

# PUBLIC HEALTH GIS NEWS AND INFORMATION

November 1998 (No. 25)

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

**Selected Contents:** Conferences (p.1); News 9); Special Reports (pp. 9-11); GIS Lectures 18); Website(s) of Interest (pp.18-19); Final



from GIS Users (pp.1-7); GIS Outreach (pp.7-12); DHHS and NCVHS Update (pp.12-19); Thoughts (p. 19)

## I. Public Health GIS (and related) Events

\*\*\*\*\*

### **SPECIAL CDC/ATSDR GIS LECTURE:**

November 5, 1998, **“Protecting Privacy of Georeferenced Contextual Data,”** Alan Saalfeld, PhD, Assistant Professor, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University. sponsored by the NCHS Cartography and GIS Guest Lecture Series, and CDC’s Behavioral and Social Science Working Group, 2:00-3:15 P.M., NCHS Auditorium, Hyattsville, MD [see extended abstract this edition]

☞ GIS Applications in Water, Wastewater and Stormwater Systems Seminar, American Society of Civil Engineers, January 8, 1999, Atlanta, GA [Contact: see ASCE website at <http://www.asce.org>]

☞ 2nd Annual National Birth Defects Prevention Network Workshop, January 26th and 27th, 1999 Washington, D.C. [Contact: Marcia Feldkamp at voice (801) 538-6953 or Linda Mitchell (CDC) at (770) 488-7176]

☞ 7th Biennial Symposium on Minorities, the Medically Underserved and Cancer, February 27 - March 3, Washington, DC [Contact: Carlotta Hancock, (713) 798-5383]

☞ Coastal GeoTools '99 conference, the NOAA Coastal Services Center, April 5-7, 1999, Charleston, SC [Contact: Steve Meador at voice (843)740-1334 or email [smeador@csc.noaa.gov](mailto:smeador@csc.noaa.gov) or visit [www.csc.noaa.gov/GeoTools99](http://www.csc.noaa.gov/GeoTools99)]

☞ Symposium 99: Workshop and Symposium on

Combining Data from Different Sources, Statistics Canada May 4-7, 1999, Ottawa, Canada [Contact: Christian Thibault at voice (613) 951-6935 or visit <http://www.statcan.ca/english/conferences/symposium99/index.htm>]

☞ 1999 ASPRS Annual Conference, “From Image to Information,” American Society for Photogrammetry and Remote Sensing, May 17-19, 1999, Portland, OR [Visit: <http://www.asprs-portland99.com>]

### **SPECIAL CDC/ATSDR GIS LECTURE**

December 16, 1998, **“Towards a Working Atlas of Model GIS Maps for State and Local Public Health Planning,”** Charles Croner, PhD, National Center for Health Statistics and Thomas Richards, MD, Public Health Practice Program Office, CDC, sponsored by the NCHS Cartography and GIS Guest Lecture Series, and CDC’s Behavioral and Social Science Working Group, 2:00-3:15 P.M., NCHS Auditorium, Hyattsville, MD. CDC/ATSDR staff please make Envision arrangements now. [See abstract in this edition]

\*\*\*\*\*

## II. News from GIS USERS

(Please communicate directly with colleagues on any issues)

### **A. General News (and Training Opportunities)**

1. From **Jay Kim**, NIOSH CDC: Jay H. Kim of Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health has prepared an **Atlas of Respiratory Disease Mortality, United States: 1982-1993**. This publication includes maps by health service area that displays age-adjusted death rates and rate ratios from 17 respiratory diseases that are mostly attributable to hazardous occupational

exposure. Thus, for each of the disease categories, occupational causes have been documented. Using Atlas GIS, Jay has prepared 34 maps based on mortality data from NCHS. The final draft has recently been sent to a printer.

2. From **Cathy Cubbin**, former NCHS CDC Intern: "Examining Spatial Correlates of Mortality Using GIS" will be presented by Catherine Cubbin and Linda Pickle at the Computer Theater session at the annual APHA meetings in November. **Abstract.** The National Center for Health Statistics has produced an **Atlas of United States Mortality** [Editor: see <http://www.cdc.gov/nchswww/products/pubs/pubd/other/atlas/atlas.htm>] which includes maps of rates for 18 leading causes of death in the United States for the period 1988-92. Mortality data from the new Atlas are being layered with data representing demographic, socioeconomic, environmental, and behavioral risk factors using ArcView software to create a geographic information system (GIS) that will allow researchers to explore correlations between the patterns of mortality and those of suspected risk factors. The usefulness of mapping in public health research for exploratory analysis will be discussed, maps with data at several layers will be illustrated, and the new GIS system with a user-friendly interface will be demonstrated using examples from the new Atlas. In addition, the mixed effects statistical model from the Atlas is being extended by adding risk factors for cause-specific mortality. Examples from the analysis of homicide in black men will be used to demonstrate how GIS can be used to complement statistical modeling for exploratory analysis of homicide and for model checking. [Contact Linda at voice (301) 436-7904, ext. 148 or email [lwp0@cdc.gov](mailto:lwp0@cdc.gov)]

3. From **Lisa Warnecke**, Consultant, National Academy of Public Administration (Announcement: Study of Limitations on Collecting and Using Disaster Information): Here's an announcement for a project that might be of interest, and we are still gathering input. You may also wish to point folks to the NAPA GI study that was completed earlier this year ([www.napawash.org](http://www.napawash.org)).

**Description of the Study.** In November 1997, a high level federal Disaster Information Task Force submitted to Vice President Al Gore a report entitled *Harnessing Information and Technology for Disaster Management*. That report recommended creating a national electronic disaster information network that could become a component of a worldwide Global Disaster Information Network (GDIN). The vision for GDIN is "a robust, interactive knowledge base of disaster-related information accessible to disaster managers throughout the United States" and other affected parties. Although the feasibility study wove together the many threads necessary to transform this vision into reality, it left many details to be worked out. Under a cooperative agreement with the U.S. Geological Survey, the Academy study will begin to tackle one such detail—the limitations on collecting, using and providing access to the many different types of data needed in all phases of disaster management. These limitations arise from such factors as copyrights, privacy protections, liability for damages arising from inaccurate data, and security. These concerns are growing because public agencies are using more commercial data, and electronic data systems have increasing potentials to combine data into dossiers that can endanger privacy. For such reasons, it may not be wise to assume that all the data needed for disaster management can or should be made readily available over the Internet. The Academy study will address such data access issues in a preliminary fashion.

**Importance of the Study.** Global disaster costs are very high and rising. The lives lost and property destroyed are in the hundreds of thousands of people and hundreds of billions of dollars. The technology to reduce these losses is available. The Academy's study would help to harness information technology to this task.

**Duration of the Study.** This preliminary six-month study will provide an initial consideration of these tough data access issues. It will scope-out the nature of the issues and indicate where further work is most needed. During the study, it is expected that the Academy's panel would meet twice. The meetings will last one day each, and the panel will set the specific dates of those meetings.

The Panel. The panel is now being formed. We expect it to have 6-7 members with outstanding qualifications to consider the interrelationship between disaster information needs and the data access policies at all levels of government.

The Study Approach. This study will examine (1) the types of information needed in the disaster management field, and where they come from, (2) existing laws and regulations governing data access at the state, local, and federal levels of government, and (3) foreseeable trends in information needs and the evolution of data-access law. Draft research findings and issue papers will be reviewed at an invitational conference to collect a wide range of views on this subject prior to consideration of the study report by the Academy's panel. The study report is expected to include preliminary findings, conclusions, and recommendations, including needs for further study. [Contact: Lisa at voice (315) 478-6024 or email lwarnecke@usa.net]

4. From **Peggy Lassiter**, Howard University (The Mississippi Delta Project: Education and Preparation of Nurses addresses the environmental risks in the Mississippi Delta Region (MDR) through train-the-trainer faculty workshops): This project is a response, in part, to Executive Order 12898 (February 11, 1994), which outlines federal action to address environmental justice in minority and low-income populations. Among the areas of the country most threatened by environmental hazards are the **219 counties within the seven states comprising the Mississippi Delta: Arkansas, Illinois, Kentucky, Louisiana, Missouri, and Tennessee.**

In 1994 Howard University Division of Nursing entered into an agreement with the Minority Health Professions Foundation, the Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry to respond to concerns about environmental pollution and its effects on health of Delta residents and to enhance the role that health professionals play in addressing these concerns. The nursing initiative was a spearhead to increase a focus on environmental health with particular attention to the MDR.

Two workshops for faculty in schools of nursing were completed in 1998. The first, for individuals in Mississippi and Louisiana, was held at Alcorn State University School of Nursing; Natchez, Mississippi in May. The second, for faculty in Tennessee and Arkansas, was conducted at University of Tennessee, Memphis College of Nursing in September. A successful learning activity utilized in both settings focused on gathering locality specific environmental information through the Environmental Protection Agency Internet site. Participants stated that EPA databases reviewed can be very helpful to health care workers and community residents as they identify environmental problems, work with a variety of resources to analyze related factors, and act to protect community health or find solutions for hazardous exposures.

Additional faculty and community nurses' workshops are planned for 1999. Questions or information about the project may be directed to the following email address: Peggy G. Lassiter, MSN, RN, Coordinator, The Mississippi Delta Project, Howard University, College of Pharmacy, Nursing, and Allied Health Sciences

5. From **Bill Henriques**, ATSDR (You might want to announce the availability of this data set in the newsletter - these folks have done a fine job collecting African spatial data sets and they certainly could be of great use of health professionals doing work in African countries): The **Spatial Characterization Tool (SCT) v 1.0** for Africa was released for general distribution in June of this year. The application was developed by the Characterization and Assessment Applications Group (CAAG) of the Blackland Research Center, a part of the Texas A&M University System. Written for use within the ARC/INFO Geographic Information System (GIS) the SCT was designed to facilitate flexible and rapid exploration and analysis of environmental and demographic data for the African continent.

The SCT is now in use at a large number of agricultural, environmental and medical research centers and university departments around the world. The authors of the SCT are interested in determining

the uses to which the application has been put since release in order to assist in planning future development. Descriptions of the projects in which the SCT has been used as well as comments on design and further analytical capabilities that would be useful are eagerly sought from the SCT user group. Users interested in collaborating in the development of specific new applications for the SCT are also invited to respond.

With the permission of each user who sends comments to the authors, these descriptions and comments will be made available to the rest of the SCT user group through the CAAG web site at [www.brc.tamus.edu/char](http://www.brc.tamus.edu/char). A general description of the SCT as well as other work currently being undertaken by the CAAG can also be found at this site.

Descriptions of projects and comments can be emailed to [caag@brc.tamus.edu](mailto:caag@brc.tamus.edu). The authors -Corbett, J.D. and R.F. O'Brien, 1997, The Spatial Characterization Tool-Africa v 1.0. Texas Agricultural Experiment Station, Texas A&M University System, Blackland Research Center Report No. 97-03, CD-ROM Publication-would like to thank respondents in advance for their assistance. [Contact: Richard O'Brien, Characterization and Assessment Applications Group, Blackland Research Center, 808 E Blackland Rd, Temple TX 76502, at voice (254) 770 6636 or email [caag@brc.tamus.edu](mailto:caag@brc.tamus.edu)]

6. From TJ Mathews, NCHS CDC: 2<sup>nd</sup> Annual National Birth Defects Prevention Network (NBDPN) Workshop- Building the Foundation for the 21st Century. This year the workshop will be conducted at the Doubletree Hotel at Pentagon City, Arlington, VA on Tuesday, January 26<sup>th</sup> and Wednesday, January 27<sup>th</sup>. A host of guest speakers will present and share their expertise on various topics, including: Putting the Birth Defects Prevention Act into practice; Update on the Birth Defects Centers for Research and Prevention; Epidemiology of Cardiovascular Malformations; Gene-Environment Interactions; Geographic Information Systems, and; Creative uses of surveillance data. [Contact: See Conferences, p.1]

7. From Anita Burney, GIS Consultant: ESRI is now

accepting paper abstracts for the 1999 User Conference. We are seeking papers describing innovative uses of GIS in the Health industry, within Health Departments (Federal, State, and Local); Health Plans (Insurance Plans, etc.); and Integrated Delivery Systems (Hospitals, etc.). Guidelines for submitting abstracts can be found at <http://www.esri.com/events/uc/index.html>. Deadline for abstract submission is November 13. [Contact: Anita at email [awburney@erols.com](mailto:awburney@erols.com)]

8. From Doug Browne, NCIPC CDC (Census Looking for Part-Timers): The U.S. Census Bureau is hiring federal employees to work part-time in temporary, intermittent Census 2000 positions (typically as a census taker or crew leader). To apply, call toll free at 1-888-325-7733. [Editor: This is one way to get familiar with TIGER and Census data collection; Source: FEDweek October 21, 1998 Issue]

9. From Richard Wright, San Diego State University: The University Consortium on Geographic Information Science(UCGIS) is hosting a Summit at the GIS/LIS meetings in Fort Worth, TX on November 10, 3:30-6:30pm, in Rm E108 of the Convention Center where GIS/LIS sponsoring societies and other major GIS organizations will tell about their particular interests in GIS education. Issues include accreditation and certification, methods of delivering GIS education, the appropriate curriculum content for different constituencies, the roles of the universities and the private sector in GIS training, educational partnerships, and distance learning in GIS. Major GIS stakeholder organizations will make short overview presentations followed by an open forum and discussion. This is an open meeting for all GIS/LIS participants. [Contact: Richard email [wright16@mail.sdsu.edu](mailto:wright16@mail.sdsu.edu)].

10. From Neil Andersson, CIETcanada: I wondered if anyone has experience and points of view of the relative merits of raster-based vs vector-based modeling for epidemiological data. CIETcanada ([www.ciet.org](http://www.ciet.org)) is introducing GIS into local public health planning in five pilot regions in Canada and

would benefit from some advice on this. [Contact: Neil at voice 1 613 241 2081 or email CIETinter@compuserve.com]

11. From Joseph Francica, Intergraph Corporation: I invite all health care professionals needing information regarding Intergraph's complete line of GIS products to visit our web site, sign up for a free subscription to our monthly corporate email-newsletter on GIS technology and receive a free evaluation copy of GeoMedia 2.0. See our web site at <http://www.intergraph.com/gis/business>. [Contact: Joseph at voice (888)343-5072 or email jrfranci@ingr.com]

12. From Cynthia Warrick, Howard University: Please visit URL [http://www.con-ed.howard.edu/gis\\_newsletter.htm](http://www.con-ed.howard.edu/gis_newsletter.htm) for the Urban Environment Institute (UEI) Newsletter on the recently convened 1998 Summer GIS workshop. Because the Summer Faculty workshop is earlier this year; we request all funding commitments be sent by February 1999. The Advanced workshop is May 15-22, 1999 and the Introductory workshop is June 14-19, 1999. If you need any additional information, please let me know. Thanks for your past support and involvement. [Contact: Cynthia at voice (301) 585-2295 or email cwarrick@con-ed.howard.edu]

### **B. Technical News**

13. From Martin Kulldorff, University of Connecticut Medical School (Spatial Statistics Software): SaTScan is a statistical software package for analyzing geographical health data using the spatial and space-time scan statistics. It is designed for any of the following interrelated purposes: (1) to evaluate reported spatial or space-time disease clusters, to see if they are statistically significant, (2) to test whether a disease is randomly distributed over space or over time or over space and time, and (3) to perform geographical surveillance of disease, to detect areas of significantly high or low rates.

SaTScan analyzes the data using either a Poisson model, based on the observed and expected number of events in different areas, or a Bernoulli

model, with 0/1 event data such as cases and controls. It is possible to adjust for various covariates provided by the user, such as age or gender. Developed at the National Cancer Institute, SaTScan is in the public domain, and version 2.1 was released in August 1998. It runs under Windows 95/NT. It is easy to install and use, and can be downloaded free of charge from the world wide web at '<http://dcp.nci.nih.gov/BB/SaTScan.html>'. [Contact: Martin may be reached at the Division of Biostatistics, Department of Community Medicine and Health Care, at voice (860) 679-1263]

14. From Jennifer Zoerkler, ESRI (S-Plus Road Shows): ESRI will be participating with MathSoft at the December 8, event at NIH in Rockville. We will be demonstrating the use of the S-Plus Extension for ArcView. Our plan at the moment is to show the integration of crime analysis data with public health issues. This seminar introduces the fundamental concepts and tools you need to perform spatial analysis of geographic data. Some 80% of data are spatially related and powerful new software techniques allow you to explore relationships and anomalies in your data. We will look at ways to explore your data, identifying significant trends, finding spatial relationships, and ferreting out potential anomalies in the data. You will learn when to use visual and statistical analysis techniques and why classical statistical techniques are often inappropriate for spatially related data.

Quantitative spatial data analysis involves the exploration, visualization and modeling of spatial data. To work efficiently, the geographic data analyst needs an environment that integrates data storage and retrieval, cartographic rendering, data visualization and quantitative spatial modeling. Using ArcView, S-PLUS, S-PLUS for ArcView and S+SpatialStats we demonstrate the methodology of quantitative spatial data analysis. For more information, see <http://www.mathsoft.com/splus/spress/roadshow/index.htm>. [Contact: Jennifer at voice (703) 506-9515, Ext. 8055 or email [jzoerkler@esri.com](mailto:jzoerkler@esri.com)]

15. From Thurman Wenzl, (list EPIDEMIO-L- Does GAM Work?): The International Agency for Research

on Cancer (IARC) study of clustering methods conducted a study in 1989-91 of all available different clustering methods, many developed by the early critics of the Geographic Area Machine (GAM). Fifty synthetic cancer data sets were created for which the degree of clustering and locations of clusters were known but kept secret. These data were given to the participants who performed their analyzes without any knowledge of the correct results. A wide range of methods were applied by users who had no knowledge of the cluster structure of the synthetic data. There were: 1.Pothoff-Whittingham Method; 2.Cuzick-Edwards two sample method; 3.GAM-K; 4.Besag-Newell's method; and 5.ISD's Original Method. This list was later extended to include four others but these were applied with knowledge of the cluster locations and of the results generated by the blind study. These are: 6.ISD revised method; 7.Cuzick-Edwards one sample method; 8.Diggle-Morris K functions; and 9.the CAS method. [For the results, which may surprise you, see: <http://www.ccg.leeds.ac.uk/smart/gam/gam8.html>]

### C. Internet News

16. From Dawn Wright, Oregon State University (UCGIS Virtual Seminar): I wanted to announce that the UCGIS Virtual Seminar has been a smashing success since it's beginning in early September. For those of you who haven't heard of this, the UCGIS Virtual Seminar is a web-based, graduate-level seminar in geographic information science, comprised of students from UCGIS member institutions. Thus far, under the able leadership of Art Getis, Lynn Usery, Nina Lam, Bob McMaster, and Dawn Wright, registered students from the U. of Georgia, LSU, Oregon State, the U. of Minnesota, and the U. of Maryland, have discussed the following UCGIS research priorities: Spatial Analysis in a GIS Environment, Geographic Representations and Distributed Computing. We will take a break the second week of November to convene a special panel at GIS/LIS, which will give some of the students an opportunity to meet each other face-to-face and to discuss how the course might be improved over the second half. After that we will reconvene and discuss:

GIS & Society, and Scale.

The discussions by the students have been so informative and compelling that we strongly urge you to "drop in" to our seminar for a quick look; perhaps stay a while. So please come and join us for a while. A special account has been set up for visitors to our course. Point your web browser to: <http://forums.library.orst.edu/forums/Index.cfm?CFApp=7> and then log in with Username: virtual; Password: visitor. Some background information on the course may be found at: <http://dusk.geo.orst.edu/virtual> [Contact: Dawn at voice (541) 737-1229 or email dawn @dusk.geo.orst.edu]

17. From Steve Matthews, Pennsylvania State University: The Center for Electronic Records, (U.S.) National Archives and Records Administration (NARA) has updated the 'Title List: A Preliminary and Partial Listing of the Data Files in the National Archives and Records Administration.' The 'Title List' is current as of September 25, 1998, and now has entries for more than 14,000 of the over 100,000 electronic records files in the custody of NARA. The 'Title List' is available via the NARA homepage. The general URL is <http://www.nara.gov/> with information about the electronic records program and holdings available at <http://www.nara.gov/nara/electronic>. The direct link to the Title List files is <http://www.nara.gov/nara/electronic/tlintro.html>. Title List files are also available via FTP. Anonymous FTP (password 'guest') to FTP. CU.NIH.GOV, directory NARA\_ELECTRONIC. A READ.ME file on the FTP site provides further information about the 'Title List' and Title List extract files. Note that the full 'Title List' file has 23,246 lines and is approximately 1.5 megabytes in size.

Among the new entries in the 'Title List' are: 1) Section III, Department of Commerce, Records of the Bureau of the Census (R.G. 29): Decennial Census of Population and Housing, 1990: Summary Tape File 4A (858 data files). 2) Section VI, Department of Health and Human Services, Records of the Centers for Disease Control (R.G. 442): National Hospital Discharge Survey, Multi-Year Files (Newborns and Not-Newborns), 1979-1992. 3) Section VI,

Department of Health and Human Services, Records of the Health Resources and Services Administration (R.G. 512): Area Resources File, electronic Record Layouts and Users Guides, 1985-1991. [Source: Steve at email matthews@pop.psu.edu]

### III. GIS Outreach

(Editor: All solutions are welcome and will appear in the next edition; please note that the use of trade names and commercial sources that may appear in *Public Health GIS News and Information* is for identification only and does not imply endorsement by CDC or ATSDR)

✉ From Tom Richards, PHPPPO, CDC: I am interested in learning about how GIS might be applied by state and local health departments to prevent (or respond to) attacks of bioterrorism. Has anyone done any work in this area? Alternatively, does anyone have any thoughts as to "where to start" if they were to try to apply GIS in this area? If so, please suggest references or contact persons so I could follow up to learn more details. Thanks. [Contact: Tom at voice (770) 488-2544 or email tbr1@cdc.gov]

✉ From Vic D'Angiolo, Canada Food Inspection Agency: I am a GIS Specialist with the National Centre for Foreign Animal Disease at the Canada Food Inspection Agency in Winnipeg, Manitoba. My role is to provide GIS support to scientists involved in Foreign Animal Disease Surveillance here at the centre. We currently use ARCVIEW GIS for data storage and analysis. As well I have access to ARC/INFO for more advanced spatial analysis. I would be very interested in joining this GIS users group and communicating with my counterparts in this field. [Vic can be reached at vd'angiolo@em.agr.ca]

✉ From Joanna McKenzie, Institute of Veterinary, Animal & Biomedical Sciences, Massey University: I'm doing a study to identify geographic and vegetation features of farms that are associated with a measure of TB in cattle on farms in a particular region of New Zealand. Let me give you a little background to put my question in context. Possums are acting as a wildlife vector of TB in NZ, spreading the disease to farmed animals. TB is very clustered in possums both spatially and temporally. Spatial clusters appear to be

associated with nesting areas in habitat that favours the spread of the disease amongst possums. Cattle become infected by sniffing at possums that are in the terminal stages of TB when they become weak and disoriented and start wandering around on the pasture during the day instead of being tucked up in their nest. What I want to find is a set of geographic features that can predict farms that are likely to have a higher number of possum TB hot spots (the colloquial term we use for the locations of possum TB clusters - scale only 50-100 meters). As we do not have an accurate measure of the incidence of TB in possums on every farm, we have to use the incidence of TB in cattle as an indirect measure of TB in the possums on the farm.

I'm using a GIS to measure the vegetation and geographic factors on farms e.g. hectares of forest, pasture, shrubs, etc, heterogeneity measures, etc, distance from National Park, length of rivers, etc, etc. I'm then modeling these using logistic regression. The problem that I'm putting to the list is finding the most appropriate outcome variable for my model. I have various measures of incidence of TB in cattle e.g. average annual incidence, maximum incidence, 5-year cumulative incidence. Most years only small numbers of cattle become infected i.e. one or two per herd. Then in farms that have resident TB hot spots they will have a certain number of years when they get a larger number of infected cattle, up to 20+. As the farms are of varying size with cattle numbers ranging from 14-1200 the impact of one positive on the resulting incidence is quite variable. Generally the larger farms have a worse possum-TB problem (more cover for nest sites, etc) but they have lower incidence rates because of their large cattle numbers. A similar number of TB-positive cattle on smaller farms gives considerably higher incidence rates. All cattle are not equally exposed to the disease because of its spatial nature. So the rates are influenced by the variable population size on farms.

We've tried adjusting the rates using Empirical Bayes estimation, but these do not make too much difference to the results. They do pull down estimates for the smaller herds that have a high incidence, but they also bring the rates up (to a lesser extent) for the smaller herds with low or no TB incidence. It seems to

me the best way of dealing with it would be to project the data onto geographic areas that are a similar size, such as described in Bailey and Gattrell's text, Interactive spatial data analysis. I have a problem here that I don't have data for every farm within my study area hence I have gaps for which I don't have TB data. This could lead to anomalies in the number of positives in each new spatial area. I have thought of amalgamating smaller farms that are contiguous but this would be very tedious. So, I'm really keen to hear your ideas about the best way to find a measure of TB in cattle that is not influenced to such a degree by the size of the underlying population. This has turned into rather a lengthy description, but I felt it was useful to put the problem in context. Thanks. [Joanna can be reached at voice (06) 350-4008 (office) or email J.S.McKenzie@massey.ac.nz]

☞ From Ric Skinner, GIS Consultant: I am interested in hearing from anyone who is using, or considering the use of GIS, in conjunction with Cancer Registry data. As we try to manage large cancer patient databases and seek to learn more about the causes of various cancers, employing GIS seems to be a natural application whereby cancer incidence might be related to environmental or demographic factors. I am particularly interested in knowing what Cancer Registry software is used, how it is linked to GIS, what GIS software is used, how patient confidentiality is addressed (e.g., geocode patient address for point location, then aggregate upwards to block group, census tract, zip+4, etc.). One specific area of interest is how to successfully (=high match rates) geocode large cancer registry databases (1/2 million or more records) that span 5, 10, 15 years or more. [Contact: Ric at voice (610) 965-3595 or email: wskinner@fast.net]

☞ From Roger Masters, through health-gis@who.ch (Subject-US census data): We are working, on a grant from the EPA, to analyze the effects of silicofluoride usage in public water supplies on lead uptake in children's blood and crime and other behavioral and health outcomes. Having found highly significant effects in several states, we are working on a relational

database that will allow us to deal with the complexity of data on dependent variables (sometimes by ZIP, sometimes by town, sometimes by county). Do you have (or no anyone who has) the US census data by census tract, on line in a form that can be sent to us by ftp? Before we spend the considerable time of inserting CD-ROMs one at a time, it seemed worth checking. [Contact: Roger at Roger.D.Masters@dartmouth.edu]

[Note early response: Census CD is the way to go. It is low-cost product that has the complete US Census listings down to the block-group level on a single CD, with an interface that pulls data and geography of interest out of compression. We have used it quite a bit with our research on characterizing populations that reside near hazardous waste sites. This is not an endorsement, simply a statement that for our work it has been quite useful. These data can be readily linked to existing GIS themes you probably already have by Tract Ids. More information on the product can be found at site: <http://www.censuscd.com/cd/censuscd.htm> and from Bill Henriques, GIS Coordinator, Agency for Toxic Substances and Disease Registry, at voice (404) 639-6088 or email wdh2@cdc.gov]

☞ From Rama Nair, University of Ottawa (through list EPIDEMIO-L): I was wondering if any of the list members are aware of any research on the effects of El-Nino: 1) The 1991 Cholera Epidemic in Peru. Any connections to El Nino through floods and/or droughts, plus biological/economic ramifications due to El Nino. 2) The plague in India in 1994...Any connections to El Nino (i.e., due to rodent or insect increases due to el nino or due to drought caused by el nino, or due to temperature anomalies from el nino). 3) 1997 floods in California which left fungi and root rot which seriously hurt the citrus industry. 4) 1998 rodent increase (mainly noticed in the states) which has health organizations on the lookout for an epidemic of hantavirus pulmonary syndrome. [Contact Rama at voice (613) 562-5800, ext 8282 or email rnair@zeus.med.uottawa.ca]

[Note early response: There is a considerable effort underway to investigate the potential health



effects associated with El Nino and ENSO events. Juli Trtanj at NOAA organized an on-going international ENSO Experiment, with the primary goal of assessing the impact of the 1997-98 ENSO event on human health. The participants are an informal group of researchers, agencies, governments and international organizations. Additional information is available from either myself or Juli (trtanj@ogp.noaa.gov). There are a number of manuscripts addressing some of your specific questions on cholera and malaria by Jonathan Patz, Rita Colwell and Menno Bouma. Please let me know if you need the citations. Kristie L. Ebi, Ph.D., MPH, Manager, Epidemiology, EPRI, 3412 Hillview Ave., Palo Alto, CA 94304 or voice (650) 855-2735 or email krisebi@epri.com]

☛ From Victoria Cavataio, San Diego State University: Does anyone know how commonly practitioners now add Hantavirus Pulmonary Syndrome (HPS) to the differential diagnosis of acute respiratory illness throughout the US? Are there quantitative regional differences? Where may I search for such information? [Contact: Vicki at voice (619-594-5433 or email cavataio@rohan.sdsu.edu]

☛ From Amy Griffin, Pennsylvania State University: I am interested in studying the relationship between climate change and hantavirus. I was wondering if anyone has done or is trying to do a historical study of hantavirus incidence with respect to El Nino events - in other words looking for cases which were misdiagnosed at the time of incidence? Thanks for your help on this. Contact: Amy at email algriffin@hotmail.com]

#### IV. Special Reports

(Submissions are open to all)

[Editor: The following is an "advance preview" of excerpts from a paper that will be submitted for publication. Lead author Jon Sperling, Bureau of the Census, has kindly allowed me to include excerpts to convey ideas developed by Jon and Steve Sharp, Vermont Center for Geographic Information. I also wish to encourage other authors to use *Public Health*

*GIS News and Information* for similar releases. We are positioned to get an abbreviated message out in a timely manner and alert the GIS public health community to forthcoming published work. Jon's work addresses a long-standing criticism of TIGER, namely, coordinate accuracy. Based on experience in Vermont, Jon proposes a model to increase coordinate enhancement nationwide with the use of high quality attributed GPS data, through bottom-up partnerships, and potentially enormous post-processing cost-effective savings]

**Excerpts.** "Tiger Strides Accurately Toward the 21<sup>st</sup> Century: A Prototype State Data Sharing Effort Leads to a new GPS-Enabled TIGER Data Base": ...The diffusion and use of geographic information systems (GIS) in state and local government and throughout the private sector during the 1990s has been well documented. While the Census Bureau anticipated and, to some extent, precipitated this growth, the development of an operational system to capture local map updates from a wide variety of files and data formats was slow to emerge. Although pieces of the system have been in place for several years, digital exchange remained unproductive; that is, the efforts to process digital files often required more resources than the existing clerical means to update TIGER. In the absence of a formal digital exchange program, Census Bureau regional office staff manually plotted and used locally provided digital files as paper reference sources...

In essence, the Census Bureau was requesting local governments to repeat work in paper format they had already done digitally. This meant duplication of effort, an increased burden on state and local resources, and the creation of data at a lower level of accuracy; not a happy solution for either party, particularly with "partnerships" being a major theme for Census 2000...

This was the situation encountered in Vermont when each municipality was requested by the Census Bureau to manually update paper census maps. At this time, Vermont had invested heavily in state of the art GIS, had developed highly accurate street centerline files for the entire State, and was nearly geographically complete with an enhanced E-911 project, which

included the capture of GPS points for structures in most towns. Fortunately, the elements to move forward were in place and key staff from the Census Bureau and the VCGI made an effort to understand each others resources, requirements, and needs...The first phase required Vermont's road centerline files to be manipulated prior to integration with the TIGER network...Following this initial processing, the files were converted into TIGER-db format files (different than the TIGER/Line public extracts) and then matched and "rubbersheeted" to the TIGER feature base using census developed routines that work in the TIGER data base environment...

Once all the automated operations were complete, the files were available to Census Bureau geographic staff to interactively add/enhance features and attributes not resolved from the automated match... Since the base for most of the feature network in the TIGER data base for Vermont were based on the USGS 1:100,000-scale files, the discrepancies and the amount of interactive work required was relatively minimal...

For the State of Vermont, which had just undergone E-911 conversion, the automated process transferred tens of thousands of new streets, street names, address ranges, and ZIP Codes from the local files to TIGER in a few hours of computer time. Increased productivity, significant cost savings, enhanced data quality, and an effective partnership were significant benefits of this initial automated update capability. This process, however, did not enhance the positional accuracy of TIGER...

VCGI and the Census Bureau extended the developments achieved in Phase 1 by moving TIGER to the coordinate accuracy of the Vermont files. The Census Bureau rubbersheeted the TIGER road network using the local road centerlines and added roads from the Vermont files not in TIGER...VCGI and the Census Bureau leveraged their individual resources to accomplish this task...Using minimal but highly focused efforts, all preparatory work was completed in a relatively short period of time. The road network in 9 of the 14 counties in Vermont were spatially enhanced in TIGER with minimal interactive effort...

Although the Census Bureau has neither the mandate nor the resources to correct positional accuracy on a nationwide basis, the experience with Vermont poses a successful model and business case to support similar coordinate enhancement activity nationwide following Census 2000. Agencies generally support the vision of NSDI, but unless short-term agency or mission-specific benefits can be derived, there is little incentive to invest resources that would accelerate the evolution of a robust NSDI...

Collaborative efforts between the Census Bureau and the VCGI provide evidence to demonstrate the short-term cost-effectiveness of moving to a more accurate street centerline base. This initiative demonstrates the potential for dramatic cost reductions, quality enhancements, and a re-engineering of geographic support for the collection, processing, and tabulation of census data for Census 2000 and beyond. This work also coincides with the overall objectives of Executive Order 12906 as being implemented by the Federal Geographic Data Committee(FGDC).

The TIGER data base is founded on relational rather than absolute accuracy objectives and contains data that vary widely in terms of spatial accuracy, within counties as well as between counties. However, a new environment characterized by GIS, GPS, digital imagery, and data sharing provides the technical and organizational framework for rethinking future maintenance strategies. The Vermont files show the promise and potential benefits for enhancing coordinate accuracy in TIGER...

Recent innovations in updating the TIGER data base with local, state, tribal, and private GIS databases will continue to be an increasingly important component of TIGER maintenance, because it enables timely and cost-effective updates. Organizational and technical advances in spatial data sharing will create the framework for continuous local partnerships and more accurate decennial and economic censuses. This process will enable ongoing updates to be transferred to the Census Bureau on an ongoing or as-needed basis throughout the decade to support the Master Address File (MAF), American Community Survey (ACS), and the Data Access and Dissemination System (DADS)-

the primary vehicles for re-engineering the collection, processing, and dissemination of data after 2000...

This Vermont initiative provides a needed business model and demonstration for enhancing the spatial accuracy in the TIGER data base and promoting a modernization of the TIGER/MAF data base and processing environment. GPS technology has demonstrated that it can replace/enhance field listing operations and help reduce duplication of effort within the public sector. The development of a more common and shared spatially accurate database will help facilitate data exchange and data collection capabilities. Most importantly, this initiative shows the critical importance of "area integrators" to provide the bottom-up flow of data to support agencies, such as the Census Bureau, and the emerging NSDI...[Contact: Jon Sperling at voice (301) 457-1100 or email [jsperling@geo.census.gov](mailto:jsperling@geo.census.gov)]

### **V. NCHS Cartography and GIS Guest Lecture Series**

(This section may include literature citations, abstracts, syntheses, etc., and submissions are open to all)

Protecting Privacy of Georeferenced Contextual Data, **Alan Saalfeld**, Ph.D., Assistant Professor, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Thursday, November 5: Public health professionals are being challenged with privacy and confidentiality precautions in the presentation and analysis of sensitive georeferenced data. Microdata (data records from individuals) are released by government agencies to the public after the records have been stripped of geographic and other potentially compromising identifiers. Data users would like to know precise geographic location of the records to perform spatial analysis on the data. Users only really need to know relative geographic information about the records to be able to perform spatial analysis. Computing and appending contextual variables to individual records is one option for facilitating spatial analysis by the user. A contextual variable provides the user with summary local information such as averages of surrounding or neighboring individuals.

There are several obvious concerns about attaching contextual variables to a data record. The

primary concern is to avoid compromising an individual's identity by inadvertently exposing too precise a location for the individual. Other key concerns are the cost of processing and the choice of context (not all users want summaries at the same level or type of aggregation).

In this talk, we offer an alternative solution to delivering contextual information as a pre-computed appended variable. We propose instead to organize and deliver the data records as an ordered list that permits the user to compute any number of contextual indicators for a wide variety of different sized neighborhoods or contexts. We show how intervals of consecutively sequenced records in our list can be made to correspond to samples from neighborhoods. This talk examines ways of ordering spatial data so that the ordering reflects different notions of proximity and different sampling strategies. We offer some ideas for randomizing resulting samples. Finally, we discuss our methods' advantages of better data protection, easier data generation, and greater opportunities for spatial analysis.

Here's a disclosure risk scenario: a file of medical records for individuals that tested positive for HIV is being prepared for release to the public for analysis. The agency releasing the file wishes to protect, above all else, the true identities of the participants. Of course, all obvious identifiers of individuals (name, address) are removed. Geographic identifiers that are too limiting, such as census block number, are also removed or left off the published file (they too narrowly restrict the area in which an "adversary" would have to search to possibly ferret out one or more participants).

In general, each additional variable added to a record serves to make the record more identifiable, more at risk, and more readily comparable to existing external data files on real individuals. Location information tends to be highly revealing because it TOTALLY rules out all instances that are NOT at that location (often 99.9+%). Although reporting the exact census block of an individual is deemed to be too risky, reporting an average income for the block may permit some types of socio-economic analysis without compromising the individual or his true location. Here

one must be careful that the income information does not serve as a proxy for block number by actually pointing to a unique block--as might be the case if the average income were computed to the penny (or in the case of Bill Gates' census block, an obvious outlier!)

Risk analysis these days can rarely work in absolutes or in isolation. To assess fairly a risk of disclosure, one must take into consideration as much as possible the variety of readily available information on individuals in the public domain. This information is what "adversaries" will try to match in order to disclose information on individuals' identities that those releasing the files are trying to protect.

[Note: Alan, Laura Zayatz, and Erik Hoel have a joint paper in the 1992 ASA Proceedings on using moving averages (of a linearized spatial order) to generate contextual variables in linear time; Contact Alan at voice (614) 292-6665 or email saalfeld1@osu.edu]

\*\*\*\*\*

Towards a Working Atlas of Model GIS Maps for State and Local Public Health Planning, **Charles Croner**, PhD, National Center for Health Statistics and **Thomas Richards**, MD, Public Health Practice Program Office, CDC, Wednesday, December 16: Many of the more than 3,000 U.S. local health departments (LHDs) are not, as of yet, GIS operational. Limited technology, staffing and training resources appear to be key impediments to current GIS empowerment. Conversely, there is no lack of interest or desire to acquire and implement GIS technology. For many LHDs, adopting GIS remains simply a matter of time.

Tom Richards and I are pursuing several activities that, hopefully, will help "jumpstart" the use of GIS for many of these LHDs. We believe GIS is an indispensable tool to cost-effectively support and improve the planning and provision of LHD services for a wide range of disease prevention and health promotion activities. The purpose of this presentation is to share preliminary findings of successful LHD GIS mapping efforts, that support health planning initiatives, and which can be easily replicated or modeled. Many of these displays illustrate a range of exploratory investigative approaches to commonly occurring community public health issues such as lead

poisoning, immunization, pregnancy and prenatal care, injury, toxic emissions and others. Some already have broken new ground and assisted community leaders in improving community health.

We foresee many uses of a working atlas of GIS maps dedicated to addressing local public health planning needs. For the moment, these GIS maps can serve to provide a much needed template for new and future public health investment at the community level. This is especially the case for many LHDs not quite yet involved in mainstream GIS.

## **VI. Related Census, DHHS and Other Federal Developments**

I. Testimony of Dr. Rita Colwell, Director, National Science Foundation, Before the Basic Research Subcommittee, House Science Committee, September 28, 1998. **Hearing on Remote Sensing as a Research and Management Tool**

Mr. Chairman, members of the subcommittee, thank you for allowing me the opportunity to testify on the use of remote sensing technology in the sciences. Before I begin, let me say how pleased I am to be appearing before this subcommittee again. While I have testified before this subcommittee several times in the past, this is my first appearance in my new capacity as NSF Director.

The Science Committee and the Basic Research Subcommittee have worked hard over the years to promote policies that have helped keep the United States a world leader in science, space and technology. The Committee is unique in that its work has been marked largely by bipartisan cooperation in pursuit of common goals. This tradition has been maintained by you Mr. Chairman, along with Full Committee Chairman Sensenbrenner, Ranking Member Brown, Vice Chairman Ehlers and especially the late Steve Schiff, the former chairman of the Basic Research Subcommittee.

The subject of the hearing today -- remote sensing -- is a special one for me, both as Director of the National Science Foundation and as an active researcher as well. Remote sensing is an wonderful example of the many exciting developments in science. First, it exemplifies how scientists across

disciplines are work and collaborate on complex global scientific problems in new exciting ways. Second it demonstrates how the next generation of computer and communications technology is able to organize, transmit, and disseminate massive amounts of data collected by satellites and remote sensing devices in ways that create benefits for scientists and citizens. Third, it shows how federal agencies like NSF, NASA, NOAA and DoD - each with their own expertise and missions - collaborate effectively to solve fundamental and complex scientific problems.

Remote sensing technology -- whether from satellites, radar facilities or other types of sensors -- is now an important tool for researchers in many disciplines. For many years, development of sophisticated remote sensors were driven in large part because of scientific and national security needs. In the 1980's and early 1990's the United States -- led by NASA -- made a major commitment to improve our ability to gather data about the Earth's systems from satellites. Now a new generation of advanced sensors allow scientists to study the earth's biochemistry and ecology on a fully global scale. Today, at NSF, we regard remote sensing as an indispensable research tool in many disciplines we support, including geosciences, physical science, biology, and social and behavioral sciences.

I have seen the power of remote sensing technology first hand during my research. As an aquatic microbiologist I have spent over 30 years studying the microbial disease, cholera. In recent years, use of satellite data by myself and my students has led to a greater understanding of how global environmental change influences the spread of this deadly disease. We have found, through the use of remotely sensed data, that cholera epidemics can now be related to climate and climate events, including ocean warming events such as El Niño. Further refinements of these studies could allow us to save thousands of lives each year through effective monitoring and prediction of conditions conducive to cholera epidemics.

Cholera is a disease caused by drinking water contaminated with a bacterium known as *Vibrio cholerae*. Cholera can cause severe diarrhea and

dehydration, and in some cases may be fatal - especially to those with weakened or less well-developed immune systems, such as the elderly and young children. Before the advent of modern water treatment facilities, cholera was once the source of major epidemics in European and American cities. Today cholera remains a serious problem in developing countries that lack sources of clean water. The link between clean water and reduced outbreaks of cholera has been known since the nineteenth century. For centuries, cholera was associated with the sea as well, since outbreaks of the disease followed the ocean coastlines. Cholera also tends to spread in erratic but massive pandemics that ravage parts of several continents over long periods of time. Despite these tendencies, we have known little about how exactly the cholera bacterium was spread until recent decades.

The key breakthrough in the early seventies came when we discovered that the cholera bacterium lives in the gut of microscopic aquatic animals, the zooplankton, This pivotal discovery occurred only through years of fundamental research by myself and my fellow researchers and students, both in the laboratory and in the field. Proving the link between plankton and the cholera bacterium allowed my colleagues and I to develop a simple and inexpensive solution to help reduce the presence of cholera bacteria in water obtained from untreated sources.

The low technology solution to combat cholera that we came up with was to filter untreated water through cloth, trapping the plankton and, thereby, the attached bacteria. This low technology approach is critical, since during the monsoon flood season in areas of India and Bangladesh, living conditions deteriorate to simple survival. Building a fire to boil water is not possible. Luckily, inexpensive cloth - used to make a women's dress called a sari - is readily available. Filtering water through three or four layers of sari cloth has been proven to remove plankton and other particulate matter and this reduces the numbers of disease-causing cholera bacteria in the water. Field studies are underway that will soon test our hypothesis that cholera outbreaks will decrease in villages using this new filtering technique.

The relationship between plankton and cholera bacterium also has caused us to study how significant changes in climate such as ocean warming may be a factor in cholera epidemics. The use of ocean climate data generated by satellites has been critical to this research link. As the ocean warms during global climate events like El Niño, we know that large plankton blooms become more prevalent. Given that cholera bacteria was proven to be present in plankton, we felt that there would be a strong correlation between ocean plankton growth and cholera outbreaks. If such a link is verified, it will hopefully lead to direct monitoring and prediction of cholera epidemics in the future.

To correlate cholera cases with changes in ocean climate, we needed enough historical data on cholera cases along with sufficient ocean climate data for a particular region. We have weekly cholera case histories for Bangladesh. Earlier studies using ocean color were valuable in developing the research. Directly collected data on ocean temperature and sea surface height -- usually made by water sampling from research ships -- has been sparse and infrequent up to this point and for the most part been very expensive and time consuming. However, archived remotely sensed data from NASA and NOAA supported satellites on global ocean temperature and height are now publicly available on the Internet.

Comparing the number of cases of cholera with changes in sea surface temperature and ocean height obtained from remote sensing, showed a strong link between changes in ocean climate and incidences of cholera. If verified, this would bring the potential of predicting conditions conducive to cholera outbreaks closer to reality. This would have enormous implications for improvements in human health worldwide. Recent cholera epidemics have swept through Latin America and Africa in the early 1990's. These recent epidemics - part of a larger cholera pandemic that began in the 1960's - killed thousands of people.

This work in developing models to enable prediction of conditions conducive to cholera epidemics...that is, to allow proactive, not just reactive, measures against cholera would be difficult

if not impossible without remote sensing. But satellites represent only half of the picture. Only through advances in high speed computing and communications technology has the data from remote sensing become readily available and usable. These advances have been supported by the National Science Foundation at their earliest stages for many years.. Continued NSF support for fundamental research on information science and technology - through our investment in the Next Generation Internet and other initiatives -- will continue to play a vital role in scientists' ability to effectively use our remote sensing resources.

While remote sensing now is viewed primarily a tool for NSF-funded scientists, it is clear that development of this important tool must proceed hand-in-hand with the development of scientific questions requiring its use. This is one area where NSF and the academic research community can cooperate effectively with mission agencies like NASA and DoD that manage remote sensing resources. Continued design and implementation of higher resolution satellite technology will be critical for improved physical, biological and chemical understanding of our land, sea and atmosphere.

It is clear also that developments in remote sensing technology will have important commercial applications as well. While I will leave it to other witnesses to discuss the details of this important area, I would like to include one or two points. It has been known for several decades -- ever since the Gemini and Apollo astronauts brought back the first photographs of the earth -- that observations from space would have tremendous use in areas such as forestry, hydrology, agriculture, and fisheries management. However, use of remotely sensed data for commercial applications - as well as many scientific ones -- lagged for many years. The reasons for this have little to do with the sophistication of remote sensing technology. Rather it has to do more with the ability to store, manage, access and use the massive data produced by satellites, radar facilities and other remote sensing instruments.

Without advanced information processing, it would take decades to compile and analyze the

incredible amounts of information that produced by many of these instruments. We still have enormous amounts of stored earth climate data from remote sensors that will require additional developments in vast database handling. This research must occur before we can understand fully how our planet's ecosystem works. Since the advent of the Internet and other advances in computing and information sciences, the use of data gathered from remote sensing increased profoundly in a few short years.

Several congressionally-mandated reports on unlocking the potential of remote sensing were written during the 1980's and early 90's. Let me conclude with a quote from the first sentence of one: "If the United States is not prepared to manage efficiently the increase in quantities of remotely sensed data, it will not be able to reap the full benefits of its investment in its satellite systems. In order to use remotely sensed data efficiently, scientists and other users will require adequate data storage and computer systems capable of managing, organizing, sorting, distributing, and manipulating these data at unprecedented speeds."

This prophetic statement is just as true now as when it was written just four short years ago. Nevertheless, the year 1994 is ancient history in terms of the Internet's development - we have come a long way from the days of the 14.4 modem. Advances in computational power, instant communication, vast databases, and extensive analytical capability have brought us to a new era of discoveries in many fields of science and engineering. Now we can explore the universe with powerful tools that unlock knowledge from the subatomic to the super-celestial level.

Some of the contributions of information systems and remote sensors will be like the seafaring ships of earlier eras -- transporting huge quantities of commodities from distant places. Other contributions from information science and technology will be to create whole new disciplines and fields of knowledge, to trigger new industries, and to find new worlds, literally and figuratively. The NSF is poised to lead those diverse expeditions. As the new Director of NSF, I'm excited to be a part of these efforts and look forward to working with you Mr. Chairman and this subcommittee to lead this country forward into the

next century. Thank you. [Source: See <http://www.nsf.gov/od/lpa/congress/rc80928.htm>; Note: Dr. Colwell's name may be familiar to many GIS Users as she delivered one of the keynote speeches at the August "GIS in Public Health" conference]]

\*\*\*\*\*

## II. Public Health Service, NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS

**Subcommittee on Population-Specific Issues**, Excerpts, June 16, 1998, Washington, D.C.

### EXECUTIVE SUMMARY

The Subcommittee on Populations met in break-out session during the NCVHS summer meeting. The Subcommittee is a consolidation of the Subcommittee on Population-Specific Issues and the Workgroup on Population-Based Data of the former Subcommittee on Data Needs, Standards and Security. The reorganization also created a new workgroup for this Subcommittee, the Workgroup on Quality.

The Subcommittee began by discussing plans for the July 14-15 meeting on the Pacific Insular Areas, Puerto Rico, and the Virgin Islands. It was agreed that a synopsis of the meeting and descriptions of the participants would be prepared, and that a governmental official would be invited to give an official welcome.

The Subcommittee was briefed by Harriet Starr about the PRO analysis and IPRO's role in analyzing race and ethnicity data. She described some of IPRO's findings about racial and ethnic differentials in the areas of cardiac catheterization, radical prostatectomy, and carotid endarterectomy.

Next, the group heard presentations from Katherine Wallman of OMB and Clyde Tucker of BLS, Nancy Gordon of the U.S. Bureau of the Census, and Jennifer Madans of NCHS on their work on vital statistics in the wake of OMB's revised standard for reporting data on race and ethnicity. Related issues include collection and tabulation methodologies for a richer set of information and ways to link it with information collected in the past. Clyde Tucker, who heads an initiative in this area, noted that tabulation issues affect data from the Census, national surveys, vital statistics, crime rates, and other information sources. Analysts have begun work to separate out

what would have been true changes over time from those that are methodologically induced by changing the standard. Ms. Gordon explained the Census Bureau's work on race and ethnicity questions and the plans for Research Data Centers. Ms. Madans described the National Center's work on numerous comparability and tabulation issues and its role in providing data for intercensal estimates. NCHS hopes to house a Research Data Center.

Dan Friedman and Ed Sondik briefed the Subcommittee on ongoing planning for a consultation on a vision of health data needs for the 21st century. The focal event envisioned is a workshop (or two), and background papers will be commissioned as well. Dr. Sondik proposed that the Committee put this workshop on its agenda for this year, target a date, and constitute a subcommittee to work on it. The draft National Information Infrastructure paper may be useful to this project.

Finally, Ms. Coltin described her vision of the mission of the new Workgroup on Quality, in which Ms. Ward, Dr. Starfield, and Dr. Lumpkin also will participate. The meeting concluded with attention to several organizational matters. There was apparent agreement that a separate Workgroup on Population-Based Data is not necessary, and that the entire Committee should serve in an advisory capacity to NCHS. The next Subcommittee meeting is September 17.

#### **DISCUSSION OF OMB GOVERNMENT-WIDE STANDARDS FOR DATA ON RACE AND ETHNICITY**

Office of Management and Budget: Ms. Wallman stated that for more than four years, government staff have reviewed the standard for collection of data on race and ethnicity and the implications of increasing intermarriage and immigration, and the resultant trend of increasing numbers in the population of individuals of multiple heritage.

Based on work by a 30-agency Committee, OMB issued on October 30, 1997, a revised standard to permit respondents to Federal Government requests for information to list more than one race if respondents wish to do so. Related issues include the

collection and tabulation methodologies for this richer set of information, and ways to link the new data with information collected in the past. Ms. Wallman noted that race/ethnicity issues mirror those recently faced by those who collect economic data when it was recognized that the U.S. economy had shifted from a manufacturing to a service orientation.

Ms. Wallman introduced Clyde Tucker, who is leading the initiative to develop guidelines for ways to present the data. She emphasized that these guidelines are not intended to be a single standard or a regulation. Mr. Tucker noted that tabulation issues involve data from the census, national surveys, vital statistics, crime rates, and other information sources.

The October 30 notice cited two types of users: those with the responsibility to carry out legislative mandates and those who monitor economic and social trends over time. Users might wish to look at the full distribution of the possible combinations at different geographic levels, or they might wish to see all-inclusive data, which provides complete distribution of all responses (adding up to more than 100 percent of the respondents); or they might wish to collapse the categories for various uses.

Analysts have begun work to separate out what would have been true changes over time versus those that may have been methodologically induced as a result of changing the standard- the "bridge" task. They have also focused on concerns about the ongoing collection of statistics when the denominator is based on census figures. Examples include: vital statistics, crime rates, and estimates from national surveys (in using census population controls).

One group is working specifically on the bridge task and another on procedural design (telephone, self-administered, personal interviews, mail, etc.). The procedural design group will also look at collecting aggregate data, such as school enrollment data. Another group is studying future trends and how to reflect them in data collection and analysis. A number of approaches are being investigated, including the collapsing of categories and algorithmic and non-algorithmic assessments. Part of the investigative process includes simulation research to determine which of these methods might be most



appropriate under which conditions.

**U.S. Bureau of the Census:** Ms. Gordon distributed a handout presenting the questions on ethnicity and race that are part of the Census 2000 Dress Rehearsal. Included are 15 check-off boxes and 3 write-in lines. Because all the data are retained, researchers can specify their own tabulations, and confidentiality is maintained.

One of the Census Bureau's most important products is used in the enforcement of the Voting Rights Act. A "huge array" of data is amassed, i.e., 63 different racial categories and up to 266 possible cells for every block in the country which shows numbers of people, but not characteristics. Work is ongoing with the Justice Department to sort out the balance between protecting confidentiality and meeting the requirements of the law.

Because the full distribution is not useful for many analytical purposes, analysts are looking at possible ways that data can be collapsed. Ms. Gordon described some possible arrays of data by which to report race, including multiple categories of race. She noted that the new standard was adopted in part because of concerns related to health research, i.e., knowing the background of individuals is important to understanding more about certain diseases linked to one's racial heritage.

The Census Bureau has developed an organizational structure called Research Data Centers, in which, under strictly controlled circumstances, researchers are permitted access to data protected by Title 13. The researchers are sworn in and are subject to both financial penalties and jail terms if they divulge data.

#### **National Center for Health Statistics (NCHS)**

Ms. Madans noted that negotiations are underway with the Census Bureau to locate a research data center at NCHS, and that an HHS data center is also anticipated.

NCHS is involved with two separate but related kinds of activities. One issue involves vital statistics. NCHS is ensuring that data collected for numerators will match with what Census is collecting

for denominators, thus maintaining meaningful rates. The other activity is to supply the Census Bureau with data to incorporate into intercensal estimates.

Related NCHS actions include: reevaluating birth and death certificates to make numerators and denominators as comparable as possible, evaluating how the collection of race and ethnicity data on provider-based surveys, the analysis of NHIS data. NHIS has collected (but not published) multiple race categories since 1982. A report will be issued to explain what happened over time using this data collection approach. Separate reports will examine health outcomes vis-a-vis single race and multiple race. If someone reports multiple races, NHIS also asks which is the race that person relates to more strongly.

For major health indicators, the plan is to run both for several years and then pick up on the new line, depending on how quickly society changes. Ms. Gordon stated that a change of around 2 or 3 percent will make little difference. Initial findings of the analysis show a large age effect; most of the multiple race data are being reported among younger people, and it will be interesting to see if that changes over time. Actual results may be forthcoming within several months.

Discussion: Dr. Friedman noted that as NCHS tries to align vital statistics data with the new census data, there appears to be both a classification issue and an issue about self-report versus observation. Ms. Madans concurred, stating that birth certificates will be easier to deal with, but it is unlikely that death certificates will give any multiple race information and may make the problem worse. To ameliorate this problem, she suggested the use of more reliance on matched files; the HIS is routinely matched to the National Death Index. In infant mortality, only the linked file is used to look at race differentials. Ms. Madans was hopeful that increased education of personnel who complete the forms and the increased use of electronic records will have a beneficial effect on quality.

In response to a question on privacy issues, Ms. Gordon explained Title 13, which governs the Census Bureau's protection of data confidentiality, and

some tactics the Census Bureau uses with its public-use files to preserve confidentiality, particularly for records detailing sparsely populated geographical areas.

With help from the National Science Foundation (NSF) in a competitive process, the Census Bureau is gradually expanding the number of research data centers. Researchers must submit a proposal, which must be accepted by NSF and the Census Bureau. One of the acceptance criteria is that the proposed analyzes must be of benefit to the Census Bureau as well as to the greater understanding of phenomena in the world. Successful researchers are given access only to the data necessary for that project, and the data must remain at the data center. Projects that produce regression coefficients, for example, are favored over those that produce tabular results, again for confidentiality reasons.

It was acknowledged that age adjustment changes will be problematical, as will the state CHIP program, which will require population-based data and not administrative data. Ms. Wallman encouraged continued input concerning the guidelines from Subcommittee members.

### **VISION OF HEALTH DATA NEEDS FOR THE 21ST CENTURY**

Dr. Friedman described the history and purpose of this proposed workshop and acknowledged the individuals who have been part of it. The planning Workgroup has developed a draft work plan that focuses on the following goals: (1) developing a 21st century vision for health statistics, (2) specifying that vision in terms of the role NCHS should play in implementing it, and (3) developing specific operational steps for NCHS. The draft work plan was distributed to members of the Subcommittee.

The group discussed developing a 21st century vision, among other things, to define health statistics in terms of its purpose, roles, and boundaries through focus groups. They also explored trying to identify (1) the criteria through which health statistics systems could be evaluated and (2) a process through which to examine recommendations that have evolved relating to health statistics and their systems, and the

implementation of those systems.

Dr. Sondik asserted that NCHS needs advice on many of its choices, especially in light of its very tight budget. Much of the advice it has received over the years has become outdated, either because of advancing technology or inaction. Sponsoring a workshop and involving the broad health statistics community is seen as a way to elicit timely advice.

Among the participants envisioned would be the Committee on National Statistics, whose proposed role would be either to prepare or to commission a synthesis of the advice received over the past several years. The National Committee on Vital and Health Statistics could focus on a draft vision statement regarding where we should be in 2005 and 2010.

Dr. Sondik suggested providing participants with documents to which they could react, to focus the workshop process. The paper(s) would be very useful in differentiating health statistics and the mission of health statistics agencies within HHS, and some of the allied agencies, such as the Census Bureau, from broader information issues. Dr. Sondik proposed that the Committee put this workshop on its agenda for this year, target a date, and constitute a subcommittee to work on it. [Source of Excerpts: Final minutes, Subcommittee on Populations, June 16, 1998]

### **Web Site(s) of Interest for this Edition**

Kate MacQueen, CDC NCHSTP, and CDC Behavioral and Social Science Working Group (BSSWG) mentor, has brought my attention to an interesting contextual presentation of theme mapping. She writes "In our work with qualitative analysis, we often look at the contextual relationships among words and concepts, using similarity/distance matrices and exploring fine-grained v coarse-grained conceptual relationships. There are definite similarities to some of the issues raised in the NCHS Cartography and GIS Guest Lecture of November 5 (see abstract in this edition). A friend recently directed me to some textual mapping applications that you may find intriguing (if you haven't already discovered them on your own). ThemeScope is probably the most visually stunning application for these purposes and can be found at [www.cartia.com](http://www.cartia.com). They have posted a mapping of the

Starr report, by way of illustration; took 3 minutes for their program to compile the full text." The underlying technology used in ThemeScape has actually been in development for over four years, beginning at the Pacific Northwest National Laboratories of the Battelle Memorial Institute. With roots in military intelligence, early development was funded by the

United States federal government to help discover patterns and trends within hundreds of thousands of intelligence documents. Among other operations, the technology played an instrumental role in analyzing Iraqi troop movements after the Gulf War. [Editor: I agree with Kate. This is worth seeing]

### **Final Thought(s): Every Effort Counts**

As we move steadily towards the passage of 1998, and a new year ahead, we have all experienced or witnessed the steady momentum of GIS activity in the field of public health. Nothing could be more satisfying professionally than to be part of the multitude of GIS activities that continues to build with time. There is a GIS culture now emerged in public health and it becomes more established with each passing day. Colleagues involved with GIS are busy with every type of activity imaginable. These activities are contributing to the overall mass and acceleration of GIS in public health.

I have the feeling that GIS is becoming one of the most centripetal technological forces for scientific spatial analytic exploration in our field. The prognosis couldn't be better as we head for the 21<sup>st</sup> century. And there remain many more potential and soon-to-be connected GIS users from a variety of settings that will continue to add to our roles. One example that I cited in September are scientists at Historical Black Colleges and Universities. The same holds true for the many state and local public health (and related) agencies that, often due to limited staffing, technology and training budgets, still constitute a large reserve for future GIS participation.

Also providing impetus to this momentum are the growing demands for public health georeferenced databases and the steady progress of the Federal Geographic Data Committee in support of the National Spatial Data Infrastructure (NSDI). Most federal agencies today, including the US Geological Survey, US Environmental Protection Agency, US Department of Agriculture, the National Oceanic and Atmospheric Administration, and others, readily accommodate the inclusion of health data into their research and reporting agendas. In fact, public health is becoming a player of equal importance among these long-established GIS entities. GIS necessitates that the environmental and earth sciences consider public health data as an integral part of their respective research efforts.

Continue doing what you do in public health and know that your work with GIS is part of this greater whole and associated force. Public health data are an essential part of our NSDI. We can take collective pride in knowing every effort counts.

Charles M. Croner, Ph.D., Editor, ***PUBLIC HEALTH GIS NEWS AND INFORMATION***, Office of Research and Methodology, National Center for Health Statistics <cmc2@cdc.gov>. Copyright Notice: This report is in the public domain but its contents are not to be altered or changed without prior written approval of the editor.

**Until 1999...Stay in GIS touch**