

PUBLIC HEALTH GIS NEWS AND INFORMATION

April, 1998 (No. 21)

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

GIS National Conference



August 17-20, San Diego

Act Now! Announcement and Call for Papers at <http://atsdr1.atsdr.cdc.gov:8080/GIS/conference/>

Selected Contents: Conferences (p.1); News from GIS Users (pp.1-5); GIS outreach (pp.5-8); Special reports (pp.8-12); GIS literature (pp.12-14); Federal news (pp.14-22); Net site of interest (p.22); Final thoughts (p.22)

I. Public Health GIS (and related) Events

☛ Ninth International Conference on Quantitative Methods for Environmental Sciences, July 3-6, 1998, Queensland, Australia [See <http://www.sa.cmis.csiro.au/ties98/>]

☛ 4th International Interdisciplinary Conference on the Environment, July 7-10, 1998, Washington, DC [See <http://www.assumption.edu/html/academic/conf/iiccall.html>]

☛ International Symposium on Spatial Data Handling, July 12-15, 1998, Vancouver, Canada [Sponsor: Department of Geography, Simon Fraser University; See www.sfu.ca/gis/sdh98.htm]

☛ 1998 Americas Conference, The Association for Information Systems (AIS), August 14-16, 1998, Baltimore MD [See conference web site <http://www.isworld.org/ais.ac.98> and geographic technologies web site <http://mennecke.business.ecu.edu/AIS/98/>]

☛ Eighth International Symposium in Medical Geography (IGU), July 13-17, Towson, MD [Contact Robert Earickson at email earickso@umbc.edu]

☛ 1998 Joint Statistical Meetings, The American Statistical Association, August 9-13, Dallas, TX [See <http://www.amstat.org> or use voice (703) 684-1221]

☛ Joint International Society for Environmental Epidemiology (ISEE) and International Society of Exposure Analysis (ISEA) conference, August 15-19, 1998, Boston MA [See www.med.ualberta.ca/PHS/ISEE]

☛ **GIS in Public Health Conference**, Agency for Toxic Substances and Disease Registry (with other federal, state and academic sponsors), August 17-20, 1998, San Diego, CA [See <http://atsdr1.atsdr.cdc.gov:8080/GIS/conference/>]

☛ Epidemiology and Community Interventions in Diverse Populations, 1998 Annual Scientific Meeting of the American College of Epidemiology (ACE), September 26-28, 1998, San Francisco, CA [See <http://amcollepi.org/ace/ace.html>]

II. News from GIS USERS

(Please communicate directly with colleagues on any issues)

A. General News (and Training Opportunities)

1. **Rita Colville**, new Director, National Science Foundation, will be an invited speaker at the 1998 "GIS in Public Health" conference in San Diego. In

a related correspondence, **David Mark**, University of Buffalo writes: President Clinton on Friday (2/13/98) nominated Neal F. Lane, Director of the National Science Foundation, to be his science adviser. To fill the vacancy left by Dr. Lane's departure, Mr. Clinton nominated Rita R. Colwell, president of the University of Maryland Biotechnology Institute, to head the science foundation. Dr. Colwell gave a talk at the CODATA meeting in Bethesda in December 1997 about linking remotely sensed data to health, specifically, sea surface temperatures and cholera. She has also been featured recently on one of the science cable television channels, talking about this line of research. Here is the abstract of her CODATA talk (Global Climate and Health: A Paradigm for an Interdisciplinary Data Network): "The effects of climate on infectious disease have not been well characterized nor are they fully understood. Through data exchange and interdisciplinary networking, it is now possible to realize substantial new information from epidemiological and satellite remote sensing data banks. The hurdles of different missions, language, and data gathering can best be overcome by establishing interdisciplinary teams. With such teams, it is now possible to develop global monitoring of infectious disease and to build predictive models for selected classes of infectious diseases, e.g., those associated with vector, water, and airborne infectious agents."

2. Adopting the Year 2000 Projected Population as the Standard for Age Adjustment of Health Statistics, Special Presentation, April 6, 1998, 10:00 am to 12:00 noon, Building 31C, 6th Floor, Conference Room 6, NIH. [Editor: See related report in Section VI. of this newsletter] The National Center for Health Statistics (NCHS) has proposed that DHHS officially adopt the year 2000 estimated population standard for age adjustment of health statistics. As you may be aware, the standard that is now commonly, but not universally, used is the estimated 1940 census population. Also being proposed is the uniform adoption of this standard within the Department of Health and Human Services for routine presentation of mortality statistics.

This proposal has been endorsed by the DHHS Data Council, and is widely viewed as a beneficial change. These changes would take effect with data on deaths occurring in 1999, although agencies would be free, under certain conditions, to implement the new standard as soon as it is approved. Age-adjusted death rates are one of the key measures used in mortality statistics to take into account the changing age distribution of the population, and thereby to make meaningful comparisons of mortality risk over time and among groups. This effort to adopt a new, uniform standard for use when DHHS releases official mortality statistics to the public will have some significant impacts on our data, since the 1940 standard reflects a much younger population structure.

Since it is important for NIH staff to understand the impact of this change as it relates to various diseases and statistics, Wendy Baldwin's office (OD's Office of Extramural Research) has organized this meeting at which NCHS and NIH staff will discuss the rationale and significance of the proposed change, and the impact and other issues associated with implementing this change. **SPEAKERS: Harry M. Rosenberg**, Ph.D., Chief of the Mortality Statistics Branch, Division of Vital Statistics, National Center for Health Statistics (NCHS), CDC; Robert N. Anderson, Ph.D., statistician with the Mortality Statistics Branch, NCHS; Teri Manolio, M.D., M.H.S., Director of the Epidemiology and Biology Program, National Heart, Lung, and Blood Institute, NIH; Benjamin Hankey, Sc.D., Chief of the Cancer Statistics Branch, Cancer Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH; and, Lynn A. G. Ries, M.S., statistician with the Cancer Statistics Branch, NCHS. [Contact: Martina Vogel at volkovm@od31tm1.od.nih.gov]

3. From **Lois Dean**, Housing and Urban Development: The 1997 national American Housing Survey data is now available from the HUD USER website at: <http://www.huduser.org/data/other/ahs.html>. The national American Housing Survey, conducted every 2 years by the Bureau of Census for HUD, provides detailed data on housing costs, home financing,

household equipment, housing and neighborhood conditions, and the households that occupy them. Data are based on a sample of 55,000 housing units and are reported by region and nationally. For current information on housing needs, market conditions, and existing programs from HUD's Office of Policy Development and Research, visit the HUD USER Web Site at <http://www.huduser.org>. If you have comments on this listserv or the HUD USER Web Site, please send email to huduser@aspensys.com.

4. From **Rita Fellers**, University of North Carolina: Special Session in Cancer Modeling -- 1998 IGU Symposium: Are you working on a cancer modeling project? Consider presenting in the special session on this topic that Dr. **Wendy Kaye**, ATSDR, and I are putting together for the IGU Medical Geography Symposium July 13-17 in Towson, Maryland [See Section I. of this edition]. We are interested in papers dealing with the problems of modeling cancer incidence or mortality, while accounting for as many effect variables as possible and attempting to control for aggregation bias, combining data with differing scales, and other methodological problems. Ecologic studies as well as studies based on individual-level data are most welcome.

Our deadline for getting the session paperwork in is April 3, so e-mail me, the organizer: Rita Fellers, rfellers@email.unc.edu, of your interest, or drop me a line at Dept. of Geography, CB# 3220, UNC-Chapel Hill, Chapel Hill NC 27599-3220 and I'll get the information back to you promptly. Wendy will be chairing the session and moderating the lively discussion which we hope will follow. If we have enough interest, we'll expand to two sessions.

B. Technical News

5. From **J.L. Waldon**, Virginia Tech (ArcView interface with GPS): As a result of my request for information (through FWIM-L@LISTSERV.VT.EDU) regarding software that would interactively link ArcView and a GPS unit, I received two leads and a lot of other interesting stuff included below. The first two are the most promising, one explaining the new ArcView extension due out this quarter and another

from Blue Marble. We've decided to wait for the ArcView extension since we already have a site license with ESRI (I have not included the resulting discussion on low tech navigation that was instigated by my original post).

Subject: Has anybody figured out how to interface a GPS unit directly to ArcView? I'm sending field crews out in January with a laptop, ArcView, and a landsat backdrop to collect ground truth data. It would be great if we could place a 'you are here' dot on the landsat scene for them using the (Trimble) GPS unit.

Response (1): I know that I've heard Trimble talk about their own software to link their units to Arcview, but finally ESRI has leaped in with a real interesting GPS-Arcview link product called the tracking analyst extension, due for release Q1 1998. I've seen it work and besides taking realtime GPS feeds from a variety of formats, it has an impressive suite of tracking, monitoring and mapping functions I've never seen in any GPS product before and I think you folks are going to go postal over this when it comes out. You can search our website www.esri.com on "Tracking Analyst", but here is one of the recent press releases: (<http://www.esri.com/base/products/arcview/extensions/trackingext.html>).

Response (2): Try the "Geographic Tracker" from the Blue Marble Co. It's relatively cheap (\$99). I think you can download a demo copy from their web site at <http://www.bluemarblegeo.com>. I have not used the software but it claims that it allows a GPS interface with ArcView (supports NMEA 0183 and Trimble TSIP/TAIP formats). If you decide to get it could you let us know how it works. We are contemplating buying it as well. We use Trimble Geo Explorer II's and Pro-Light's. I'm not sure about how the physical connection between the Laptop and GPS unit would work. We have always accomplished this by using the older "field notes" software by Penmetrics. It's adequate but is WIN 3.1 based. We normally save the Field notes data as a DBF file and export it to ArcView (or Arc-Info) after differential correction.

Response (3): There are several add-on and stand-alone commercial programs, some which utilize an Arcview interface, and others which use their own desktop layout. They all allow one to plug in a GPS

receiver and display the current position, relative to a background image (satellite image, digital air photo, etc.), and some will allow the inclusion of vector data as well. (Obviously this assumes you have a differential beacon hooked up, or are willing to accept the averaged, non-corrected position). Trimble themselves have a program called ASPEN. PENMAP is another, produced by CONDOR, out in (OR, WA?). GEOLINK is made by GeoResearch, in Maryland (<http://www.georesearch.com>) and there is one more (I will try and dig up the name, that actually uses Arcview for an interface (probably a link within the ESRI home page could steer you towards them.)

Response (4): For point data you can take the coordinates and place them into a delimited text file then add them as an events theme. You need a field for the lat, long, and ID to do this. In my class, the students just store waypoints and type them up. In a real setting, they could be downloaded after conversion and post-processing. Most software can convert to GENERATE format, and generate format points with their comma delimiters are easy to read into an ArcView table. I have not done lines yet.

[Editor: Jeff Waldon is Project Leader, Fish and Wildlife Information Exchange, Dept. of Fisheries and Wildlife Sciences and may be reached at voice (540) 231-7348 or email fwixchg@vt.edu email]

C. Internet News

6. From **Anjani Chandra**, NCHS : For those interested in the geography of health and in health applications of GIS, the 'Just Another Medical Geography Page' at http://www.geocities.com/Tokyo/Flats/7335/medical_geography.htm has been updated. This page is devoted specifically to the field of medical geography and contains sections on the geography of disease, health, and health care, and examples of recent research, as well as an e-mail list of researchers. The new page contains a medical geography discussion forum, a list of electronic discussion groups in medical geography, recent articles and studies, tools and software, a conference calendar, and more.

7. The Northwest Center, in cooperation with the

Washington State Department of Health, is pleased to announce the launching of another email listserv: **waphgis**. This electronic mailing list is intended to foster discussion of issues surrounding the use of Geographic Information Systems (GIS) for public health purposes. The GIS and Health listserv sponsored by the Northwest Center for Public Health Practice at the University of Washington is being set up initially to start a discussion about applications, GIS based public health studies, technical, organizational, policy and funding issues around using GIS for public health matters in WA State. In short, the goal is to see how best to apply GIS in solving public health problems whether in Washington State or elsewhere.

We strongly encourage people interested in GIS as applied to public health whether actually doing GIS work or not to participate and listen from anywhere in the world. What is going on in Washington State may be useful to others and vice versa. The working language will be English, but responses in other languages will be accepted. We need input from anyone who has an answer or point of view. All you need to subscribe to the list is an email account. To subscribe to the list, send a message to listproc@u.washington.edu with the request "subscribe waphgis" followed by your name in the body of the message, like so: subscribe waphgis Jane Doe. You will receive a message shortly after that asking you to confirm your subscription request. Respond to that message, and you will then receive a message confirming your subscription, along with a welcome message telling you about the list. [Editor: If anyone needs help subscribing, or if you have any other problems with the list, contact Phillip Dunham, the list manager: Phillip Dunham, Training Coordinator, INPHO Project, Northwest Center for Public Health Practice, University of Washington at voice (206)616-9245 or email pdunham@u.washington.edu. The list moderator will be **Dick Hoskins**, WA State Department of Health, GIS and Spatial Epidemiology Unit, and can be reached at voice (360) 705-6050 or email REH0303@hub.doh.wa.gov]

8. From Raj Singh, MIT (through

scope@igc.org): I used to work as an environmental consultant and one of the things I did a lot of was site assessments of properties for banks. The idea was that they wanted to make sure there were no environmental problems with the site before they loaned someone money to buy it. Part of our research involved doing a background check on the property to see if there were any past industrial uses. As you can imagine, Sanborn maps were wonderful to have for this. Now to the question of obtaining them. I'd always ask the city if they had a set, and most people would have no idea what I was talking about, but I'd often find some old engineer who understood me and sometimes they'd even have the maps. I've heard that you can also buy them from Sanborn, which is still in business, but I've never tried that. Remember these maps were made for fire insurance, so many rural areas won't have been mapped, but most cities were done and the maps still float around. Universities have them too. A quick search of altavista (search for "sanborn maps") turned up 209 links. [Raj is with the Dept. of Urban Studies & Planning and can be reached at voice (617) 253-0607 or Email rajsingh@mit.edu]

Response from Steven Romalewski, Columbia University: The Sanborn Map Company would know how best to obtain the maps, and they're also in the process of developing their GIS services using electronic versions of their maps. Sanborn can be contacted at: 629 Fifth Avenue, Pelham, NY 10803, at voice (914) 738-1649 or (800) 930-3298. Mark Radin is the GIS Manager at Sanborn, and has worked extensively in GIS over the years.]

Response from **Sally Alhoms**, Bureau of the Census: Mr. Romalewski, thank you for sharing this very valuable (and direct!) source for the Sanborn maps. Ah, the possibilities in getting them in electronic form for use with a GIS! I conducted a neighborhood landuse/population change study in Charlotte, NC, during graduate school. I can remember my delight at having come upon the Sanborn maps in the UNC-Charlotte library. It was a very fascinating and graphic source to see patterns of housing unit and commercial/industrial changes in that neighborhood during the first three decades of this century. Should anyone be in the Washington, DC area, I do know that

the Geography Division of the Library of Congress is also a rich source of insurance maps from Sanborn and also other companies.

9. From Peter Morrison (through FWIM-L@LISTSERV.VT.EDU): I would like to encourage all of you who deal with GIS to tap into the extensive discussions that are held about GIS, biology and conservation on CONSGIS. Many of you already know about this listserv - but for those of you who don't here's the info: Here is how to sign up long version; To subscribe to the CONSGIS LIST send a message to: LISTSERV@URIACC.URI.edu, Leave the SUBJECT field blank (empty) and Enter a one line message which reads: SUBSCRIBE CONSGIS; <SEND> the two word message. Your message should be acknowledged very quickly and you will be asked to confirm that you wish to subscribe. Confirm your subscription request and you will be subscribed. There is no fee for subscription. [Contact Peter at peterm@METHOW.COM]

10. From **Dave Morton**, Virginia Tech: ERDAS is pleased to announce the launching of "ERDAS-L". The ERDAS-L discussion list is intended to foster free and efficient communication among the ERDAS user-base and the extended Geographic Imaging community. For advice and experience-based insight from your professional peers concerning any ERDAS products or concepts, submit your questions, responses, and comments by E-mailing to the ERDAS-L discussion list. To subscribe please visit the ERDAS website at: <http://www.erdas.com> [Contact: Dane Williams at dwilliams@erdas.com]

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 **Cynthia Brewer**, Penn State University: I received a terrific response to my note in the last issue of the newsletter. I was seeking participants for a

funded study on visualizing epidemiological data. I already have 6 data sets that have been sent to me, and I'm discussing final details with three other researchers. I received inquiries from 23 others, and 10 of these people seem interested enough to follow through with sending data. Given that I was seeking a minimum of 10 participants, I am well set. The people I have talked to are from all around the country and I even had a few inquiries from far away places like Peru and Sweden. Data has been enticingly varied, with topics like delinquent behaviors, stage of onset at diagnosis for cancers, occupational mortality, and racoon nematodes. Geographic coding is usually by address, zipcode, or county. Most of the researchers have specific hypotheses about what they are hoping to see in the data and are keen to have it mapped. Thanks go out to everyone who responded and also thanks Chuck for getting my research off to a good start. [Editor: Cindy is Assistant Professor of Geography, Penn State and can be reached at voice (814) 865-5072]

✉ From **Maxia Dong**, NIOSH: Terry Wassell suggested you may be able to help me with the following questions: I am doing a comparison on occupational history and residential history, to see if there is any inconsistency in address given from completed questionnaires. I want to know if a participant lived in one place which he/she was able to commute to his/her workplace. I have heard there are programs which can figure out the distance between two places by Zip code known. I'm wondering if you could give me any information about who would be the expert. [Contact Maxia at mfd7@cdc.gov]

✉ From **Art Getis**, San Diego State University: Although I had not a chance to talk to you at Boston [annual meetings of the Association of American Geographers], I did hear your presentation and read through your working paper (No. 23). I am extremely interested in what you and your associates are doing. I have worked a bit with CDC in Puerto Rico on dengue fever. At Boston, a student of mine and I presented some early work on the spatial distribution of hantavirus syndrome. I did note that you have a

guest lecture series. I would be delighted to have an opportunity to talk to interested researchers on the role of local statistics in ferreting out disease clusters using epidemiological data. J. Keith Ord (Penn State) and I have developed a new family of statistics that we think can be useful in this endeavor. At SDSU, I conduct a seminar on disease clustering that is attended by graduate students in geography and public health. [Contact Art at arthur.getis@sdsu.edu]

✉ From **Ted Hull**, National Archives and Records Administration: Dear Ms. Flock: This is in response to the item posted in the latest issue of "Public Health GIS News and Information," Section III 'GIS Outreach,' regarding the Area Resources File (ARF) data. Per that section, I am forwarding a copy of this reply to Chuck Croner, editor of that publication. The ARF has been identified as a permanent record of the Federal government requiring long-term preservation by the National Archives and Records Administration (NARA). The most recent version of the ARF available through NARA is dated February 1995. We have in our custody a number of historical versions of the ARF data, which we can also make available for a cost recovery fee. You can locate a listing of the various versions of ARF in our custody in our "Title List" available through our homepage at <http://www.nara.gov/nara/electronic>. The ARF files are listed in the Records of the Health Resources and Services Administration (Record Group 512). Machine readable documentation is available for most of the versions of the ARF in our custody; otherwise documentation is available in paper form. Also available via our homepage is information about how we make

copies of electronic records available and our cost recovery fee structure. The cost for a copy of a file on CD-ROM or 9-track tape is \$90.00 and \$80.75 on 3480-class tape cartridge. Additional files added to the same output media are \$24.50 per file, unless (for 9-track tape or 3480-class tape cartridge) you can accept an exact copy of the files as we have them stored where the charge is the basic copy fee. I hope this information is useful. Please contact me if you have any questions. Ted Hull, Electronic and Special

Media Records Services Division, National Archives and Records Administration, 8601 Adelphi Road, College Park, MD 20740-6001, (301) 713-6645, ext. 253

☞ From **Dorothy Wigmore**, (Getting street maps for Mapinfo): I'm trying to do some mapping work with some folks in Florida and wanted to show them what could be done with mapping info about their community as a prelude to looking at workplace mapping. When I have tried to use TIGER street maps and other Census Bureau data in Mapinfo, I can't import them. I remember finding reference to a program that could help with this, somewhere in cyberspace, a while back, but of course can't find it now. Have you any suggestions? I'm looking for Hillsborough County in Florida. Is this something that ATSDR folks might already have in a dbf form I could beg, borrow or steal? Or someone else? Appreciate any leads you can give me.

☞ From **Sid Ganesan**, Vietnam Veterans of America Foundation: The mapping and surveying of areas containing land mines throughout the world to create the Geographical Landmine Information System(GMIS) is a very important step being taken by the International Campaign to Ban Land mines to prevent future victims of antipersonnel and antitank land mines, and to clear mined land vital to the growth of hindered post-war economies. The project that is to be proposed to the Department of State in the coming weeks involves teams of field agents carrying out 3 levels of survey of countries scourged with land mines. GIS technology will be used along with information gathered during the first 2 levels of the survey to map 3 general types of areas: mined; unmined, and mined area which is favorable for settlement and development. These maps will serve to focus mine clearance efforts on areas of greater economic and social importance in the short term. Information obtained during the first two levels of the survey can be found at the following website: "www.un.org/Depts/Landmine/Standard/s-index.htm."

It is important for readers to see this survey report prototype to appreciate the extent of mapping

and surveying that mine clearance requires. Level 1 and 2 surveying consists of field agents using GPS (Global Processing System) units and traditional mapping/ surveying equipment to construct grid references of all mined areas. GIS will help us using these grid references and the digitized demographic (i.e., population, population density, water courses, transportation routes, public utilities etc.)/ topographical (i.e., physical terrain, physical elevation etc.) base data of respective regions to differentiate between and map three types of areas: mined; unmined, and mined area which is optimal for settlement and development. Level 3 of this survey is mine clearance of prioritized mined areas. Digitized base demographic and topographic data of countries around the world will be essential to the mapping of the three types of areas mentioned above. Where can we find this type of data? We will use GIS to create and manipulate maps from grid references of mined areas acquired in level 1 and 2 surveys, topographical base data and demographic base data. The GMIS, along with further use of GIS, will allow the movement to focus scarce resources by the International Campaign to Ban Land mines on mine clearance of areas key to economic progress of war-hindered states. An important result of the mine clearance of such areas is the discouragement of individuals vulnerable to antipersonnel and antitank land mines from areas that are mined, but of little value to their way of life. The Vietnam Veterans of America Foundation would greatly appreciate knowledge of where we can find digitized topographical/ demographic base data, and other ideas about this project. [Contacts: Bob Eaton and Bill Barron, Vietnam Veterans of America Foundation, 2001 S Street NW, Washington DC 20009 at voice 202-4383-9222 or email(s) bob@vi.org and barron@vi.org]

☞ From **Recinda Sherman**, Oregon Health Sciences University: I am starting my MPH thesis project on dengue/GIS and am looking for guidance. I have a degree from Smith College in Biology and Public Policy, and I am very excited to use my MPH thesis to integrate these disciplines. However, I am having a tough time getting started on my thesis. I would like to

use GIS to demonstrate risk for SE US states (Florida, Louisiana, Texas) for dengue transmission by comparing environmental and geographic risk factors with the Caribbean. I am planning an internship to CAREC, Trinidad & Tobago to start collecting incidence data, however, at this point, I don't know the extent of my analysis because I don't even know what data will be available to me. I am currently working with a geographer from Portland State, as well as a medical entomologist, a statistician, and two public health physicians, who are interested in both international medicine and infectious diseases, from OHSU. In addition to the epidemiology/biostatistics program, I have completed a medical entomology independent course and am currently in an independent GIS course. I have strong support from all members on my thesis committee, however, none of them are able to help me with a practical approach to my project. I would really appreciate your help, and many thanks, in advance, for your time. [Recinda can be reached at shermanr@ohsu.edu or voice (503) 494-0591]

IV. Special Reports

(Submissions are open to all)

Clackamas County Department of Health and Human Services, Community Health Mapping Engine (ChiME): Geographic Information Systems Project, Alan Melnick, M.D., M.P.H., Health Officer, Clackamas County, Oregon

Problems Addressed:

Although local and state government routinely collect data related to community health status, for several reasons local health consumers and planners rarely use the data. First, the data is not timely. For example, up to two years may elapse before vital statistics data is released in hard copy form. Once the data is released, the hard copy report contains limited county level analysis and is not amenable to further data manipulation. Local planners are left to ask the responsible state agency to make specific data runs, requiring additional time and staff support. Second, a variety of health related data is collected and maintained in different formats by many different

agencies at the local, state and federal level and is not available in one convenient location accessible to community health planners. Third, health data is analyzed and reported at the county, state and federal level. Larger counties often contain many diverse and sizable communities whose borders do not necessarily coincide with other political boundaries and whose characteristics are not captured accurately by summaries based on these boundaries. Data presented at the county and state level most often fails to address these concerns and consequently is not useful for local communities in conducting health assessment and planning.

Mission and Vision:

The Clackamas County Geographic Information Systems (GIS) Project is designed to overcome these problems and allow local government to engage diverse communities in a partnership to make improvements in community health and well being. An interdisciplinary team is developing a software system, the Community Health Mapping Engine (ChiME), that allows a rapid and easy incorporation of multiple data sets. Through our Healthy Communities partnership process, we hope that multiple agencies, private and governmental, will begin to share data and allow additional data sets to be incorporated into the system. Data sets already incorporated are geographically referenced to allow analysis in a geo-spatial format at the local, sub-county community level. Interested community members and agencies can apply a user-friendly, interactive mapping function to assess a variety of health and social demographic factors and benchmarks related to community health. The system is **flexible** and **modular** so that as additional data sets become available, they can be easily incorporated into the system for use by community members. Safeguards are being placed to protect confidentiality during small area analysis. The tool includes statistical analysis, including confidence intervals, that allows community members to compare their community indices with county, state, national rates and benchmarks, and follow trends over time. Although current datasets and functionality are limited to Clackamas County, the

application is designed to allow expansion to accommodate other regions and geographic scales (counties, states and nations). The application will eventually be web-based so that community members can access the application from a variety of locations.

Specific Features:

*User-friendly interactive system eventually available on a web site. The prototype development environment is "ArcView GIS." We will develop future versions using application-independent languages such as Visual Basic, Java and Map Objects.

*An initial screen will contain text that describes the project, lists data and data sources and provides instructions on how to use the system.

*Help icons and screens will be available at all times. We are designing the system for **two user skill levels**: community members without formal epidemiologic skills and advanced epidemiologic investigators. We will build an epidemiology tutorial into the system for those unfamiliar with epidemiologic concepts. Besides providing instructions on how to use the system, screens will provide easy to understand explanations of concepts such as incidence rates, prevalence, confidence intervals and the need for age adjustment when evaluating mortality rates. Pop-up help screens will contain messages discussing the concept of ecologic fallacy and the need to avoid drawing causation conclusions from the data.

*Users can analyze data at **state, county and community** levels and present their findings in **table, chart and polygon/map** format.

*Census data (1990 and projected) serves as the numerator (social demographic data) and denominator (population data). A commercial product (Equifax) provides inter-census data.

*Users can **compare** community measures with county-wide data, statewide data, benchmarks and (eventually) national data. CHiME will automatically calculate confidence intervals for means/rates allowing statistical comparison.

*Users can **compare** measures for each geographical area over **time** using automatically calculated confidence intervals. Users can also do analysis for

single years or with time aggregated data, allowing study of outcomes with small numbers (such as infant mortality).

*Comparison data (means and rates) can be displayed in map polygons, charts, graphs and tables.

*Table, chart and graph displayed data, including absolute numbers, rates, means and confidence intervals are available by clicking on state, county or community. Users can "zoom" in or out among the levels.

*Because raw data, stripped of identifiers, is incorporated into the system, users can perform stratified analyses, enabling them to evaluate health outcomes by social demographic factors. For example, prenatal care measures can be evaluated by stratifying by income, ethnicity or both.

*Users can assess two well-being variables simultaneously, allowing them instantly to identify communities where the variables are **discordant**.

For example, users can easily determine communities that have both statistically high teen birth rates and low juvenile crime rates.

***Confidentiality** will be assured by restricting analysis, reporting and depiction of very small numbers, especially when multiple stratification is performed.

*For compatibility with population data sources, we defined community areas as an aggregate of a number of census block groups. We plan to share these initial community maps with community groups. We can easily redraw community boundaries based on **community feedback**. The new boundaries will still conform to census block group borders with accurate denominator data, allowing easy analysis.

*We have included a "**Data Wizard**" which allows project administrators easily and rapidly to incorporate additional data sets into the system. The wizard contains a series of steps to reconfigure the new data base and add it to the application. A variety of common data formats are supported. The primary criterion for the new data is that it includes an address field. Data sets already included in the system are listed in Appendix A. Other data sets will be added at the sub-county level once address fields are completed. These include mortality (the system could

calculate Years of Potential Life Lost (YPLL) and age adjusted mortality rates at the community level), two-year-old immunization rates, cancer registry data, high school dropouts, commuting time and domestic abuse (elder, child and spouse). We also hope to work through our partnerships with health care systems and providers to include morbidity data, such as hospital discharge diagnoses.

***Links** to appropriate county health officials will be included, allowing users to ask **questions** and obtain **consultation**. Links to other on-line information sources will also be provided.

Clackamas County:

Clackamas County, with a population of 315,000, extends from the southern corner of Portland eastward to the crest of the Oregon Cascade mountain range and 35 miles to the south. It encompasses an area of nearly 2000 square miles. This economically diverse county contains some of the most impoverished urban and some of the most affluent suburban communities in Oregon. Industry ranges from agriculture, including a large number of nurseries, to a growing high technology sector. The largest minority group is Hispanic, at 3%, but other minority groups include significant numbers of Russian and Southeast Asian immigrants.

Level of Involvement and Interaction with Affected Community Residents:

We designed the CHiME application to increase the involvement of community residents in the health of their community. As mentioned in features above, once the software is on the web, any resident with access to the Internet, including sites such as local libraries, will be able to use the application. Help screens will guide community residents through the system and will teach some basic epidemiologic concepts along the way. Links will be available so that residents can easily contact health department representatives if they have any questions. Other links will identify relevant community health resources and services. Community boundaries can also be changed based on community input.

Even before being placed on the web, the

Clackamas County CHiME has involved community residents through a variety of venues, including the Local Public Safety Coordinating Council, the Reduce Adolescent Pregnancy Project, the Healthy Communities Council and the Robert Wood Johnson Turning Point Partnership.

The Juvenile Crime Subcommittee of the Local Public Safety Coordinating Council is interested in mapping and analyzing juvenile crime rates and associated risk factors at the sub-county level. Because the Clackamas County Sheriff has provided the raw reported crime data, the committee can look at reported community juvenile and adult crime in relation to demographic factors, specific crimes committed and in comparison to other community health indicators such as the teen birth rate. The local Reduce Adolescent Pregnancy Project (RAPP) group is particularly interested in looking at teen birth rates by high school attendance area and by legislative district. Both groups are interested in looking at trend data.

Healthy Communities is a partnership involving community residents, local governments, hospitals, health plans, business, schools, religious leaders and other agencies in the Portland metropolitan area. The Clackamas County CHiME is working with the Healthy Communities Council to expand the number and variety of data sets available and ultimately to build an infrastructure for cooperation and data sharing across organizational boundaries.

Clackamas, Multnomah and Washington Counties (the three counties in the Portland metropolitan area), in conjunction with the Healthy Communities Council, have developed a local partnership funded through Robert Wood Johnson's Turning Point Initiative to study how public health services are delivered and to make recommendations for improvements. One goal of our Turning Point initiative is to develop an integrated data system. Healthy Communities and Turning Point have expressed an interest in using the CHiME as a way of integrating and sharing data among all of our partners.

Impact on the Health of the Community:

The Clackamas CHiME will provide

communities a tool to help themselves in at least two ways: (1) by enabling them to assess a variety of factors related to community well being and (2) by allowing them to evaluate any actions they take in improving their health status. Residents can compare their community health status measures with county and Oregon state benchmarks. The sub-county level of analysis will allow residents and planners to develop specific programs for communities within a large county that have significant problems in specific areas.

Because of the built in time trend analysis, local communities will be able to evaluate the effectiveness of these targeted local initiatives over time. In addition, the two variable analysis will allow planners to evaluate assets and barriers to well-being within and between specific communities. For example, planners may be interested in identifying communities with low socioeconomic indicators and high crime rates that have been successful in lowering teen birth rates, in the hopes of replicating this success elsewhere.

Replicability by Others in the Public Health Field:

The CHiME demonstration uses data sources that are readily available (see appendix A), derived from state vital statistics, the U.S. Census and county government collected data such as reported crime. A freely available commercial product provides inter-census population data. The data wizard allows us to easily incorporate other data bases for use by our communities. Regional support for our CHiME demonstration is already strong. We are convinced that this demonstration will encourage our local

partners, such as local hospitals, to participate in the project by making their data more available as they find it useful for their planning efforts. Software used, “ArcView GIS” and its integrated programming language, is commonly used in many government offices and is an inexpensive desktop application designed to run on personal computers. Future versions of CHiME will be centrally available on-line and will not require any additional investment beyond web-access equipment.

Lessons Learned:

*Communities cross political boundaries and can be defined in many ways. Our starting point was to use high school attendance areas drawn to conform with census block groups. These definitions may need to be changed following community input. Also, the most successful GIS project will require the participation of neighboring counties.

*Polygons depicting incidence rates and prevalence are more useful in evaluating health status than points representing events.

*We need to be careful to employ confidentiality safeguards when small numbers are involved. Also, for some rare events, although there are disadvantages, time aggregated data is more appropriate for analyses.

*Data we collect should include addresses that can be accurately geo-coded. We hope to convince our public and private partners that all health related data collected should include an accurate geo-spatial element.

Appendix A: Data sets Initially Included

Variables	Level of Analysis	Years/Aggregates
Age, Gender and Race	County and Community Levels	Single years: 1990,91,92,93,94,95,96
Personal Income	County and Community Levels	Single years: 1990,91,92,93,94,95,96

Births (Including repeat births)	County and Community Levels	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Abortions	County Level	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Pregnancies	County Level	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Deaths	County Level	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Suicides	County and Community Level	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Arrests	County and Community Levels	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995
Reported crimes	County Level	Single years: 1990,91,92,93,94,95,96 Aggregate: 1991 through 1995

[Contact: Alan also holds the position of Director, Joint Residency in Family Medicine/Public Health & General Preventive Medicine, Oregon Health Sciences University and may be reached at voice (503)494-0756 or E-mail melnicka@ohsu.edu]

V. Public Health GIS Literature

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

Many of the following citations are adapted from **Don**

P. Albert's "Medical Bibliography" Section of the *Medical Geography Newsletter*, Medical Geography Specialty Group, AAG, Vol. 15, No. 1, March, 1998; Don is Research Fellow, Department of Veterans Affairs, Durham, NC and Affiliate Professor, George

Mason University)

Economic Geography: A. J. Bailey et al. (1998) A tale of two counties: Childhood lead poisoning, industrialization, and abatement in New England [Extra Issue:96-111]

Cartography and Geographic Information Systems: T. Nygeres et al. (1997) Geographic information systems for risk evaluation: Perspectives on applications to environmental health 24(3):123-144; J. Chakraborty and M. P. Armstrong (1997) Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment 24(3): 145-157; M. S. Scott and S. L. Cutter (1997) Using relative risk indicators to disclose toxic hazard information to communities 24(3): 158-171; R. B. McMaster et al. (1997) GIS-based environmental equity and risk assessment: Methodological problems and prospects 24(3): 172-189.

Geographic Information Science:

Y. Zhou et al. (1996) GIS based network models of Schistosomiasis infection 2(1-2): 51-57.

Health and Place:

P. Congdon (1997) Bayesian models for spatial incidence: A case study using the BUGS program 3(4): 229-247; N. Ford et al. (1997) The hidden dimension: Sexuality and responding to the threat of HIV/AIDS in South Sulawesi, Indonesia 3(4): 249-258; S. Karvonen and A. Rimpela (1997) Regional diversity in smoking among Finnish adolescents: A comparison of conventional administrative and cultural categorizations 3(4): 271-280.

Journal of Rural Studies:

G. Higgs and S. D. White (1997) Changes in services provision in rural areas: Part 1: The use of GIS in analyzing accessibility to services in rural deprivation research 13(4): 441-450.

Social Science and Medicine:

W. J. M. Martens (1998) Climate change, thermal stress and mortality changes 46: 331-334; K. Judge et

al. (1998) Income inequality and population health 46: 567-579.

American Journal of Tropical Medicine and Hygiene:

A. Khan et al. (1997) An outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates 1994-1995 57: 519-525; R. Thompson et al. (1997) The Matola Malaria Project: A temporal and spatial study of malaria transmission and disease in a suburban area of Maputo, Mozambique 57: 550-559; M. Y. Kosoy et al. (1997) Distribution, diversity, and host specificity of Bartonella rodents from the southeastern United States 57: 578-588.

Annals of the Association of American Geographers:

C. A. Brewer et al. (1997) Mapping mortality: Evaluating color schemes for choropleth maps 87: 411-438.

Annals of Epidemiology:

K. Kafadar (1997) Geographic trends in prostate cancer mortality: An application of spatial smoothers and the need for adjustment 7: 35-45.

Epidemiology and Infection:

T. Baozhang et al. (1997) Infection with human immunodeficiency virus and hepatitis virus in Chinese drug addicts 119: 343-347; M. Haim et al. (1997) An outbreak of Trichinella spiralis infection in southern Lebanon 119: 357-362.

Indian Journal of Medical Research:

A. V. Kondrachine and P. I. Trigg (1997) Global overview of malaria 106: 39-52; V. P. Sharma and A. Srivastava (1997) Role of geographic information system in malaria control 106: 198-204.

Journal of Medical Entomology:

Y. Rubio-Palis and R. H. Zimmerman (1997) Ecoregional classification of malaria vectors in the neotropics 34(5): 499-510; M. W. Service (1997) Mosquito (Diptera: culicidae) dispersal - The long and short of it 34(6): 576-588.

The Lancet:

S. Openshaw (1996) Geographical information systems and tropical diseases 90: 337-339; S. I. Hay (1997) Remote sensing and disease control: Past, present and future 91: 105-106.

Wilderness and Environmental Medicine:

R. L. Langley and W. E. Morrow. (1997) Death resulting from animal attacks in the United States 8: 8-16; J. Stepanek and D. W. Claypool (1997) GPS signal reception under snow cover: A pilot study establishing the potential usefulness of GPS in avalanche search and rescue operations 8: 101-104.

Journal of Rural Health:

L. D. Baer, K. D. Johnson-Webb, and W. M. Gesler (1997) What is rural? A focus on urban influence codes 13(4): 329-333.

Business and Health:

W. R. Brieger et al. (1997) Eradicating guinea worm without well: Unrealized hopes of the Water Decade 12(4): 354-362.

Other: C. M. Croner (1998) *Public Health Geographic Information Systems (GIS) News and Information, 1994-1997*, Cognitive Methods Staff Working Paper Series, No. 23, Office of Research and Methodology, National Center for Health Statistics, is available (while supplies last) by writing to Chuck Croner at cmc2@cdc.gov. This document contains all editions (Nos. 1-18) of *Public Health GIS News and Information* during the initial three years of reporting.

VI. Related Census, DHHS and Other Federal Developments

Report to the Analytic Methods Forum, Summary of the Second Workshop on Age-Adjustment, Harry M. Rosenberg, National Center for Health Statistics, Centers for Disease Control and Prevention, March 15, 1998: The Second (1997) Workshop on Age Adjustment proposed revised procedures to be used by federal agencies to age-adjust death rates.* These procedures will also be recommended to state health

departments to ensure uniformity and comparability of data presentation, beginning with deaths occurring in 1999.

Background

Since 1943, the National Center for Health Statistics and state health departments have used the 1940 population as a standard when age-adjusting death rates. However, DHHS agencies are currently using at least three different standards. Most agencies along with NCHS use the 1940 standard, but some agencies are using standards based on the 1970 and 1980 populations. In two departmental workshops, one in 1991 and one in 1997, there was consensus that DHHS agencies and states need to be consistent when publishing mortality statistics in order to minimize confusion and misunderstanding by data users and the media. Multiple standards also create burdens for the states, who attempt to make their data consistent with Federal practices. Thus, when DHHS releases official mortality statistics to the public, age-adjusted death rates should be based on a single, uniform standard.

In 1991, NCHS convened the first workshop on age adjustment to examine technical issues and problems related to the calculation and interpretation of age-adjusted death rates. Participants included representatives from NCHS, the Centers for Disease Control and Prevention, the National Institutes of Health, the National Academy of Sciences, state health departments, and academia. The first workshop recommended the continued use of the 1940 standard by NCHS and encouraged other federal and state government agencies to use this standard when publishing official mortality statistics. In addition, the first workshop recommended that NCHS study issues that might lead to the introduction of a new or additional standard by the year 2000.

A second workshop in 1997 was convened to examine policy issues related to age adjustment. In both workshops, there was agreement that when it is appropriate to standardize, the choice of population standard is arbitrary, and that trend comparisons and group differences tend to be similar regardless of the standard used. However, many perceive that the 1940 standard is grossly outdated. The second workshop

considered this perception problem sufficient reason to replace the 1940 population with a standard that better reflects the current population. Thus, for both technical and communication reasons, the second workshop recommended that the standard population be changed from the 1940 population to the 2000 population. Although there is no overriding statistical reason for choosing the year 2000 standard over other age distributions, the participants considered it appropriate to choose a year that represents the current population and with which data users can be comfortable.

The second workshop proposed that a new year 2000 population standard be implemented with data year 1999, i.e., beginning with publications reporting 1999 data. This necessitates the use of a projected 2000 population. The recommendations of the second workshop are listed below.

Recommendations

1. The population standard for age-adjusting death rates should be changed from the 1940 standard million population to the projected U.S. 2000 population to be published by the Census Bureau in the spring of 1998. A single standard should be used by all agencies for official presentation of data. For special analyses, alternative standards may be used as appropriate to the research.
2. Agencies should implement the new population standard by data year 1999.
3. Agencies should continue to use and publish their current standards until the official implementation data year (1999) when the new common standard will be adopted. To avoid confusion, agencies implementing the new standard prior to data year 1999 should simultaneously publish rates adjusted to both the old and new standards.
4. After the implementation date, agencies should use the new standard in all press releases and other communication with the public.
5. NCHS will be responsible for selecting a name for the new standard and will determine the number of significant digits.
6. Agencies should continue to use the current eleven age groups (less than 1 year, 1 to 4 years, 5 to 14 years,

15 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 to 74 years, 75 to 84 years, and 85 years and older) for calculating age-adjusted rates using the new standard.

7. NCHS will convene an implementation committee that will be responsible for developing a time table and strategies for implementation and for commissioning papers to publicize the change in standard.

8. NCHS will publicize the new standard in NCHS publications, the *Morbidity and Mortality Weekly Report*, *Public Health Reports*, and appropriate professional newsletters. Scholarly papers could also be published in appropriate professional and technical journals.

9. NCHS will convene a workgroup to evaluate the age-adjustment standard at least every ten years.

Issues

Agencies that wish to implement the year 2000 standard may do so prior to the publication of 1999 data if they also continue to show data adjusted to the 1940 population. This will serve as a transition period from the old to the new standard. Death rates adjusted to the new standard will not be comparable with older age-adjusted rates, and as a result, all agencies will need to retrospectively age-adjust their time series to the new standard.

The targets for the Healthy People objectives for the year 2010 will need to be modified using the new standards. Key staff involved with this project have been informed of the proposal to change the standard. Use of the new standard should not have a major impact on the trends in mortality toward the new targets. However, the targets themselves will look different. Further, there will also be a small impact on race and ethnic differentials. The differences in mortality between the white and black populations will be reduced from 1.6 using the 1940 standard to 1.4 using the year 2000 standard. Similarly, the change in standard will slightly reduce the differential between the Hispanic and white non-Hispanic populations. This will need to be explained.

Concerns have been raised regarding the use of a projected 2000 population rather than the actual 2000 Census population. However, the 2000 Census

will not be available until 2002 or 2003 which would substantially delay implementation. Also, differences between the actual and projected populations should be small. Thus, weights based on the projected population calculated to five or six decimal places will be, in all likelihood, identical to weights based on the actual Census population. It is worth noting that the current 1940 standard is based on a projected 1940 population.

Implementation Plan

The second workshop proposed that the recommendations be widely publicized among federal agencies and States. This will involve the preparation and dissemination of educational materials, presentations, and publications designed to inform key audiences of the change in standard.

[Contact: Harry Rosenberg, Chief, Mortality Statistics Branch at voice (301) 436-8884, ext. 175 or email hmr1@cdc.gov]

*Age adjustment or standardization is one of the key tools used in mortality statistics to take into account the changing age distribution of the population, and thereby to make meaningful comparisons over time and among groups in the risk of mortality. The age-adjusted death rate should be viewed as a construct or an index rather than as an actual measure of mortality risk. Age adjustment by the direct method requires a standard age distribution or "standard population." Statistically, the age-adjusted rate is a weighted average of the age-specific death rates, where the weights represent the standard population proportions by age.

1999 CDC and ATSDR Symposium on Statistical Methods: Emerging Statistical Issues in Public Health for the 21st Century: The Centers for Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR) will sponsor a Symposium on Statistical Methods on January 28-29, 1999 in Atlanta, Georgia. The theme for the symposium is "Emerging Statistical Issues in Public Health for the 21st Century." A short course entitled "Privacy, Confidentiality, and the Protection of Health Data—A Statistical Perspective" will be

offered on January 27, 1999 in conjunction with the Symposium. Both the Symposium and short course are open to the public.

The Symposium will include invited talks and contributed papers. Scientists are encouraged to submit abstracts for papers related to one or more of the session content areas listed below:

- (1) Data collection and storage, including questionnaire and survey design, the use of data registries, and issues related to patients' rights and data privacy and confidentiality.
- (2) Modeling and analysis of complex and/or dependent data structures, including techniques and software for spatial, clustered, longitudinal, survey, and genetic data, hierarchical and causal modeling, and data mining.
- (3) Modeling and analysis of sparse data structures, including issues related to missing values, limits of detection, low dosages or exposures, and rare conditions.
- (4) Design, modeling, and evaluation of public health interventions.
- (5) Applications of statistical methods in public health arenas including infectious and chronic disease prevention, injury and violence prevention, occupational and environmental exposures, and immunization.

Abstracts will be considered for either oral or poster presentation and must be postmarked no later than July 1, 1998. Authors of papers accepted either for oral or poster presentation will be notified by September 30, 1998. To request registration and abstract information and forms, or for additional information regarding the scientific content of the Symposium, please contact **Bradford A. Myers**, 1999 CDC and ATSDR Symposium on Statistical Methods, 1600 Clifton Road N.E. (MS-D01), Atlanta GA 30333 (404-639-3806, fax: 404-639-4463, or email bam6@cdc.gov).

Subcommittee on Privacy and Confidentiality of the National Committee on Vital and Health Statistics (Roundtable Discussion): Identifiability of Data, Excerpts, Wednesday, January 28, 1998, Hubert H.

Humphrey Building, Robert Gellman, J.D., Chair (Note: This transcript is unedited - for the full transcript of this and the following days' discussion on health and medical registries, see <http://aspe.os.dhhs.gov/ncvhs/980128tr.htm>). Mr. Gellman: I think the way we are going to start is I am going to ask some of the people from government to spend a few minutes talking about their data -- data activities; what do you collect; what do you do with it; how do you put it out. Al, do you think you could begin? Take five or ten minutes and give us a good sense of what you do with data, what your statutory constraints may be, and how you function.

Dr. Zarate (Confidentiality Officer, NCHS): I handed out a couple of pages in which I lead off by describing the twin mandates that we always have to balance in assessing what we can release, and what we ought to release. Our agency is charged with disseminating statistics on as wide a basis as is practicable, from our creating legislation, the Public Health Service Act. In the following provision of that act, and I paraphrase it here -- I had to paraphrase it, because it has been described as some of the most tortuous language that was ever put forth in legislation. It basically says that no identifiable -- and you can see why we're very interested to be here -- information may be used for any purpose other than what respondents were told when it was collected, or prior to its collection actually. Nor may it be shared with anyone that the respondent was not aware of or not made aware of prior to its collection.

The term "consent" is vital to this. We are limited by what the respondent has consented to use it for, and to share it with. We have no discretion in this law. It is not like some other provisions which would allow us to release data for certain purposes. We may not release identifiable information.

Mr. Gellman: Is there a statutory definition of identifiable here?

Dr. Zarate: No, unfortunately there is not, but what we have -- you can look at these notes, and peruse them at your leisure. What we normally do is we make of course a distinction, and this appears over and over again, and it is subject to a little interpretation. When data come in, when data are

gathered, of course our field representatives are given strict procedures to follow for the maintenance of confidentiality at the initial point of collection, and to be careful to be able to explain to individuals when we are getting their consent, what the limitations are, what our statute says. So they are aware before they give the data to us. What is also of course not specified and never has been in our statute is the exact meaning of "consent," because that phrase under regulations to be developed by the secretary, and they have never been developed.

So we use a level of consent that is basically consistent with other survey research. It is the implied consent or constructed consent, as I understand it. When an individual, knowing how we intend to use it, who we intend to share it with, gives us the information anyway, that is construed as consent. In other portions of our surveys, particularly which are more invasive you might say, in our health examination surveys we do obtain written consent before an individual undergoes any blood tests, or any kind of measurements.

Mr. Gellman: You collect all of your data, most of your data, some of your data directly from individuals, or do you get it from intermediaries who collect data?

Dr. Zarate: All of our survey data are directly from -- no, sorry, I take that back. We have the Health Examination Survey from individuals. The better known ones, the National Survey of Family Growth, directly from individuals; the National Health Interview Survey directly from individuals. The Hospital Discharge Survey, we obtain samples of medical records from the institution which maintains those records. So that is an example of indirect.

Mr. Gellman: Are those records identifiable?

Dr. Zarate: No, they are not identifiable.

Mr. Gellman: We don't know what that means yet.

Dr. Zarate: The thing is that the only way back to them would be through the provider, and we don't want to know. I'm trying to think now in the other cases where we--the other major case is vital statistics, where this is information on births, deaths, marriages, et cetera, where the consent refers to the state office of

vital statistics, which provides us the data, not to the individual themselves. So it does vary. We observe standard procedures. Now I don't know if you are all aware of them, and one interesting new feature of our data collection efforts is that where previously we used paper questionnaires, now those data are collected electronically. In the case of one survey, it is collected for us by the Census Bureau as our contractors, and they move the data around electronically from region to region.

Finally, it gets to our office in electronic form. It is edited there. The first thing that is done there is that the direct identifiers are stripped, and an analytical file is made. That analytical file, which contains identifiable data, but not the so-called direct ones, explicit ones, then that is used. We keep a master copy of the file with the identifiers in a secure place, and then what our analysts use is what we call in-house files. Those files are still regarded as confidential. At a later point when we decided that this is something that we want to release in a public use basis, it then comes to me, but not after -- we have had the people in the program go through the file. We use a checklist first developed by the Census Bureau, and I guess Easley Hoy will tell you more about that, but which we have now further developed with more health examples in it.

What we do is we get information on the file. We want to know how old it is. We want to know how big it is. We want to know the information that is contained, but more than that, we want the program people to give us some advance information that we'll use to decide the level of identifiability with what they are proposing. We ask them a lot about geography. We ask them not just the level of geography that is identified, but whether or not there are any implicit geographic measures there that they think about. For instance, it is common practice to embed in an internal identification number, some geographic detail; the block on which the respondent was found, the county or the primary sampling unit, which may be a single county, et cetera.

So that the knowledgeable intruder could look at that -- probably go there first -- and figure out what county, what state, and they are off and running. So we

ask them about those kinds of details. Is there any other kind of information that might be used? We ask about detail. With regard to ordinary demographic barriers, which are either commonly available elsewhere, or which might result in a rare and highly visible case. So for instance, recently I had to ask one survey to reduce the amount of detail it provided on height and weight, because some very large and very tall people, and small people--anyway, rare and visible kinds of people even on a regional basis might have been identified. So we ask that kind of information.

We ask whether or not they know of any other files which may be in general or that are matchable into what they have done. Sometimes we tend to forget that where we got the sampling frame originally is going to have this information. In one of our cases where we had a sample of institutions from Dunn and Bradstreet, someone pointed out that well, they've got the sample. So we've got to completely block the information that would describe that sample from them unless the respondents have been told they could have it, and they want it. So that kind of information. So we go through this excruciating detail.

Mr. Gellman: When you do this, do you have a formal checklist or a formal procedure?

Dr. Zarate: Yes, it's in a checklist which actually -- I'm getting ahead of myself, but recently, about three years ago was formulated in the government; a group called the Interagency Confidentiality and Data Access Group. There are people like myself, a lot of them are mathematical statisticians or Easley's colleague Laura Zayatz is now the chair. I'm the vice chair. Jenny Dewolf for the Bureau of Labor Statistics, got this thing going. We are now an interest group; a subgroup if you will of the Federal Committee on Statistical Methodology.

What we have done is we have put our experience together in this checklist, and we have designed it so that it is in an electronic version. You can adapt it to whatever use you want use it for. You can take things out, add some examples. It is intended to educate the people in the programs. In my case, I want people to know why we are so concerned about certain things. It tells them for instance some of the general ways in which the information could be used

to match in other files. We explain to them why we want certain information. It is a standardized checklist. I require all of the people who want to release a file for public use, to fill that out.

I don't regard it, nor does anybody else regard it as the complete answer to the definition of identifiability. It is a start. It means that all of the surveys are treated in the same way. We are sure to ask the same questions that we have agreed are pertinent, and to educate the people who are asking it. It is designed to provide information on sample surveys, complete counts, tabulations, that is the tables, as well as micro data tapes, and it is flexible. So I use a version at NCHS, and we ask people to fill this out. After that is done, I still look at the code to see if they have missed something. They do from time to time, miss some things. If there is a particularly difficult case, I have the discretion -- we have not instituted this on a regular basis like the Bureau of the Census has -- we also have a disclosure review board, where we have some mathematical statisticians, the chairman of the Confidentiality Committee, and people from the program sitting, who review the files, and are available to me to discuss any real problem areas.

From time to time we have even had people come in from the outside to help us look at a particular, really problematic case where I have said, you can't release it. They say, gee, we've got to get this out. Is there any way we could do it? We bring people in to look at it, and to see if there is any way in which it can be recoded. Again, this was alluded to before, the issue of if you put in all of the tools at disclosure limitation arsenal, then you generally make a file unusable, and the researchers are unsatisfied. You may satisfy yourself that you guarded the individual's privacy well, but then researchers can't use it.

You have to say, well, is this one of those cases where I simply can't release this file? We are not going to go anywhere with this, and there have been cases like that, or is there something more we could do to release it. In that situation, after we have gone through all of these processes, we make a judgment. We do not have an absolute way of defining identifiability. When it comes down to it, what we are saying is that where we use that reasonableness test

quite frequently. Where I know that there is another database out there, it is a no brainer. You can't release it. We are constrained.

Then there are situations where we say, well, you know these are common characteristics, demographics which are generally available at almost every level, and we know that states and municipalities are constantly putting together databases which could be matched with ours and we err on the side of precaution. There are other cases where we have to say, who knows? Then we think of well, what would it take to get in there? Who would it take to get in there? How much effort? Who would be interested? I don't think there have been any cases in which we have said, well, it doesn't look -- because what I really try to get people to do is say, well, if you are really unsure, let's do some more tabulations. Let's look at what a potential intruder would be presented with. Usually people will just say, yes, you know, we can't let that out.

So in the face of this, I must say that we have tried to turn to other means of provider researchers -- and I indicated some of those -- the kind of detail that they need. This is very difficult to do, because we do not have the capacity to allow researchers in general -- we very rarely, if ever, have gotten consent from individuals to let a generalized class of people to look at information, like all researchers. We always said that until now, we have never done that.

What we have tried to do is to look at ways in which information in the amount of detail that researchers normally want it, is available to them without compromising confidentiality that we have set up. There are some normal ways in which do it. There is collaboration with researchers in our agency, where the researcher is the one who accesses, or our NCHS staff is the only one who accesses, and his or her colleague then looks at cleaned up tables. We have recently gotten and implemented a visiting scholar program, the American Statistical Association Visiting Scholar Program. We have the administrative meetings to allow that person to access confidential data, but very few people can take advantage of that kind of thing.

So we have turned to other methods of

permitting more people to use the information, three in particular. One, we have an analytical programming service which has been set up to examine tabulations from the National Health Interview Survey and the National Survey of Family Growth. This is an extension of cooperation really. We get tabulations. We give people a dummy file, and they can use that dummy file to get their programs ready. The data don't represent anybody in our samples. They are fictitious cases. They can use that to debug their programs.

They then submit those programs to us, and we run them for them. After we have run them--and of course none of their programs can have instructions to list individual cases, or to ask for geographic detail. We pre-program that so it would permit them to get key items of information for individuals. We then look at their tabulations, and we do tabular disclosure limitation practices, and then give it back to them. So basically, they have never looked at identifiable data, but they have the use of it. We are working on extending that to what we call a remote access system, where there would be fewer people involved, and this would be automated.

Finally, we are trying to develop something that the Center for Economic Studies in the Census has done in several places, and that is to develop a research data center, where individual scholars or researchers can come, and in the capacity of a sworn employee -- which we don't have yet, but we are working on it, and we are hoping to have that, and other security measures -- can access confidential data on our site in such a way that they can't bring in any data with them to match up with it and achieve an identity. Of course, these are files which have been stripped of the obvious identifiers, so that any identification would require some analysis. Effectively what we do is block them from doing that analysis, and making sure that they don't bring anything with them, or they don't take anything out. What they do take out is reviewed by us before they take it out. So we are working on setting up those kinds of systems where people can come, and once again, have the use of confidential data without being able to identify any of our respondents.

Mr. Gellman: Let me ask you to talk for a

minute or two about what you publish. The kinds of information that you put out on data tapes; who uses them; what they use them for. Give us a better substantive sense of why this data is important.

Dr. Zarate: I can tell you that we put out--I think the figure is more than 400 public use data tapes in the last five or six years. It's an enormous number. Some of these are just reiterations. Like the National Health Interview Survey does a survey every year. Where it hasn't varied, it's the same survey every year, but it is more recent data. Then there are data for vital statistics. Data on our Health Examination Survey is done in a cyclical basis, roughly once every four or five years.

Ms. Greenberg: It's going to continuous.

Dr. Zarate: It's going to continuous, that's right. So that information, with all of its supplements--I mean it's not just one file, but it's many different files focusing on particular aspects of the data that they gather. As I said, vital statistics before, and hospital discharge statistics, but our Health Care Statistics Branch is a family, so that there is information on nursing homes, on doctors' visits, ambulatory care. So there is a whole host of information of that variety, where we get information on physicians, and information on the providers, as well as samples of their respondents.

Mr. Gellman: Describe the users.

Dr. Zarate: Users, wow. It's everybody, isn't it? I can't think of any person--periodically we have people reminding us that our principal user is Congress. Then we have the American taxpayer. Then there are the researchers, of every description. There are pharmaceutical people who are interested in especially our cause of death statistics. I think offhand of cases of people who want to use our information for commercial purposes. Some of the standards, the physical dimension data that are put out by the Health Examination Survey are used by commercial interests in establishing seat size on airliners and things like that. This is what I hear around the agency. Other people could give you more detail. I don't think there is any limit or end to the kinds of people who would use our data.

Mr. Gellman: Has it come to your attention

from time to time that the information you have been releasing has been used in a way you didn't intend, that is, has been used to identify people? Is this something that comes up occasionally, rarely? Would you know?

Dr. Zarate: I guess you have to be realistic that there could be ways in which that has happened, and that we don't know. We know of none. None has come to our attention. There have been cases where we say, whoops, let us have that back, but we have always managed to make sure that the circle is unbroken. We get I think challenges in the form of Freedom of Information requests. Our agency has an exemption from those on two counts; one, that it is an unwarranted invasion of person privacy, and it is also covered by another statute. So although we get them periodically, we have only once had to worry about what we were going to give out.

So that from time to time we get worried, because one of our significant is we maintain what is called a National Death Index. This is a system for the convenience of researchers to be able to use the information in our mortality data sets, so that they can really follow people. They are usually people who are following cases and want to know, are the dead, and if so, what did they die of? They may have moved out of the study area. In that case, it would be nice to go to a central spot and locate the death certificate and get that information. What we do is we tell them where they can find it. We don't give them the data. We say, okay, here is where it is likely to be. We don't tell them for sure. So they put in a request, we know this person, and such and such description.

We put that description in and we do a match, and we say here are the ones that are likely to be positive. There may be more than one, and here is where they are. The researcher's task is to go to that state and ask for them, and it's the state's decision to let them have that information, whatever information they want to let them have. Sometimes we get worried that this information is out. There are legal interests who would like to have that information. There has been one case where a lawyer successfully sued to get the information, but basically what had to be done was that the holder of that information had to go back to the states and get permission. So once again, we did

not allow any information out that weren't supposed to. It was adjudicated and taken care of. That's as close as we have come to releasing any identifiable information.

Ms. Breitenstein: My question to you was is there any sort of evaluation about the data elements that are requested? In other words, is a tape a tape a tape, or is there any sort of request for certain data elements which is then answered, or is it like here you go, you can look at whatever you want to look at.

Dr. Zarate: We get these requests all of the time principally for geographical detail. People want to know about their own area or their own region. Because of this, one of the things that we have tried to do is the National Survey of Family Growth I mentioned before is a file. It was specially created with lots of geographic detail. That is, with information about the local area in which women live. Information about the family planning services, and what we call contextual data.

You put this all together -- and this that case, Latanya, where you said we can't make this a public use tape, because there is just too much information on it. It is huge. It is a huge record such that if you took the value of any one individual, there would be none other like that anywhere in the country. So we decided that that would be a good candidate for a remote access system, where people could use that information, but never really see it.

Ms. Breitenstein: I was just wondering if it's like here is the data, or if it's you are requesting these certain data elements; here are these certain data elements.

Dr. Zarate: We have to evaluate them all, and part of my job is to make sure our staff are aware that they just can't grant an interesting request. It's something that you have to be on top of all of the time. I think that one of the things that we try to do is I took seriously, there was section the private lives and public policies, which basically said that confidentiality is not just a question of assuring someone that you are going to take good care of the information they provide, but you've got to actually do it.

We go to, I think--our colleagues in the Census Bureau and other agencies go to extraordinary lengths

not only making sure that informants understand what we are going to do with their information, but making sure that is exactly what we do with it, and that we don't provide any information to anyone -- we've shot ourselves in the foot sometimes where another agency has come in and funded something, and then I had to turn around and tell them, sorry, you can't have the data in the form you wanted it, because we didn't tell the respondents we were going to share it with you.

We told them it would never leave the building, and so that's what is going to happen. It's not going to leave the building. There is a lot of pressure on me from time to time. Fortunately, it's not brought by my superiors, but by a lot of people who want to release this information, and by researchers who will say, isn't it enough for us to tell you that we are nationally recognized, and have been for many years, a professional organization, and we have all the means to take good care of your data. Isn't it enough? I have to tell them no, it's not enough, because we have told the respondents that we wouldn't give it to people like you in the form that would identify them. We can't do it.

Dr. Harding: Just a question and a request. The Interagency Confidentiality and Data Access Group,

you are the vice chair?

Dr. Zarate: Yes.

Dr. Harding: Would it be possible for us as a committee to get some of the thinking and minutes and so forth from that group?

Dr. Zarate: We can give you the thinking. The way we have set it up is that --

Dr. Harding: I don't mean all right now.

Dr. Zarate: Okay, but in order for us to really discuss issues like this, one of our features of our charter is that our minutes are not public documents; legitimately so. If a contractor is there, we ask the contractor to leave if it might be a problem. We can certainly share with you our products and our thinking. That would be no problem.

Dr. Harding: I think that would be very helpful to us, and even if we could give you some feedback on that.

Dr. Zarate: Oh, yes.

Mr. Gellman: A question from the audience, and then we are going to move on to the Census Bureau.

Net Site of Interest for this Edition

The National Oceanic and Atmospheric Administration (NOAA) holds descriptions of over 30,000 data sets held by NOAA and other organizations. One can search the data descriptions via the Internet using full text search software. The Directory was developed as part of NOAA's Environmental Services Data and Information Management Program (ESDIM) in the Environmental Information Services Office and can be accessed at <http://www.esdim.noaa.gov>. [Source: G.S. Barton, "Searching for Historical Climate Data Publications in the NOAA Environmental Services Data Directory," proceedings of the Ninth Symposium on Global Change Studies and Namias Symposium on the Status and Prospects for Climate Prediction, pp. 91-93, American Meteorological Society, January 11-16, Phoenix]

Final Thought(s)

I want to especially encourage all of our GIS User Group colleagues who practice the science of GIS in state and local governments to share experiences through "Public Health GIS News and Information." In this issue, the

report by Alan Melnick on GIS in Clackamas County, Oregon, is instructional, replicable to other local health department settings, and includes important lessons learned. GIS developments at state and county levels are fostering local and neighborhood public health disease prevention planning. We now can envision a hierarchical sharing, at all levels of government, of georeferenced health and environmental data bases that will foster our understanding of disease etiologies and lead to more cost efficient prevention interventions. GIS will contribute to improved public health assessment and disease prevention parity among us as we share not only the data but the successful spatial and spatial analytic techniques that can be uniformly applied across these settings. These are exciting times as more state and local governments become GIS empowered in public health.

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.

August 17-20, 1998 “GIS in Public Health” Conference