

## Improving exposure assessment in environmental epidemiology: Application of spatio-temporal visualization tools

Jaymie R. Meliker<sup>1</sup>, Melissa J. Slotnick<sup>1</sup>, Gillian A. AvRuskin<sup>2</sup>,  
Andrew Kaufmann<sup>2</sup>, Geoffrey M. Jacques<sup>2</sup>, and Jerome O. Nriagu<sup>1</sup>

<sup>1</sup> Department of Environmental Health Sciences, School of Public Health,  
University of Michigan, Ann Arbor, MI 48109-2029, USA (e-mail: jmeliker@umich.edu)

<sup>2</sup> BioMedware Inc., 516 N. State St., Ann Arbor, MI 48104, USA

Received: 13 November 2003 / Accepted: 25 October 2004

**Abstract.** A thorough assessment of human exposure to environmental agents should incorporate mobility patterns and temporal changes in human behaviors and concentrations of contaminants; yet the temporal dimension is often under-emphasized in exposure assessment endeavors, due in part to insufficient tools for visualizing and examining temporal datasets. Spatio-temporal visualization tools are valuable for integrating a temporal component, thus allowing for examination of continuous exposure histories in environmental epidemiologic investigations. An application of these tools to a bladder cancer case-control study in Michigan illustrates continuous exposure life-lines and maps that display smooth, continuous changes over time. Preliminary results suggest increased risk of bladder cancer from combined exposure to arsenic in drinking water ( $>25 \mu\text{g}/\text{day}$ ) and heavy smoking ( $>30$  cigarettes/day) in the 1970s and 1980s, and a possible cancer cluster around automotive, paint, and organic chemical industries in the early 1970s. These tools have broad application for examining spatially- and temporally-specific relationships between exposures to environmental risk factors and disease.

**Key words:** GIS, exposure, Epidemiology, Arsenic, STIS

---

This study was supported by grant R01 CA96002-10, Geographic-Based Research in Cancer Control and Epidemiology, from the National Cancer Institute. Development of the STIS<sup>TM</sup> software was funded by grants R43 ES10220 from the National Institutes of Environmental Health Sciences and R01 CA92669 from the National Cancer Institute. Access to cancer case records was provided by Michigan Cancer Surveillance Program within the Division for Vital Records and Health Statistics, Michigan Department of Community Health. The authors thank Michigan Public Health Institute for conducting the telephone interview and Stacey Fedewa and Lisa Bailey for entering written surveys into a database. The authors thank 3 anonymous reviewers for their helpful comments.

## 1 Introduction

A key component of environmental epidemiologic research is the assessment of historic exposure to environmental contaminants. Exposure assessment efforts are complicated by temporal variation in contaminant concentrations and human behaviors, and by residential and occupational mobility patterns. Despite temporal variation in exposure, existing assessment tools generate exposure estimates that under-emphasize time by ignoring variation within pre-defined time intervals (Steinmaus et al. 2003; Charles et al. 2003). For example, tools such as cumulative exposure calculations, time-window cumulative exposure estimates, and peak exposures over definitive time intervals are unsatisfactory because a critical span of time could be overlooked or attenuated. Other methods rely on recent estimates of exposure (e.g., levels of contaminants in blood) and assume those estimates to be temporally invariant (Gammon et al. 2002). These types of exposure assessments can be misleading due to extended latency periods for chronic diseases, such as cancers.

Collecting and constructing detailed temporal datasets is an arduous undertaking in environmental epidemiologic research, and for this reason, investigators often under-emphasize the temporal dimension. For instance, in a New York cancer study, serum organochlorine (OC) levels measured shortly after breast cancer diagnosis were used as a key measure in an exposure assessment (Gammon et al. 2002), even though periods of historic OC exposure are of more etiologic importance. In another study, cancer cluster investigations relied on residential location at time of diagnosis or time of death (Kulldorff et al. 1998), despite the likelihood that exposure to potential etiologic environmental agents occurred years prior to diagnosis. Quality historical data, such as exposures that occurred 20, 30, or 40 years ago, are extremely difficult, if not impossible, to obtain. Other historical datasets, however, can be gathered with some degree of assurance: residential mobility patterns, shifting locations of polluting industries, changes in behavior such as water or food consumption habits, and historical pollutant concentrations. Existing exposure assessment efforts do not adequately consider temporal variation because temporal datasets are seldom collected; when they are collected, suitable visualization tools are not available.

Visualization tools are necessary to explore relationships between variables, and how those relationships change over time. Despite modern computer technologies for storing and managing temporal and spatio-temporal datasets, surprisingly few tools are available for visualizing the “what, where, and when” of events (Andrienko et al. 2003; Chittaro et al. 2003). One visualization tool, geographic information system (GIS) software, enables users to visualize what happened, and where; and has augmented assessment of exposure to environmental contaminants. For example, researchers have geocoded locations of industries, industrial waste sites, pollution plumes, as well as homes, schools and jobs of study participants. From these geocoded features, disease maps have been created and spatial analyses performed (Brauer et al. 2003; Maantay et al. 2002; Meliker et al. 2001; Reif et al. 2003; Swartz et al. 2003). A frequent criticism of GIS, however, is its inability to support temporal data structures (Beaubroef and Breckenridge 2000; Dragicevic and Marceau 2000). This can be problematic

if, for example, investigators wish to explore whether residences of cancer cases cluster at any time in the past 50 years; or whether living in close proximity to a chemical industry at any point in time is associated with subsequent cancer development. With GIS, spatial patterns at different isolated moments can be examined, using animation tools. However, each map, or snapshot, must be created independently, requiring a substantial amount of effort and introducing greater likelihood of data entry error. Furthermore, information about change is not available in the interval between two consecutive snapshots. Visualization of a map that displays smooth, continuous changes over time can generate additional insights about spatial disease patterns.

Spatio-temporal visualization tools support evaluation and query of spatio-temporal datasets, and also can enrich analysis of temporal datasets that are devoid of geographic coordinates. A backbone of these tools is their temporal data structure, which enables non-geographic attributes, such as temporally-variant exposure estimates, to be visually examined. An example of a visualization tool, exposure life-lines, is introduced in this paper. Exposure life-lines facilitate scrutiny of continuous exposure estimates, permitting examination of whether exposure at any point in time is associated with subsequent disease development. Such an analysis cannot be undertaken using existing analytical software.

In this paper, space-time visualization methods are illustrated using preliminary data from a bladder cancer case-control study in Michigan. Established causes of bladder cancer include cigarette smoking and exposure to aromatic amines in occupational settings; however, many cases of bladder cancer remain unexplained. One possible cause of bladder cancer is exposure to arsenic in drinking water. Concentrations of arsenic in drinking water exceeding World Health Organization (WHO) and US Environmental Protection Agency (EPA) guidelines ( $10 \mu\text{g}/\text{L}$ ) have been identified in ground-water supplies of 11 counties in southeastern Michigan: Genesee, Huron, Ingham, Jackson, Lapeer, Livingston, Oakland, Sanilac, Shiawassee, Tuscola, and Washtenaw (Kim et al. 2002; Kolker et al. 2003; Slotnick et al. 2003). Previous individual-level studies of arsenic in drinking water and bladder cancer have been criticized for imprecise exposure assessments (Cantor 2001) which failed to account for (1) changes in arsenic concentration in water over time, (2) individual residential mobility patterns, and (3) behavioral changes in drinking water consumption. These shortcomings are familiar to many investigations of environmental exposures and cancer; spatio-temporal visualization tools are essential for alleviating some of these shortcomings.

## 2 Data and methods

A sample of 39 cases and 39 controls from a bladder cancer case-control study in Michigan was selected to demonstrate applications of spatio-temporal visualization tools for exposure assessment. This size dataset allows for straightforward manipulation and visualization of complex exposure scenarios, enhancing communication of the intricacies of these visualization tools. Cases were recruited from the Michigan State Cancer Registry and diagnosed in the year 2000. Controls were frequency matched to cases by age ( $\pm 5$  years), race, and gender, and recruited using a random digit dialing

procedure from an age-weighted list. To be eligible for inclusion in the study, participants must have lived in the eleven county study area for at least the past 5 years and had no prior history of cancer (with the exception of non-melanoma skin cancer). Participants were offered a modest financial incentive and research was approved by University of Michigan IRB-Health Committee. Participants answered a telephone questionnaire concerning drinking water habits, smoking, and medical history, and completed a written questionnaire describing residential mobility history. Information obtained from these questionnaire instruments was used to create spatio-temporal datasets.

This report describes the key functionalities of a software being developed to facilitate spatio-temporal visualization. An early version of the software, STIS<sup>TM</sup> (TerraSeer, Ann Arbor, MI), supports spatio-temporal datasets but does not yet generate all of the visualization capabilities that are described here.

## *2.1 Temporal and spatio-temporal datasets*

### 2.1.1 Residential mobility

Participants provided written residential histories of each home lived-in for at least one year (for a total of 519 homes). The duration of residence and exact street address were obtained, otherwise the closest cross streets were provided. Each residence in the eleven county study area was geocoded and assigned geographic coordinates in ArcGIS; residences outside the study area were not geocoded. Participants resided at 288 homes within the study area, with time spent averaging 66% of their lifetimes. Street files were downloaded from Michigan Center for Geographic Information website, and were part of the Michigan Geographic Framework. Michigan Geographic Framework datasets use the Michigan GeoRef System, based on an Oblique Mercator projection. Residences within the study area were successfully geocoded: 78% automatically matched using ArcGIS settings of spelling sensitivity equal to 80, minimum candidate score equal to 10, and a minimum match score equal to 60. The unmatched addresses (22%) were manually matched using cross streets with the assistance of internet mapping services. If cross streets were not provided, best informed guess placed the address on the road (3%), and as a last resort, residence was matched to town centroid (2%).

### 2.1.2 Water supply history

Participants provided written information about primary water supply and any changes in water supply at each residence (e.g., public surface, public well, private well, or bottled water). Managers of 135 public water supplies in the study area answered questions about quality of drinking water, source of water, changes in water supply, changes in extent of coverage of water supply, and changes in treatment procedures. Each residence was classified by its primary water supply for a span of time, based on accounts provided by participants. At approximately 3% of the addresses, participants did not assign a water supply; STIS<sup>TM</sup> was used in those cases to assign a water supply (see Sect. 3.3).

### 2.1.3 Arsenic data

The water sample that provided current arsenic exposure was collected from the kitchen tap, or primary source of water for drinking and cooking at each participant's current home. All plasticware was acid-washed for trace metals determination following modification of a previously described protocol (Nriagu et al. 1993). Blanks and replicates were collected for at least 10% of the samples. Water samples were stored on ice, acidified with 0.2% trace metal grade nitric acid, and refrigerated until analysis. Water samples were subsequently analyzed for arsenic using an inductively coupled plasma mass spectrometer (ICP-MS, Argilent Technologies Model 7500c).

Historic databases were used to estimate arsenic concentrations at past residences. Michigan Department of Environmental Quality (MDEQ) maintains a database of arsenic measurements (1993–2002) in private (N=11,615 arsenic measurements) and public well water supplies (N=1675 arsenic measurements) in the study area, analyzed in a state laboratory with graphite furnace atomic absorptions spectrometry (GF/AAS) (1993–1995), hydride flame (quartz tube AAS) (1993–1995), and an ICP-MS (1996–2002). Private well water measurements from MDEQ database were utilized to generate city or township averages (means) of arsenic concentrations for private well waters not yet monitored for arsenic. This MDEQ database of public well water supplies was used to calculate a mean value of arsenic for each community's ground-water supply. Community supplies relying on surface water were assigned an arsenic concentration equal to 0.3  $\mu\text{g/L}$ , the mean level detected in tap water samples that rely on surface water in the area. Residences outside the study area were similarly assigned an arsenic concentration of 0.3  $\mu\text{g/L}$ . Arsenic concentrations in private and public water supplies were assumed not to change over time, since evidence suggests limited temporal variability (Ryan et al. 2000). The bladder cancer case-control study is in progress and, therefore, arsenic concentration estimates from an ongoing modeling effort are not yet available; however, the arsenic database presented here is adequate for highlighting some of the benefits of spatio-temporal visualization tools for generating insights into exposure-disease relationships.

### 2.1.4 Water consumption patterns

Estimates of water consumption (liters/day) were calculated based on answers to a series of questions from a telephone interview. Participants were asked to self-report the number of glasses of water and beverages made with water drank at home during the past year (the year prior to cancer diagnosis for cases), the previous ten years, and changes in drinking water consumption over the course of a lifetime.

### 2.1.5 Industrial locations

Industries believed to emit contaminants that have been associated with xbladder cancer or believed to emit arsenic were identified using the Toxics Release Inventory (USEPA 2000) and the Directory of Michigan

Manufacturers (Manufacturer Publishing Co., 1946, 1953, 1960, 1969, 1977, 1982). Standard Industrial Classification (SIC) codes were adopted, but prior to SIC coding, industrial classification titles were selected. Characteristics of 245 industries, including, but not limited to, cigarette manufacturing, fabric finishing, wood preserving, pulp mills, industrial organic chemical manufacturing, and paint, rubber, leather, and battery manufacturing, were compiled into a database. In total, 21 three-digit SIC codes and 17 four-digit SIC codes were used to select industries. Industries were geocoded following the same matching procedure as described for residences: 89% matched to the address, 5% were placed on the road using best informed guess, and as a last resort, 6% were matched to town centroid. In addition, a MDEQ list and ranking of 549 contaminated sites in the study area was successfully geocoded (100%).

## *2.2 Calculating exposure estimates*

Exposure was calculated by multiplying arsenic concentration ( $\mu\text{g/L}$ ) by volume of water drunk at home and used for making beverages at home (L/day). Each change in water consumption was used to estimate exposure for a particular time-window, assuming that the arsenic concentration was constant. Results are presented as average exposure to arsenic in units of  $\mu\text{g/day}$ . Exposure calculations were performed for 587 unique space-time periods, with each space-time period defined by a unique combination of residential location, water source, water treatment, and water consumption behavior. These exposure estimates are preliminary and used to illustrate potential of spatio-temporal visualization tools for exposure reconstruction. Other components of a detailed exposure assessment, such as water consumed at work, occupational exposures, and uncertainty associated with estimates of arsenic concentration, are being assessed but are not reported in this paper.

## **3 Features and architecture of spatio-temporal visualization tools**

### *3.1 Data structure*

Spatio-temporal visualization tools support datasets in which time is a principal feature. For example, the data structure used by STIS<sup>TM</sup> requires that: (1) each row in a dataset represents a space-time intersection and a variable of interest and (2) when a geographic location or a variable of interest changes value, a new row is created. Data tables can be linked to spatial features, and the following information must be specified: whether the data are time-dependent, unique ID, start date, and end date. In the example of drinking water history and residential mobility of participants, each home and water source occupy a unique row with a start year, end year, geographic coordinate, and participant ID number. Any change in location of residence or source of drinking water is characterized by a new row with a defined duration, using the same participant ID number (Table 1). Other variables, such as water consumption patterns, use of home water treatment systems, and concentrations of drinking water contaminants, are stored

**Table 1.** Spatio-Temporal Dataset Format for STIS Point Features: Residential Mobility History and Water Consumption History for Two Participants

Residential mobility history					
Sample ID	Year Moved In	Year Moved Out	Address <sup>a</sup>	X-Coord <sup>a</sup>	Y-Coord <sup>a</sup>
001366	04/08/1951	01/01/1963	Address #1	694980	264132
001366	01/01/1963	01/01/1971	Address #2	687299	268878
001366	01/01/1971	01/01/1972	Address #3	694161	272042
001366	01/01/1972	01/01/1975	Address #4	680421	278791
001366	01/01/1975	01/01/2004	Address #5	649645	275342
001397	01/01/1933	01/01/1937	Address #1	692980	168978
001397	01/01/1937	01/01/1950	Address #2	687699	174042
001397	01/01/1950	01/01/1953	Address #3	692161	176791
001397	01/01/1953	01/01/1957	Address #4	660421	177342
001397	01/01/1957	01/01/1964	Address #5	684656	274665
001397	01/01/1967	01/01/1969	Address #6	694766	278743
001397	01/01/1969	01/01/1993	Address #7	686910	274183
001397	01/01/1993	01/01/1998	Address #8	692830	280704
001397	01/01/1998	01/01/2004	Address #9	685618	270049
Water consumption history					
Sample ID	Start Period	End period	Estimated Water Consumption (Liters/day)		
001366	04/08/1951	01/01/1995	1.5		
001366	01/01/1995	01/01/2004	2.75		
001397	04/08/1933	01/01/1949	0.25		
001397	01/01/1949	01/01/1982	1.0		
001397	01/01/1982	01/01/2004	0.6		

<sup>a</sup>Address and geographic coordinates are altered to protect participant confidentiality.

in separate datasets, including rows with defined durations and ID numbers. Despite different durations for variables, participant datasets can be joined together, using participant ID numbers. Analogous to GIS, using spatio-temporal visualization tools, a participant's mobility history can be visualized by displaying specified attributes of a participant. In effect, changes in water consumption patterns, water supply, contaminant concentrations in water, or other variables can be illustrated as participants' move through time.

From a data input perspective, datasets of point, line, or polygon features are structured similarly. As with point features, visualizing changes in community attributes requires geographic coordinates, a duration, unique community ID number, and characteristics of a community. These variables are recorded in unique rows for each space-time intersection. The ID number remains constant, even if the geographic area of the community changes—analogue to a unique ID number for each participant, even though residences change. In this manner, polygons can change shape and attributes over time.

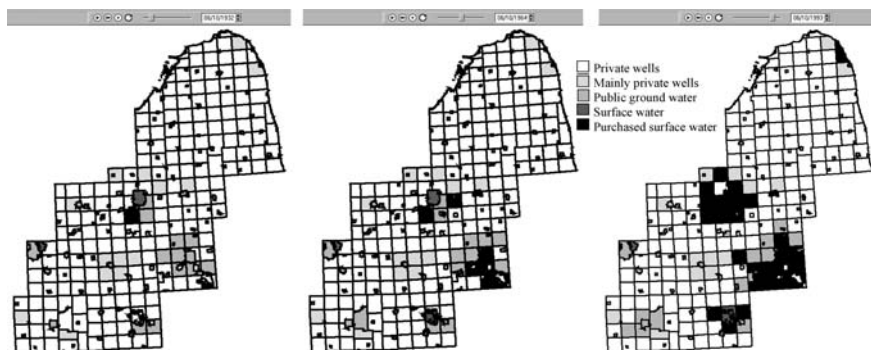


Fig. 1. Public and private water supply status in Southeastern Michigan in 1932, 1964, and 1993

### 3.2 Space-time maps

Smooth, temporally continuous, space-time maps are created in STIS<sup>TM</sup> using the data structure described above. To illustrate the changes in the source of drinking water, a series of snap-shots are presented, representing the region's water supply status in 1932, 1964, and 1993 (Fig. 1). The slider bar is dragged to the left (distant past) or right (more recent past) to display different years. With limitations of the printed page, only static images can be presented here; however, STIS<sup>TM</sup> produces continuous space-time animations.

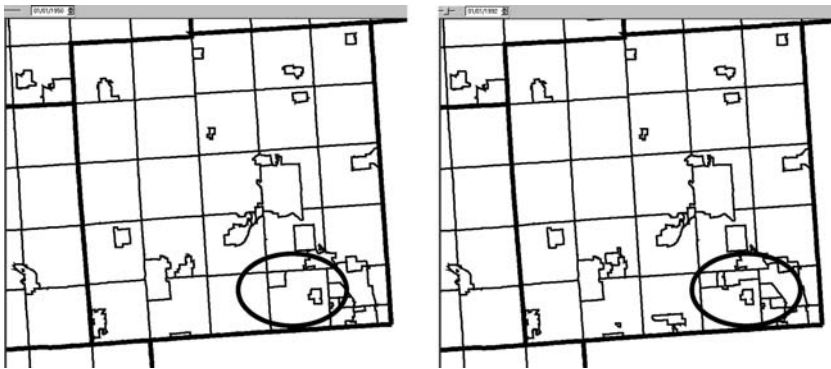
In the public water supply example, in addition to assigning a water supply source to a participant's residence, valuable information can be gained by visualizing water supply changes. Water supply is designated as: private wells, public ground-water, public surface water, purchased surface water, or mainly private wells (i.e., some small residential developments in the community have public ground-water supplies). From 1932 through 1993, most of the region was served by private wells (white color in Fig. 1). In 1932, only a few communities were served by surface water; by 1993, several communities purchased surface water from the city of Detroit water system. Visualization of changes shows that over time, small communities developed public ground-water systems and some public ground-water systems changed over to public surface water distribution systems.

In addition to attributes changing, variations in town boundaries are displayed, as when new towns become incorporated, communities expand their borders, and when communities merge (Fig. 2). Between 1950 and 1992, several new communities were incorporated in Oakland County. In the database, each community is assigned a unique ID number that remains the same; any other variable, including geographic coordinates, is permitted to change.

### 3.3 Space-time queries

Given space-time maps, a natural extension is for the software to process space-time queries. This procedure would reduce data entry time and error





**Fig. 2.** Feature of spatio-temporal visualization tools. Town boundaries change with time (1950, 1992)

associated with manually assigning an attribute to a space-time object. For example, in the bladder cancer study, participants have been unable to recall their drinking water source at approximately 3% of their lifetime residences. The public water supply map (Fig. 1) from 1920–2003, was overlaid with the residential history map, and each residence without an assigned drinking water source was visually compared with the public water supply map and assigned a drinking water source. A space-time query feature would make this assignment automatic.

As a further example, space-time query tools could be used to assign an arsenic concentration to drinking water at each residence. Distinct space-time maps of arsenic concentration could be created for residences on public water supplies and private wells. A space-time query procedure could differentiate between residences on private well water and those on public water supply, and assign arsenic to each residence, at any point in time. This procedure highlights the broad application of spatio-temporal visualization tools in facilitating exposure calculations and therefore improving exposure assessment.

## 4 Visualization examples

### 4.1 Spatio-temporal patterns: Cancer clusters based on residential histories

Space-time maps allow for examination of cancer clusters using former residences of cases and controls. The ability to scan a continuous space-time map for potential cancer clusters at any point in time is a valuable exploratory tool with considerable potential for research. A continuous space-time map of public water supply history was overlaid by locations of current and past residences for 39 cases and 39 controls in STIS<sup>TM</sup>. The map was scanned to investigate if cases or controls drank different water sources at any point in time. In 1946, eight cases and four controls lived in Genesee County and six of the cases and two of the controls drank well water; the other cases and controls imbibed surface water (figure not shown to protect participant confidentiality). Association between residential location and

drinking water source among cases and controls will be explored in finer detail when additional data are collected and geocoded.

In addition to data on public water supply history, other spatio-temporal databases were added to the space-time map. Current and former locations of selected industries and contaminated sites were overlaid to investigate if the cancer cases cluster around these sites. In 1973, former residences of four cases appeared to cluster together, near automotive, paint, and organic chemical industries in Oakland County (figure not shown to protect participant confidentiality). These results are suggestive of an association between bladder cancer and industrial pollution in the community. As additional data are collected, these findings will be investigated in greater detail.

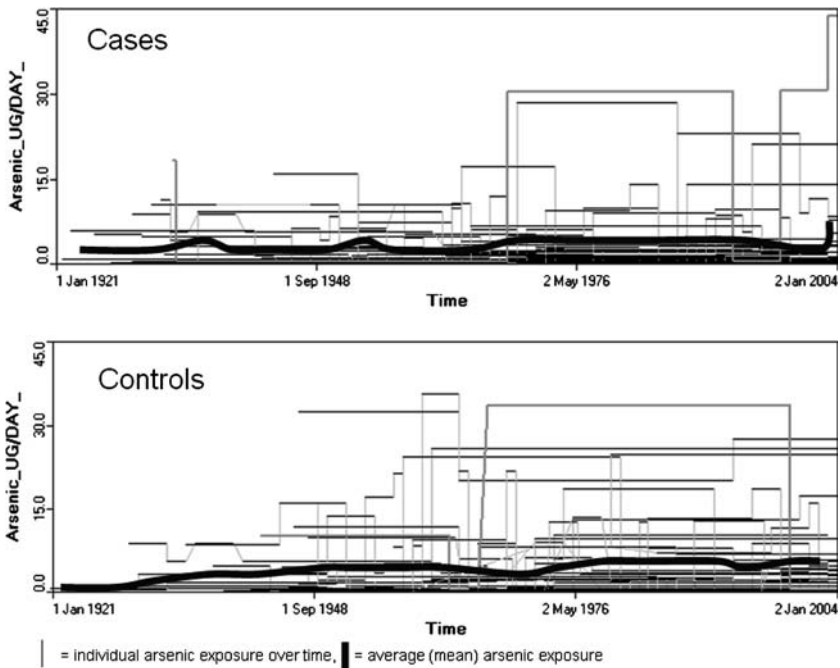
Spatial patterns based on participant's age may generate fresh insights, compared with those based on specific dates. For example, a cluster of former residences of cancer cases at a specific year may indicate environmental contamination during that year. A spatial cluster of former residences of cancer cases when they were children, irrespective of year, may indicate early-lifetime vulnerability to an environmental exposure in the area. Spatio-temporal visualization tools could be used to display cancer clusters based on the years a participant lived at a residence, or to display cancer clusters of similarly aged participants. In this manner, clusters of children could be visualized, even if the children are born in different generations.

The spatio-temporal visualization tools being developed will enable the user to scan a smooth, continuous space-time map to look for patterns at any slice of time. Using a space-time map, however, differentiating each participant's movement in time is difficult. One method for tracking an individual on a map is by connecting lines between residences, creating geospatial life-lines which can be used to trace individual movement in space and time (Hornsby and Egenhofer 2002; Kwan 2000). With a large study sample, say of over 1000 participants, life-lines intersect and monitoring individual movement becomes confusing. Visual comparison of individual life-lines on a map, therefore, is challenging. Alternatives to geospatial life-lines that may be easier to visually comprehend are tools highlighting the temporal component, without regard for geographic information. Since a backbone of spatio-temporal visualization tools is a temporal data structure, non-geographic attributes, such as temporally-variant exposure estimates, can be visually examined.

#### *4.2 Visualizing temporal patterns*

One of the most commonly used temporal visualization tools is the time series graph. A time series of arsenic exposure for each case and control is shown in Fig. 3. Each line represents a different participant's arsenic exposure trajectory. The thick line depicts average arsenic exposure for cases and controls, in respective graphs. Individual participants' trajectories are difficult to follow, visually, because the lines intersect. The average arsenic exposure trajectory for cases and controls may generate insights but information is lost by simply averaging participants' exposures.

Other traditional tools, such as histograms and scatter plots, display variables at slices of time in STIS<sup>TM</sup>. Similar to space-time maps, these tools allow for scanning smooth, continuous, temporally-variant figures for



**Fig. 3.** Time series graphs of arsenic exposure: Cases and controls. Each line represents a participant's average daily exposure to arsenic ( $\mu\text{g}/\text{day}$ ) over his/her lifetime

relationships between variables at any moment in time. Histograms display one variable, and scatter plots display two variables, at any time slice. For example, histograms for cases and controls were compared for drinking water source. In 1965, 15 cases and 11 controls were drinking well water, while 24 cases and 28 controls were drinking surface water (Fig. 4). Scatter plots were used to compare arsenic exposure and cigarette smoking for cases and controls (Fig. 5). In 1972, controls with arsenic exposure exceeding  $20 \mu\text{g}/\text{L}$ , smoked fewer than 21 cigarettes/day. Cases with arsenic exposure exceeding  $20 \mu\text{g}/\text{L}$ , in comparison, smoked greater than 30 cigarettes/day. Limited evidence suggests that cigarette smokers, exposed to elevated levels of arsenic 30–40 years ago (Bates et al. 1995) or 40 or more years ago (Steinmaus et al. 2003) are at an increased risk for bladder cancer. But the temporal relationship between cigarette smoking and arsenic exposure has not been well documented. For example, when does smoking cigarettes interact with arsenic exposure: if exposure is simultaneous; if heavy smoking precedes a period of elevated arsenic exposure; or if heavy smoking occurs following a period of elevated arsenic exposure? Is there a critical time when cigarette smoking and arsenic intake interact to increase risk of bladder cancer? Scatter plots help researchers address these types of questions, and early results using this preliminary dataset suggest that a combination of heavy smoking and elevated arsenic about 30 years prior to cancer diagnosis may exacerbate risks for bladder cancer.

While scatter plots, histograms, time graphs, geospatial life-lines, and space-time maps, can each be employed to generate insights about temporal

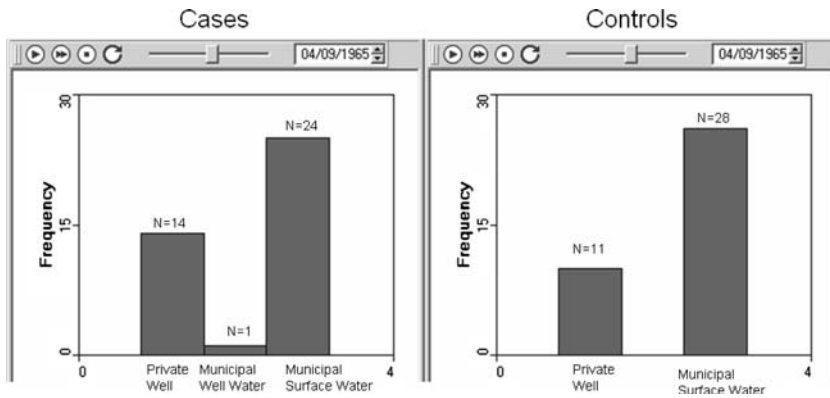


Fig. 4. Histogram of source of drinking water in 1965: Cases and controls. Slider bar controls the year displayed

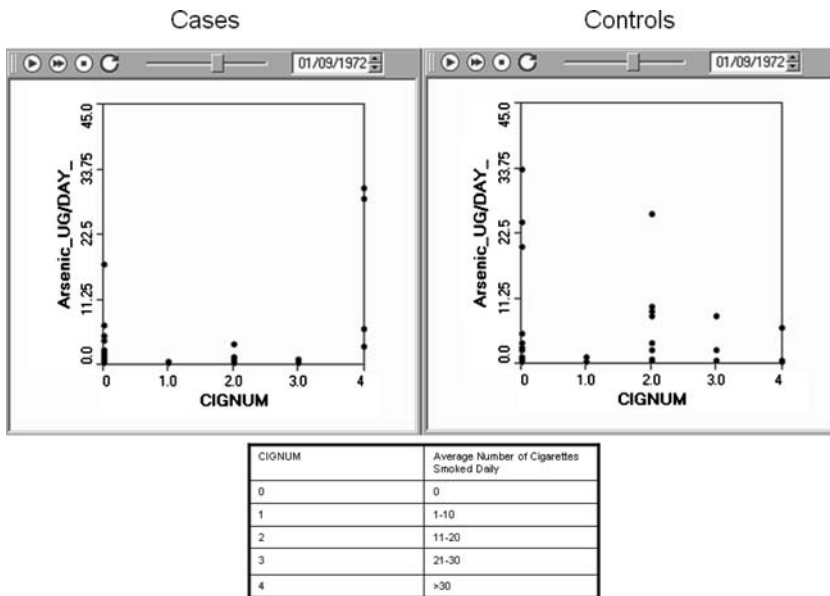
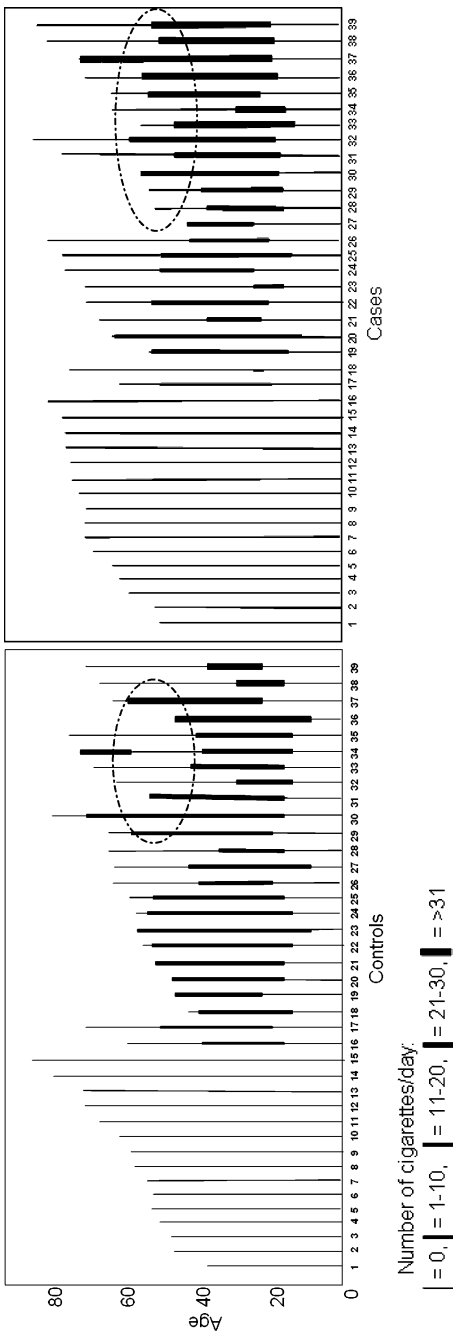
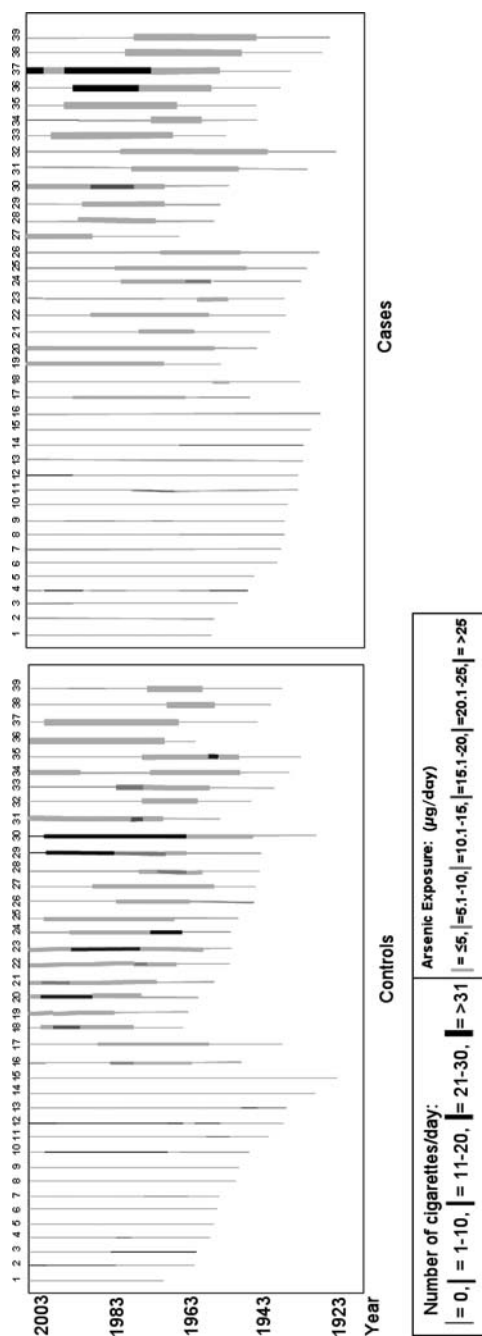


Fig. 5. Scatter plot of arsenic exposure and cigarettes smoked in 1972: Cases and controls. Two cases exposed to elevated arsenic exposure and > 30 cigarettes/day

and spatio-temporal datasets, these tools do not display the entire temporal dataset in a straightforward manner. The tools rely either on time slices of continuous maps or crisscrossing trajectories of participants, in which the relationships between participants, their exposure, and how exposure changes with time, are difficult to visually comprehend. One solution is exposure life-lines, which display participants on the X axis, time on the Y axis, and measures of exposure in the life-lines' color or thickness.



**Fig. 6.** Exposure life-lines for cigarette smoking: Cases and controls. Thickness of life-line increases with higher frequency of cigarettes smoked. There appear to be more heavy smokers around 40–50 years old among cases, compared with controls



**Fig. 7.** Exposure life-lines for arsenic exposure and cigarette smoking: Cases and controls. Thickness of life-line increases with frequency of cigarettes smoked. Darkness of life-line increases with higher arsenic exposure. A pair of heavy smoking cases was exposed to arsenic  $>25\mu\text{g}/\text{day}$  in 1970s and 1980s

Exposure life-lines (Fig. 6) are presented to illustrate changes in cigarette smoking exposure among cases and controls. The thickness of the life-line reflects the average number of cigarettes smoked each day, with thicker lines indicating heavier smoking. Using these life-lines, investigators can compare cases and controls to evaluate whether cases began smoking at a younger age, quit at an older age, or whether there was a particularly vulnerable period in which heavy smoking appears to be strongly associated with bladder cancer. Similar numbers of cases (16) and controls (15) never smoked cigarettes, but seven cases smoked  $> 30$  cigarettes/day, in comparison with four controls. In addition, nine of the twelve cases who smoked  $> 20$  cigarettes/day, were smoking in their 40s and 50s, in comparison with just three of the nine controls who smoked more than a pack a day. While cigarette smoking is an established risk factor for bladder cancer, the timing of when cigarette smoking might exacerbate bladder cancer risk is not well understood. The apparent temporal cluster of heavy smokers among cases in their 40s and 50s could shed light on the temporal relationship between cigarette smoking and bladder cancer.

As Fig. 6 illustrates, exposure life-lines may be used to investigate temporal clusters of high exposure at any point in time. They are particularly useful for investigating the latency period of exposure-disease relationships. These life-lines are a substantial improvement over such methods as cumulative exposure estimates or pre-defined time windows of exposure, which rely on temporally-aggregated exposure estimates. Exposure life-lines are helpful to investigate whether any years of high exposure are more prominent in cases compared with controls.

Bivariate exposure life-lines can be constructed to represent historic exposure levels from two variables with shades of grey for the first variable, and thickness for the second variable. Exposure life-lines depicting arsenic exposure and cigarette smoking history for cases and controls are shown in Fig. 7. A life-line's thickness increases with frequency of cigarettes smoked; increases in darkness correspond to higher levels of arsenic exposure. It is seen that more controls, compared with cases, were exposed to elevated levels of arsenic in drinking water over the course of their lives. Only one control, however, experienced simultaneous exposure to drinking water arsenic  $> 25 \mu\text{g}/\text{day}$ , and  $> 20$  cigarettes/day, and that was only for three years, from 1953-1956. In comparison, two cases were simultaneously exposed to arsenic exposure  $> 25 \mu\text{g}/\text{day}$ , and  $> 30$  cigarettes/day in the 1970s and 1980s, suggesting the possibility of interaction between arsenic and cigarette smoking, twenty-to-thirty years prior to cancer diagnosis.

Without exposure life-lines, an a priori hypothesis with a specified point in time is required to investigate when multiple variables interact. For example, researchers could not visualize and evaluate the relationships between arsenic exposure at any point in time, cigarette packs smoked at any point in time, and subsequent disease development. Exposure life-lines can be used to evaluate these relationships and shed light on interactions between variables, and when those interactions may be occurring.

In comparison to Fig. 6 where age was used to track time, Fig. 7 uses years. Displaying different measures of time, such as years and age, is beneficial in highlighting distinct trends in the data. In some situations, years

of exposure may be important, since they indicate the amount of time between exposure and subsequent disease development. In other situations, age at exposure may be critical, because individuals may be more vulnerable to toxic agents when they are at a specific age. Exposure life-lines can be flexible to present data using either measure of time.

## 5 Future directions

Application of spatio-temporal visualization tools for exposure assessment and environmental epidemiology is in its infancy. Processing and supporting visualization of spatio-temporal datasets is essential for exploring patterns in data. Beyond visualization, however, statistical analyses are necessary to identify significant relationships between timing or location of exposure and subsequent disease development. Statistics deemed important include: spatial and spatio-temporal clustering techniques; focused cluster tests to examine if a cluster is associated with a point source, such as an industry or landfill; temporal analyses to investigate if exposure at a point in time is associated with subsequent disease development; and epidemiologic odds ratio analyses.

Another area of future research is the propagation of uncertainty associated with calculation of exposure profiles. Specific to the bladder cancer example presented here, uncertainty in exposure assessment arises at several levels in data collection and manipulation, including water consumption estimates, measured arsenic concentrations, estimated arsenic concentrations, and geographic location. Efforts are underway to assess sensitivity of arsenic exposure estimates to different sources of uncertainty, and to incorporate uncertainty estimates in spatio-temporal visualization tools.

## 6 Conclusions

Despite the common perception that cancers are often caused by environmental contaminants, limited evidence exists to support widespread associations. Reports of weak, non-significant associations between environmental agents and cancer (Gammon et al. 2003; Bates et al. 1995; Steinmaus et al. 2003) may be attributed to exposure assessments that inadequately incorporate temporal variability. Exposure assessors often under-emphasize temporal variation and construct cumulative exposure estimates over pre-defined time intervals. Continuous exposure assessments are an improvement and their construction is aided by spatio-temporal visualization tools. Smooth, continuous space-time maps may be used: to assign exposure estimates through a spatial query procedure; and to reveal cancer clusters at any moment in time. Similarly, smooth, continuous, temporally-variant histograms and scatter plots allow for examining relationships between variables at any moment in time. Exposure life-lines, a type of temporally-explicit exposure profile, may convey critical information regarding timing of exposure and disease. Spatio-temporal visualization tools facilitate evaluation of spatially- and temporally-explicit exposures and ensuing disease development.



## References

- Andrienko, N, Andrienko G, Gatalsky P (2003) Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing* 14:503–541
- Bates MN, Smith AH, Cantor KP (1995) Case-Control Study of Bladder-Cancer and Arsenic in Drinking-Water. *American Journal of Epidemiology* 141:523–530
- Beaubroef T, Breckenridge J (2000) Real-world issues and applications for real-time geographic information systems (RT-GIS). *Journal of Navigation* 53:124–131
- Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer P, Gehring U, Heinrich J, Cyrys J, Bellander T, Lewne M, Brunekreef B (2003) Estimating Long-Term Average Particulate Air Pollution Concentrations: Application of Traffic Indicators and Geographic Information Systems. *Epidemiology* 14:228–239
- Cantor KP (2001) Invited Commentary: Arsenic and Cancer of the Urinary Tract. *American Journal of Epidemiology* 153:419–421
- Charles LE, Loomis D, Shy CM, Newman B, Millikan R, Nylander-French LA, Couper D (2003) Electromagnetic fields, polychlorinated biphenyls, and prostate cancer mortality in electric utility workers. *American Journal of Epidemiology* 157:683–691
- Chittaro L, Combi C, Trapasso G (2003) Data mining on temporal data: a visual approach and its clinical application to hemodialysis. *Journal of Visual Languages and Computing* 14:591–620
- Dragicevic S, Marceau DJ (2000) A fuzzy set approach for modeling time in GIS. *International Journal of Geographic Information Science* 14:225–245
- Gammon MD, Wolff MS, Neugut AI, Eng SM, Teitelbaum SL, Britton JA, Terry MB, Levin B, Stellman SD, Kabat GC, Hatch M, Senie R, Berkowitz G, Bradlow HL, Garbowski G, Maffeo C, Montalvan P, Kemeny M, Citron M, Schanbel F, Schuss A, Hajdu S, Vinceguerra V, Niguidula N, Ireland K, Santella RM (2002) Environmental Toxins and Breast Cancer on Long Island. II. Organochlorine Compound Levels in Blood. *Cancer Epidemiology Biomarkers & Prevention* 11:686–697
- Hornsby K, Egenhofer M (2002) Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence* 36:177–194
- Kim MJ, Nriagu J, Haack S (2002) Arsenic species and chemistry in groundwater of southeast Michigan. *Environmental Pollution* 120:379–390
- Kolker A, Haack SK, Cannon WF, Westjohn DB, Kim MJ, Nriagu J, Woodruff LG (2003) *Arsenic in southeastern Michigan*. In: Welch AH, Stollenwerk KG (eds) *Arsenic in Groundwater: Geochemistry and Occurrence*. Kluwer Academic Publishers, Norwell, Massachusetts, pp 281–294
- Kulldorff M, Athas WF, Feuer EJ, Miller BA, Key CR (1998) Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health* 88:1377–1380
- Kwan MP (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research C* 8:185–203
- Maantay J (2002) Mapping Environmental Injustices: Pitfalls and Potential of Geographic Information Systems in Assessing Environmental Health and Equity. *Environmental Health Perspectives* 110:161–171
- Manufacturer Publishing Co. (1946, 1953, 1960, 1969, 1977, 1982) *The Directory of Michigan Manufacturers: Michigan Manufacturer and Financial Record*. Detroit, Michigan.
- Meliker JR, Nriagu JO, Hammad AS, Savoie KL, Jamil H, Devries JM (2001) Spatial clustering of emergency department visits by asthmatic children in an urban area: south-western Detroit, Michigan. *Ambulatory Child Health* 7:297–312
- Nriagu JO, Lawson G, Wong HKT, Azcue JM (1993) A protocol for minimizing contamination in the analysis of trace metals in Great Lakes waters. *Journal of Great Lakes Research* 19:175–182
- Reif JS, Burch JB, Nuckols JR, Metzger L, Ellington D, Anger WK (2003) Neurobehavioral Effects of Exposure to Trichloroethylene Through a Municipal Water Supply. *Environmental Research* 93:248–258

- Ryan PB, Huet N, MacIntosh DL (2000) Longitudinal investigation of exposure to arsenic, cadmium, and lead in drinking water. *Environmental Health Perspectives* 108:731–735
- Slotnick MJ, Meliker J, Nriagu J (2003) Natural sources of arsenic in Southeastern Michigan groundwater. *Journal de Physique IV* 107:1247–1250
- Steinmaus C, Yuan Y, Bates MN, Smith AH (2003) Case-Control study of bladder cancer and drinking water arsenic in the western United States. *American Journal of Epidemiology* 158:1193–1201
- Swartz CH, Rudel RA, Kachajian JR, Brody JG (2003) Historical Reconstruction of Wastewater and Land Use Impacts to Groundwater Used for Public Drinking Water: Exposure Assessment Using Chemical Data and GIS. *Journal of Exposure Analysis and Environmental Epidemiology* 13:403–416
- US Environmental Protection Agency (2000) Toxics Release Inventory. Washington, DC.