

GIS NEWS AND INFORMATION

December 1995 (No. 7)

Dedicated to CDC/ATSDR scientific excellence and advancement in disease control and prevention using GIS

Selected Contents: Dr. Anselin's course now showing (p.1); GIS Users help EIS officer Tom Hennessy (pp.3-4); Keeping up with GIS (pp.4-5); Census minority statistics and the undercount (pp.5-8).

I. Public Health GIS Training Opportunities

NCHS Cartography and GIS Guest Lecture Series

The half-day course "Spatial Data Analysis Using GIS", taught by Dr. Luc Anselin on September 21 at NCHS, was videotaped as a loanable resource for interested CDC/ATSDR staff. It has been distributed first to **Chet Moore** (Ft. Collins) and **Jay Kim** (NIOSH) and is currently available (through December) for your viewing pleasure. Next (January-February), it will go to **Shelba Whaley** (Chamblee) and **Peter Kilmarx** (Corporate Square). Then (March-April), it will be sent to **Louise House** (Lexington, MA) and **Lisa Bulkow** (Anchorage). If you wish to view the tape, please call any of the above contacts in your respective location for arrangements. If you are located elsewhere and can serve to facilitate a viewing at your site, please let me know.

II. News from CDC/ATSDR GIS USERS

(Please communicate directly with colleagues on any issues)

A. General News

1. From **Chet Moore**: By the way, our GIS workshop in connection with the Society for Vector Ecology meeting [October] went very well. We anticipated about a dozen people (given that it was all-day and on Sunday), but we wound up with 36! Some last-minute scrambling for laptops gave us enough machines to handle everybody, but it was a bit like a 3-ring circus. Quite a few state and local health departments and mosquito control programs are beginning to get into GIS for their programs.

2. Inquiry from **Susan Schechter** (NCHS): Are

any GIS Users aware of any GIS statistics that show interracial or mixed race households by geographic area? I checked with Vital Statistics and others at NCHS but no one seems to be aware of any. Thought I'd check with you. Thanks, Susan [Editor: See Dr. Robinson's related comments, III.C. below]

3. From **Mike Mungiole** (picked up from "To: Multiple recipients of list <gis-l@urisa.org>") Subject: Re: GIS and medical geography: Phillip (Phillip@news.cais.com) wrote:"Does anyone know of epidemiological studies utilizing GIS? How about anything close? I am doing a study of how GIS has been used to track, predict, and otherwise study disease patterns. Any directions would be helpful."

Since my response to Phillip bounced back and this may be of interest to others, I'm posting for reference by all. "GIS for Health and the Environment", proceedings of an International Workshop held in Colombo, Sri Lanka, 5-10 Sept. 1994. You can obtain a copy from International Development Research Centre, P.O. Box 8500, Ottawa, ON, Canada K1G 3H9. Contents include:

- *The Present State of GIS and Future Trends.
- *Geographical Information Systems (GIS) from a Health Perspective.
- *Spatial and Temporal Analysis of Epidemiological Data.
- *Towards a Rural Information System.
- *A GIS Approach to Determination of Catchment Populations Around Local Health Facilities in Developing Countries.
- *GIS Management Tools for the Control of

Tropical Diseases: Applications in Botswana, Senegal, and Morocco.

*The Use of Low-Cost Remote Sensing and GIS for Identifying and Monitoring the Environmental Factors Associated with Vector-Borne Disease Transmission.

*Geographical Information Systems for the Study and Control of Malaria.

*Spatial Analysis of Malaria Risk in an Endemic Region of Sri Lanka.

*Diagnostic Features of Malaria Transmission in Nadiad Using Remote Sensing and GIS.

*Monitoring Zoonotic Cutaneous Leishmaniasis with GIS.

*Use of RAISON for Rural Drinking Water Sources Management.

*Interests, Problems, and Needs of GIS Users in Health: Results of a Small Survey.

*GIS, Health, and Epidemiology: An Annotated Resource Guide.

Also, in the journal *Emerging Infectious Diseases*, Vol. 1, No. 4, page 156-7, there is a brief article entitled "Building a Geographic Information System (GIS) Public Health Infrastructure for Research and Control of Tropical Diseases". I am very interested in this topic. I'd like to hear what others may be doing in this area. Ric Skinner, wskinner@fast.net.

4. From **Bob Bernstein**: I want to inform you about two very powerful, and potentially very useful public health-oriented softwares that have been created through research, development, and field-testing (in collaboration with several US states and developing countries) by Professor Steven Seitz, Director of the Computational Modelling Laboratory at the University of Illinois. Prof. Seitz can be reached by phone/Fax at 1-217-244-5043 / 244-5712 and email at stseitz@vmd.cso.uiuc.edu.

I was VERY impressed with the quality, creativity, user-friendliness, flexibility, and

prodigious power (both descriptive and analytical) of the Geographic Information Management System ("GIMS") and the Population Dynamics Modelling system ("MINERVA") softwares -- what makes them especially attractive is that they can be dynamically linked with each other and with standard database and analytical softwares like EpiInfo, SAS, SPSS, dBASE, etc. The GIMS and MINERVA softwares can be used on laptop or notebook computers that are capable of running OS2; yet, MINERVA is apparently capable of replicating or even improving on the accuracy and reliability of dynamic population modelling that the US Bureau of Census can only accomplish on a mainframe computer.

The longitude-latitude "boundary" files which have been acquired and the GIMS software are capable of resolving US county and national polygons down to 500 feet (including islands) -- generally a level of resolution much finer than that permitted by the accuracy and reliability of available geographically linked data. The Computational Modelling Lab can provide technical support to potential users in order to prepare (or convert existing) longitude-latitude boundary files for use with GIMS -- at present, only US boundary files are included below the national level. I would encourage anyone to open a dialogue with Prof. Seitz, perhaps inviting him to make a presentation of these softwares.

5. From **Tom Arner**: I am looking for someone who might have a Canon BJ 600 color printer that I can use for a few minutes to test an enhancement to Epi Map 2.

B. Technical News

1. From **Linda Pickle**, NCHS: Kriging can be implemented in S+ by using functions (in both Fortran and S+) supplied by Brian Ripley. These can be obtained free from the StatLib resource on

Internet at Carnegie Mellon. [Editor: Also, *S+Spatial Stats* is a new software package for analyzing spatial data, available with the release of ARC/INFO 7.0. It includes the following features-kriging, variogram visualization and estimation, spatial smoothing, hexagonal bin plots, point maps, spatial correlation measures, spatial regression models, spatial randomness tests, Ripley's K functions, simulation methods and nearest neighbor routines].

2. From **Morris Maslia**: You may wish to pass this along to those interested in GIS and applications to exposure assessment and/or exposure-dose reconstruction. ATSDR's Exposure-Dose Reconstruction Project (EDRP) now has a "Home Page" on the World Wide Web (WWW) and can be accessed at the following address: http://www.ce.gatech.edu/WaterResources/water_proj.html

It also can be reached by accessing the CE department "Home Page" through GA TECH as follows: <http://www.ce.gatech.edu>, then select RESEARCH, then select Environmental Hydraulics and Water Resources, then select Exposure-Dose Reconstruction Project. The EDRP "Home Page" also has a direct link back to the ATSDR "Home Page".

III. Other Items of related Interest

A. GIS Users Respond

The following request was received from **Tom Hennessy**, EIS officer, and distributed to CDC/ATSDR GIS Users. I wanted to share responses from our group which are also included below. Special thanks are extended to **Andy Dean**, **Marcia Taylor**, **Kevin Liske** and **Allen Hightower** for their kind advise.

Request

From **Tom Hennessy**: Hi. I'm an EIS officer in Minnesota where I work in infectious diseases. I was given your name by the respiratory branch as someone who could possibly assist me. Recently, we investigated an outbreak of Legionnaire's disease in a small town and implicated the hospital cooling tower. However, one competing hypothesis I'd like to test involves comparing the residence of cases vs. controls with respect to distance and direction from a different cooling tower. Essentially, this second tower was suspect because of its maintenance history and because it was in a position to disperse water droplets via the prevailing winds. I'd like to do what I'm referring to as a 'vector analysis' using distance and angle from the suspect cooling tower along with wind rose data to compare the location of the homes of cases and controls. Do you know of a way to do this? Would you be willing to offer some advice? If I haven't made myself clear and you have some suggestions as to where I could begin you could call me at (612) 623-5414. Although I'm in Minnesota now, I will be in Atlanta next week and could meet with you to show you the maps we've created, if you're interested.

Responses

From **Andy Dean**: I don't know of a way to do this mathematically or graphically because of the vagaries of winds, obstructions, particle size, and other unmappable objects, but I once did a quantitative assessment of dose from a point source experimentally in an outbreak of histoplasmosis originating in a courthouse tower. We waited until the wind was exactly the same direction and velocity as on the day of infection, and generated an aerosol of cooking oil, using a microdroplet generator that insect spray people provided. We placed oil-droplet sensing cards throughout the courthouse. The cards were collected and droplets counted in a dissecting microscope to give relative counts of the "dose" (in this case of bird/bat guano) from the buckets that

were dumped from the roof during a cleaning operation. The correlation with infection was only so-so, but perhaps because many occupants of the courthouse were already immune to histo at the time of the epidemic. If you would like to be the second to employ this splendid technique, a description is found in Am. J. of Epi 108:36-46, 1978.

From **Marcia Taylor**: In my former life as an Army officer I had some responsibilities for monitoring Nuclear, Biological and Chemical hazards. It's been over 10 years since I had training in dispersion of chemical agents, but I am sure the US Army Chemical School located in Fort McClelland, Alabama (near Anniston) (unless that base has already been closed) has lots of experts on measuring the effects of droplets across vectors under varying wind conditions. Fort McClelland is within 2 hours' drive of here and you may suggest this epidemiologist contact someone there. I am sure he could find someone to help him.

From **Kevin Liske**: Basically, it sounds like all he needs to do is figure out his prevailing wind direction, give himself some number of degrees to allow for dispersal and orient to the PWD, and see where the arc falls in relation to the documented incidents. If he really wants to get technical, he could calculate the distance decay of the water droplets and other factors such as thermal updrafts (lakes) in relation to the time of day. I'd be willing to talk to him about it if he hasn't gotten it figured out before he comes to Atlanta (639-6066 x 71).

From **Allen Hightower**: Tom - This is how I'd do it.

(1) Get a GIS package.

(2) Get the Census Tiger Files of the area where the cases are. Try the "Archive of Census Related Products", located at <ftp://ftp.ciesin.org/cd/pub/census> or the Lawrence Berkeley Labs at [www-http://cedr.lbl.gov/mpub/cdrom/install/](http://cedr.lbl.gov/mpub/cdrom/install/)

[contents.online.html](http://cedr.lbl.gov/mpub/cdrom/install/contents.online.html) FTP - <ftp://cedr.lbl.gov>. These are addresses for the data in Atlas GIS ASCII format. There are probably ArcInfo, ArcView, and MapInfo versions of these data/map files out there, too.

(3) Locate the cooling tower on the map, get the lat, lon, and, if you like and can, the altitude.

(4) Use the address matching feature of your favorite GIS product to get the lat, lon, and altitude (if relevant) of your cases and controls. Add them to your c/c database.

(5) Now you've got everything on an X-Y(-Z) plane. You can compute distances (and directions) from the households to the cooling towers in any number of ways and store the result in your c/c database.

(6) Now, analyze. Univariate: pooled t-test or the Wilcoxon rank-sum test (for unmatched studies) on distance vs. case/control. For multivariable analysis (logistic regression): add distance to the cooling tower as an independent variable.

(7) You can work direction into this analytic approach, too. I've probably skipped a step or two, but this is the general way to do this sort of thing. Good luck!

B. Keeping Up with GIS and Related Happenings

1. National Institute for the Environment (NIE) proposed: Representative James Saxton (R-NJ) was to have introduced an NIE bill by now. The rationale is that NOAA's Coastal Oceans Program, EPA's Environmental Monitoring and Assessment Program, and DOE's Bureau of Environmental Research would receive more support for environmental science (to improve environmental policy) than in their current agencies.

2. NOAA's Environmental Data Index (NEDI):

NEDI development is a key element of the National Information Infrastructure (NII), sort of a “yellow pages” (index) for environmental data. It is intended to provide a single place for one-stop access to environmental information, for the public, industry and academia, beginning with the federal sector. To begin, key database players will be USGS, EPA, DOE, NASA and NOAA. The prototype is under construction but several of the databases are available. For example, the USGS Water Resources Information Home Page is: ‘<http://hto.er.usgs.gov/public/nawdex.html>’ with selected on-line data.

3. EPA’s Spatial Data Library System (ESDLS): The purpose is to provide on-line access to EPA’s national spatial data (ARC/INFO format). The tabular side, “Envirofacts”, STF3A Block Group and P.L. 94-171 Block data, is available in Oracle database form. The ARC/INFO library, on the spatial side, is under development and will contain a variety of themes e.g., TIGER 90 Block and Block Group, TIGER 92, Envirofacts point coverages, Geographic Names, NPL site boundaries, 1:2M DLG (roads, hydrography, boundaries), and a variety of others.

C. Special Attachment: Selected Testimony on Minority Statistics and Census Undercount
 NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS, June 14, 1995
SUBCOMMITTEE ON HEALTH STATISTICS FOR MINORITY AND OTHER SPECIAL POPULATIONS, Hubert H. Humphrey Building

Steve Botman, NCHS: Mr. Botman noted the important distinction in the levels of population universes, between the state and national on the one hand and the below-state on the other. Adjusted population estimates/projection are potentially done for the former and not the latter. By adjusted population estimates we mean

population estimates for 1990 or subsequent years based on 1990 Census where the Census data have been adjusted for net-Census under- enumeration. There are three universes of population data: civilian non-institutionalized, civilian, and resident. Coverage in the decennial census has not been uniform, and it is lower for minorities than for the non-minority population. That is minorities are less likely to participate in the decennial Census. Several important decisions have been made in regard to the 1990 Census. The Secretary of Commerce ruled that official population estimates based on 1990 Census data would not be adjusted for net Census underenumeration. In 1992, the Director of the Bureau of the Census decided that the Bureau's official post-1992 population estimates and projections would not be adjusted for Census underenumeration. This may have been due to discomfort about using modeling for small geographic areas. The Director of the Bureau of the Census also decided that the sponsors of ongoing surveys conducted by the Bureau (e.g., the NHIS) would decide for themselves about using adjusted population estimates to calibrate survey estimates. Mr. Botman noted that changing the overall survey population estimates reduces the credibility of specific finely-tuned estimates such as those made on the labor force participation. Then in late 1993, it was decided that the Bureau of Labor Statistics could use adjusted population estimates for Current Population Survey calibration.

The Office of Management and Budget assembled most of the agencies together and determined that as of 1994, all of the ongoing surveys conducted by the Census Bureau would be calibrated to population estimates that are adjusted for net Census underenumeration.

Greg Robinson of the Bureau of the Census said the undercount in 1990 was about 4 million. Mr. Botman commented that the Bureau's refusal to release the controls (to be used for estimator calibration) for surveys it does not conduct is "a

point of contention" between the Bureau and the user community. He added that NCHS conducts several surveys (e.g., NHANES) that do not use the Bureau of the Census as the data collection agent. Dr. Robinson said the Bureau makes available the controls that go into the CPS at the national level and one number per state, but nothing at the county or city level.

Dr. Williams noted that the adjustments for the undercount factors are based on the Population Enumeration Survey. He asked about the impact of the undercount and subsequent decisions on statistics on minority populations. Mr. Botman said that the undercount does not produce a significant bias in sampling. The "nub of the issue" is the post-stratification used for calibrating surveys, which among other things addresses the survey's undercoverage of the population. One issue in poststratification is being able to partition the sample into subdomains that have relatively similar survey undercoverage rates. We do not know, for example, if persons with poor health are disproportionately undercovered in the Census or in surveys relative to the Census. For the Vital Statistics Program, however, NCHS does not use adjusted population estimates in calculating rates, in part because of the difficulty that would cause below the national level. Opinions differ about whether it is better to have comparable rates all way from the national level to the finest geographic level, or to have the best possible rate at the state level and above. For the calculation of rates, the Vital Statistics Program uses official population estimates as modified by the MARS.

Mr. Botman said some respondents do not provide all the information requested on race and ethnicity, and as a result, a significant amount of Census data are not suitable for collapsing into the four race categories (White, Black, American Indians and Alaskan Natives, and Asian and Pacific Islanders) specified by the Office of Management and Budget.

J. Gregory Robinson, Ph.D., Bureau of the Census: Dr. Robinson commented that these matters are being addressed, and a mixed race category will be tested in the national content test next year. The two programs used to measure coverage are demographic analysis, which uses historical aggregate data from the registration system and immigration estimates, and the Post-Enumeration Survey (PES). The latter involves micro-level case-by-case matching. The two have produced the same basic profile of undercount, but the PES does not have comparable numbers from 1980 or 1970.

Although the undercount has been reduced in the last 50 years, the differential undercount has not changed despite all the efforts to that end. He remarked that it is actually surprising that the differential has not increased, given the mobility of society and the influx of immigrant groups. In response to a question, he said that breakdowns of nonblack minorities cannot be obtained from demographic analysis with any assurance, because of problems of race classification and inconsistent reporting.

The largest undercounts from 1970 to the present are for black men ages 20 to 50. An average of 10 to 15 percent were missed--more in some geographic areas. One possible reason, Dr. Robinson said, is that people do not realize or believe that the Census is confidential. He has been working on sex ratios, which can be computed for states and counties, and has found that the differential between men and women persists across all counties in the U.S.

Unlike the demographic analysis, the PES provides information about Hispanics and Asians. It measured an undercount of only 6 percent for adult black men, in contrast with the 10 to 15 percent found by the demographic analysis, because some of the factors that apply to omissions from the Census have an effect on the re-interview. Dr. Williams asked about variations among subgroups, a particular interest of the Subcommittee, and Dr. Robinson said much more

information is needed on this.

The PES involved in-person interviews with a sample of some 150,000 households. As much as possible, it was an independent roster. The results were reconciled with the Census, yielding information on undercount, omissions, and duplications, and the net. Dr. Robinson said the 1980 and 1990 Censuses have greater duplication or erroneous enumerations than 1970 or earlier, possibly due to "over-zealousness" in the coverage improvement programs.

The PES served as a validation of the Census. Demographic analysis measured a 1.8 percent net undercount for the general population, and the PES, 1.6. The PES measured an over count in the population of 50 and over, which would have an impact on vital statistics.

For Hispanics, the demographic analysis made an allowance for undocumented persons. The undercount is estimated at 4.4 for blacks and 5.5 for Hispanics. For Asians/Pacific Islanders, the undercount was about 2.3 percent; Dr. Robinson noted that the undercount of Laotians and Cambodians could be much higher. The September, 1993 issue of the *Journal of the American Statistical Association* (JASA) has the PES results.

Asked about gender differences, Dr. Robinson said the gender differences cut across every group, with a greater undercount for men than for women.

The PES found an undercount of 4.5 percent for American Indians. He explained that the estimate for Indians living on the 10 largest reservations was 12.2 percent. It was not possible to develop a separate estimate for those spread out in cities off the reservation. Dr. Thompson observed that there was an over count of Indians in 1980, and a similar over count probably continues in 1990, assuming that the 1970 numbers were more accurate.

Dr. Robinson discussed the general difficulties with classification identification, which

cause many of these problems. Ten million persons, 4 percent of the population, did not check off either "white" or "black" or "American Indian" or "Chinese" (for example); they circled "Other race" and wrote in a category not recognized by OMB. Thus the numbers for "other race" were considerably larger than expected. When it is necessary to compare to other data, the 10 million people in this category are divided among the four race groups.

He discussed the reasoning used in assigning race in demographic analysis, traditionally based on the race of the father. There are increasing problems with consistency in the classification of multi-racial children, and the trend is moving away from the race of the father as the determinant. Dr. Robinson referred again to the JASA article, noting that Jeff Petzel had a good critique of demographic analysis. The author's conclusion is that it is becoming very difficult to make estimates by race with any accuracy.

In response to a question, he said the changes in inter-racial unions might affect the way race is imputed in the future. He remarked that the addition of a mixed race designation would "throw demographic analysis out the window in terms of clear-cut differential."

He then described his work to determine rough geographic indicators, through which he is finding distinct differentials. The demographic estimates are independent indicators of the differential undercounts. He noted that another problem with the PES is that below the national level, the numbers for states and counties are based on synthetic estimates. For the states with the larger net undercount (California, Texas, New Mexico, and Florida), there is confidence in the size of the differential undercount, but not for other states.

Dr. Robinson discussed the characteristics that underlie the differences in counts among groups. The PES found that one important determinant is whether a given residence is

occupied by an owner or a renter. There is essentially no undercount of owners, even for non-whites. He noted a compositional effect: a much greater proportion of minorities live in renter-occupied units than do non-Hispanic whites, and this contributes to the higher undercount of minorities. The undercount is disproportionately concentrated in certain geographic areas, such as big cities and certain areas of the country, and among certain demographic groups. The goal is to identify these areas in advance and then to use special enumeration tools for them.

The Bureau is therefore developing a targeting database. He referred to a paper discussing the underlying factors causing undercount, such as tenure, low educational attainment, language difficulty, and living in a complex household. The targeting database applies algorithms to Census data, scaling all tracts in the country. This will aid in identifying the areas

that require special methods, which might include sending a team of people into dangerous neighborhoods, supplying interviewers with a Spanish questionnaire, and doing special outreach and publicity.

Mr. Botman noted the evidence of a high response variability on race/ethnicity questions in surveys. In doing the calibration, NCHS assumes that people will answer survey questions in the same way they answer Census questions. He questioned whether this assumption is valid, and noted that the effort to "put people in boxes" raises serious conceptual issues. Dr. Williams expressed appreciation for the two presentations, and noted the Subcommittee's concern about the impact of these issues on health statistics for minority populations. The Subcommittee will continue to study this issue.

Chuck Croner, Editor, **GIS NEWS AND INFORMATION**
Office of Research and Methodology, National Center for Health Statistics
Centers for Disease Control and Prevention

Enjoy the passage of 1995...and stay in GIS touch