind direction and speed at the time of application.

Table 1 Definitions of Cartographic Variables

Cartographic Variable	Definition
Cartographic scale	Relates size of a feature on the ground to size of map feature
Operational scale	Scale at which process of interest occurs
Spatial resolution	Grain or smallest distinguishable unit
Geographic extent	Size of study area

Geographic principles concerning scale and resolution must be considered when using GIS in exposure assessment for epidemiology studies, especially if a surrogate variable is being used to define exposure. Definitions of cartographic variables that could affect the utility of a GIS in exposure assessments are presented in Table 1 from Lam and Quattrochi (10).

Resolution

Resolution is a very important concept in the application of GIS in environmental epidemiology. Suppose, for example, that health data used in the Nebraska study cited above were only available as cancer incidence rates at the census tract level instead of having residence location at the time of diagnosis. The exposure data would also have to be aggregated to the census tract level for the data analysis. This would have greatly compromised the utility of having exposure assessment data at a resolution of less than 500 meters.

The converse of this situation can also occur. In a recent study concerning childhood leukemia and pesticide use by Reynolds (6), the researchers were able to map health outcome data at the residence level. However, the exposure metric used in the study was pesticide use reported at a resolution of 1 square mile (640 acres), which is the reporting unit for the California Pesticide Use Reporting database (PUR). Thus, the exposure metric used in the study was much coarser than the health data because of the difference in resolution of the two datasets.

An example of the effect of resolution on exposure assessment in an agricultural production landscape is presented in Figure 1. In this figure, we demonstrate different methods that could be used for estimating potential exposure to a residence from agricultural chemicals as the spatial resolution of information increases. Plate A in Figure 1 is an example of the data resolution available to the Reynolds study described previously (6). In this case, pesticide use data are reported at the level of resolution of 1 square mile (640 acres). From the PUR, we know that atrazine was applied to 300 acres out of a total of 400 acres of crop land. However, no information is available from the PUR on the location of crops. As a result, even if the residence can be geocoded, the probability of exposure can only be defined as equal for all residences because crop location data would be necessary for further refinement of the exposure metric. Thus, as shown in Plate A, the probability of potential exposure (defined as the ratio of acres where pesticides were applied to total possible acres) is 300/640, or 47%. In Plate B, resolution of exposure data is refined by including crop map data. In this case, probability of exposure can be based on whether a residence is located within an agricultural production area (400 acres). Because the total possible acres is now reduced from 640 to 400 in the study area, the probability of potential exposure for the residence located within the agricultural land use zone is 300/400, or 75%. By this method, the residence located outside the agricultural land use zone is considered unexposed.

Further refinement is achieved in the example in Plate C. In this case, a proximity metric is employed to ascertain potential exposure to a residence. Application of this procedure is described by Ward and Nuckols et al. (3). In their study, the proximity metric was based on the distance of potential drift of pesticides for the type of agriculture used in the study area (4,5). By this method, residences that would be classified as “unexposed” by the method in Plate B could be assigned a probability of potential exposure based on the extent of pesticide use within the designated buffer zone around the residence. Plate D in Figure 1 is an example of how the area of exposure to agricultural chemicals can be further refined using computer-based fate and transport analysis. In this example, a dispersion model for fugitive chemicals migrating from an agricultural field is employed to determine the gradient of concentration in the local environment. Other models that could be used include dispersion models for chemical drift in the atmosphere and hydrologic models for estimating the dispersion of chemicals in

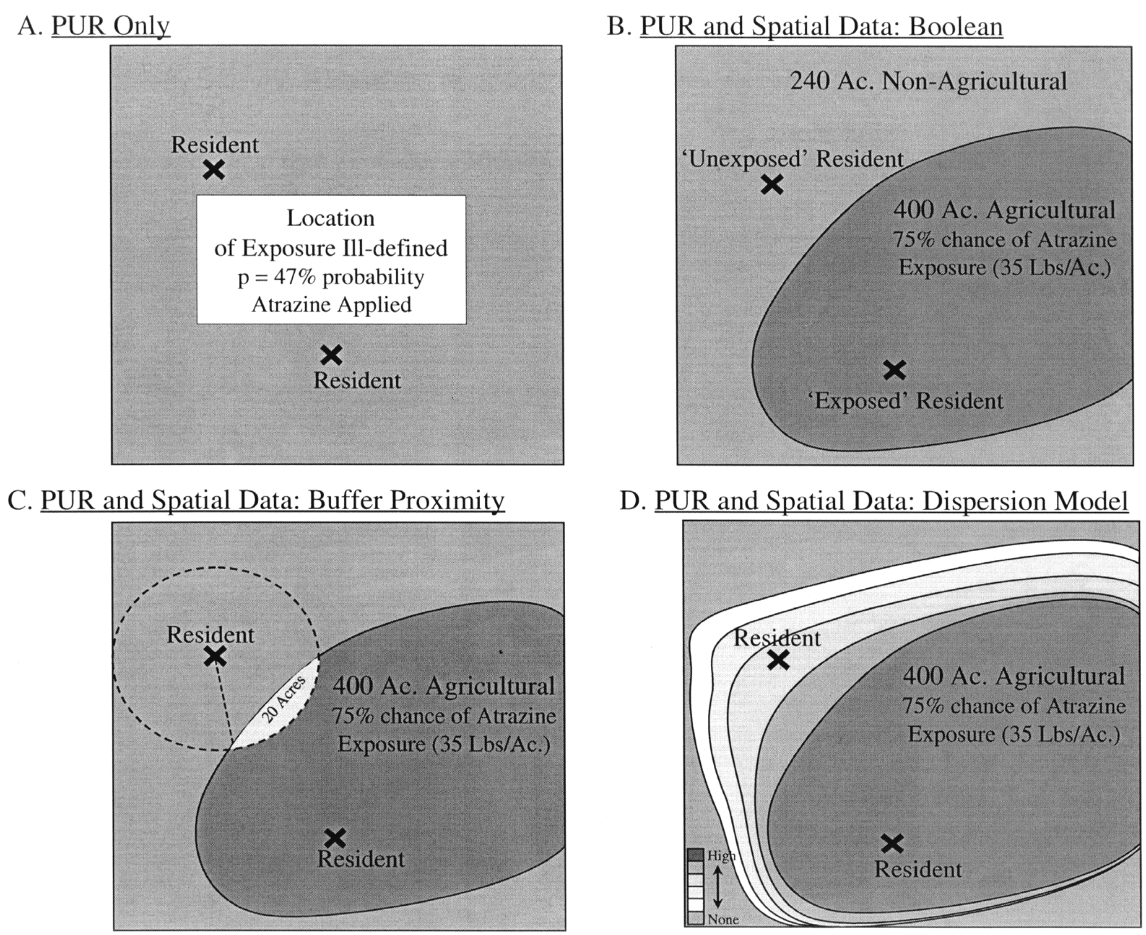

Figure 1 Methods for refining probability of exposure.

groundwater. Application of this technique for primary drift of agricultural chemicals from a spraying operation is described by Miller et al. (7).

Operational Scale

Operational scale is inherently tied to the resolution of the data that one uses because it dictates the resolution at which the exposure metric needs definition. For example, in the Nebraska study we used scientific literature concerning pesticide drift to determine that 500 meters was reasonable as an assumed distance at which drift would occur for the agricultural spraying practices that were prevalent in our study area. Thus, 500 meters became the operational scale at which we needed to be able to detect a change in cropping patterns in the area around each of our target residences. This example also points out the importance of selecting an operational scale based on scientific information about the exposure of interest, not just some arbitrary cutpoint.

Geographic Extent

Geographic extent is another important concept relevant to the application of GIS technology to exposure assessment and environmental epidemiology. Use of a GIS forces the placement of boundaries on the system being studied. Because exposure is in most cases a very dynamic process, the location of these boundaries can significantly affect the outcome of a study and the conclusions one might draw from the results. We demonstrated this in a recent health assessment study concerning a hazardous waste site near Denver, Colorado (8).

In this hazardous waste case, we were charged with determining whether residents living in the community adjacent to the site were being exposed to fugitive contaminants from the site. We concentrated on groundwater as the principal route of exposure. Over a period of several years, we used a series of metrics to classify exposure in this population. Each metric was a refinement of the previous one (i.e., starting with proximity and ending with modeling of contaminants in the water supply). With each refinement, the evidence became more convincing that contaminants that had been identified on the site were indeed present in the environment of the study population. Some of the groundwater modeling data, however, did not confirm the hypothesis that our site was the source. By extending the geographic extent of our GIS by just a few census blocks, we found that the actual source of the contaminant was another hazardous waste site located nearby. [A workshop on this issue, which uses this study as an example of the issues of scale and resolution in the application of GIS technology in exposure assessment, can be viewed at http://ehasl.cvmbs.colostate.edu (9).]

Systems Analysis

GIS is not only a tool with applications to exposure assessment and environmental epidemiology, it is a process. Once the geographic extent of the study area is defined, a GIS can be used to characterize the system of interest in terms of geophysical variables that might affect the study. For our agricultural chemicals example, this might include building layers of data in the GIS that describe the soils, geology, water supply, and meteorological and topographic factors related to pesticide transport phenomena. Sources of the target contaminant(s) can then be located within the system, and fate and transport algorithms can be applied using input data from the GIS. Some functions in GIS software (such as network modeling) can be used to predict fate and transport, but for the most part, simulation models for the particular transport medium under consideration will be required. Most simulation models have output files that can be imported into a GIS and the results displayed with some programming effort. The resulting maps of results for different media or sources can be overlaid and a composite exposure metric derived using standard functions in most GIS software. In a like manner, demographic and other information concerning the study population can be stored and manipulated in a GIS. The investigator can then test different scenarios using these exposure and other datasets to conduct epidemiological analyses.

It is important, however, to follow a standard scientific protocol in applying GIS as an analytical process. By this we mean that the study hypothesis should be defined and have biological plausibility. Having biological plausibility means that a biological basis for an association between the target substance in the exposure assessment and the disease or health outcome proposed for the study can be demonstrated. This plausibility can be based on evidence from toxicological studies or previous epidemiological studies. The exposure assessment should also take into account other factors that may be correlated with the exposure and health outcome of interest (confounding factors). An example of a confounding factor in a study of an association between a specific pesticide exposure and a disease would be another pesticide that had a similar pattern of use and was also associated with the disease.

Other important considerations in epidemiological studies using GIS are data considerations and validation of the exposure metric. There is a rule of thumb that upwards of two-thirds of the resources in a GIS project can be consumed in database preparation, geocoding, and quality assurance/quality control. Thus, it is critical in the design phase of a study to have a clear understanding of the data that are available and the data collection effort that is required. Validation of the exposure metric can be accomplished by comparing predicted versus simulated exposure variables in a field study. Validation in most cases should be site-specific. That is, the researcher should avoid the assumption that because a simulation model worked in one study area, it works in all study areas.

Discussion

Strengths

The use of GIS in public health applications is in its early stages of development, and there are many considerations that should be taken into account as one attempts to use the technology. There are also a number of research issues that need to be resolved by the scientific community. Our experience indicates that GIS can strengthen an environmental epidemiological study. When appropriately used, the technology allows the investigator to take the subject out of a numerical format and into a mapped database that can be more reflective of the subjects' environment. This can result in better study design and exposure assessment.

Caveats

The multiple databases in a GIS that describe the environment can be used as input to more precise exposure models. However, there are a number of caveats in this application of GIS technology. Perhaps the foremost caveat is that, if after calculating the geographic extent at which the study needs to be conducted[1] there are insufficient data at the scale and resolution necessary to correlate the exposure metric with the disease outcome, one should be very cautious in using a GIS.

Another important caveat in the use of GIS in exposure assessment and epidemiology is consideration of the uncertainty associated with the data. The power of a GIS is the ability to handle multiple datasets, or layers of data. It should be understood, however, that each of these data layers contains a certain degree of uncertainty. As the user adds more and more layers to the exposure metric, this compounds the uncertainty associated with the final product. How to express this uncertainty in a GIS database and carry it through the analytical process is an important research issue for the GIS community. It is important that every effort to incorporate uncertainty in the metric be made and that this information be provided to the epidemiologist. Misclassification of exposure, when it is nondifferential by disease or exposure status, dilutes the risk estimates and causes associations to be missed. An assessment of the uncertainty in the exposure variable is important so that the effects of misclassification of exposure on the risk estimates can be assessed.

Data interpolation, defined as the estimate of data values between locations of actual measurement, is another important issue in the application of GIS technology to exposure assessment. An example of such error is the interpolation of water quality data from wells across a geographic region. In an epidemiological study, the location of the study participants using wells and the locations of the wells for which there are water quality data may not coincide. One approach for assigning exposure is to apply an interpolation algorithm to the well data, creating isopleths of water quality values for points in between. These derived water quality values are then assigned to the wells of study participants where measured water quality data are not available. A caveat in using such techniques is that the spatial distribution of a substance in groundwater is highly dependent on the geophysical and hydrogeologic characteristics of the aquifer medium. Thus, if the investigator does not include this information in the interpolation procedure, the exposure metric assigned to a subject with missing data can be significantly in error.

Research Issues

There are a number of issues to consider in applying GIS technology to exposure assessments for environmental epidemiological studies. To date, most of the applications of this technology have been "retrofitted" to previously conducted epidemiological studies where a GIS was not considered in the original study design. To truly evaluate the potential use of this technology, its utility in both the design and analytical phases of epidemiological studies should be considered. One means for conducting such research would be to establish a set of long-term research sites such as those established for ecological studies by the National Science Foundation (NSF). Long-term ecological research sites were established by NSF to inventory ecological resources and to understand ecological processes that affect these resources within a specified geographic area. Subsequent research projects can then be conducted on the site to develop tools for understanding changes in the ecology. The technology for these tools can then be transferred for use in other ecological regions. By applying this approach to public health,

[1] This could be a statistical power consideration, or based on the location of a target study population.

GIS-based surveillance and analytical methods—for both long-term and "rapid-response" needs—could be developed that might improve intervention efforts.

Database development is a huge front-end expenditure for a GIS. It stands to reason that such an investment, especially for an application that is still very much in the research stage, should be made in sites that have long-term and multipurpose research potential. In fact, NSF has allocated $1 million to the development of a system of observatories in human-dominated ecosystems where long-term studies of critical ecological processes will be initiated (11). Additional support will be applied to the study of urban communities. Perhaps this is an opportunity for NSF and the National Institutes of Health to collaborate on a GIS-based research initiative.

Finally, there is a need for a support mechanism for strong interdisciplinary collaboration in the field of exposure assessment and environmental epidemiology. GIS provides a powerful platform that can be used to bridge disciplines. There are few opportunities, however, for obtaining funding for research concerning the development of exposure assessment methods with direct application to epidemiological studies. Though most agencies tout interdisciplinary research as a planning objective, little effort has been made to establish programs and proposal review sections that incorporate an interdisciplinary perspective, much less a knowledge of GIS technology.

Conclusion

There have been many advances in public health over the last century. GIS technology has the potential to revolutionize the way we approach exposure assessment in environmental epidemiology, the way we conduct health surveillance programs at the local, state, national and international levels, and the way we report health and environmental data to our citizens. To succeed with such lofty goals, we must be sure that the limitations of the technology are considered and that the use of GIS is based on sound scientific principles.

Acknowledgments

Funding for the projects cited in this paper was provided in part by the National Cancer Institute's Division of Cancer Epidemiology and Genetics (Contract NO2-CP-71100), and the Agency for Toxic Substances and Disease Registry (PA505(HARP)H75/ATH881505). Special acknowledgment to Dr. Stephanie Weigel, Mr. Chris Skinner, Mr. David Ellington, and Mr. Ryan Miller in the Environmental Health Advanced Systems Laboratory at Colorado State University for their contribution to this manuscript. Mr. Chris Skinner produced the graphic in Figure 1.

References

1. National Research Council. 1991. *Environmental epidemiology: Public health and hazardous wastes.* Washington, DC: National Academy Press.
2. Nuckols JR, Berry JK, Stallones L. 1994. Defining populations potentially exposed to chemical waste mixtures using computer-aided mapping and analysis. In: *Toxicology of chemical mixtures: Case studies, mechanisms, and novel approaches.* Ed. RSH Yang. New York: Academic Press. 473–504.

3. Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Identifying populations potentaily exposed to agricultural pesticides using remote sensing and a geographic information system. *Environmental Health Perspectives* (In press).
4. Byass JB, Lake JR. 1977. Spray drift from a tractor-powered field sprayer. *Pesticide Science* 8:117–26.
5. Frost KR, Ware GW. 1970. *Pesticide drift from aerial and ground applications.* Agricultural Engineering 51:460–7.
6. Reynolds, P. 1997. *Childhood cancer—etiologic clues using GIS*. Bethesda, MD: National Cancer Institute. RO1 CA71745-02.
7. Miller RS, Nuckols JR, Ward MH. *Estimating historical pesticide exposure using pesticide transport modeling in a GIS*. Fort Collins, CO: Environmental Health Advanced Systems Laboratory, Colorado State University. Unpublished report.
8. Reif JS, Nuckols JR, Ellington D, Weigel SJ, Burch JL, Stone AE. 1999. *Evaluation of priority health conditions at the Rocky Mountain Arsenal: A re-analysis based on geographic information system-derived exposure assessment for trichloroethylene. Final Report*. Atlanta, GA: Agency for Toxic Substances and Disease Registry. PA505(HARP)H75/ATH881505. (In review).
9. Nuckols JR, Weigel SJ. 1998. *Exposure assessment using GIS technology.* Fort Collins, CO: Environmental Health Advanced Systems Laboratory, Colorado State University. http://ehasl.cvmbs.colostate.edu. (See "GIS in Public Health—San Diego, 1998" at this Web site.)
10. Lam NS-N, Quattrochi DA. 1992. On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer* 44(1):88–98.
11. National Science Foundation. 1997. *NSF funds first long-term studies of urban ecology.* NSF PR 97-63. Washington, DC: National Science Foundation.