NC-CATCH:
North Carolina Comprehensive Assessment for Tracking Community Health

James Studnicki, ScD; John W. Fisher, PhD; Christopher N. Eichelberger, MS

Although the United States invests more resources than any other nation in the world for health, we are far from the healthiest country. In fact, the US languishes in the lower half of industrialized nations for a range of indicators used to measure health status and even trails nations considered to be economically underdeveloped for selected health indicators.[1] Accompanying this disappointing level of overall health status are the enormous disparities in the health of groups defined by race, ethnicity, socioeconomic status, and geography.[2] Despite decades of scrutiny by researchers and health policy analysts and numerous reports and white papers on the subject by prestigious and influential organizations such as the Institute of Medicine (IOM) of the National Academies, these vast differences are pervasive and seemingly intractable.[3,4]

For example, poor White females in Mississippi and in Appalachia experienced declines in life expectancy between 1982 and 2001.[5] Similarly, in North Carolina between 1975 and 2005, despite reductions in absolute rates, Black infant mortality steadily increased from 1.8 to 2.3 times the White rate.

The Measurement Mandate

While the problem of variation in community health status is complicated and has many underlying causes (known and unknown), there is nearly universal agreement that if we are to have any hope of improving it there must be a monitoring system to measure and benchmark community health status in a systematic way. There have been frequent attempts to provide a framework for community health status monitoring systems. The Planned Approach to Community Health (PATCH) developed in 1985 by the Centers for Disease Control and Prevention (CDC) sought to determine root causes and key intervention points for a few selected health issues.[6] In 1991 the American Public Health Association developed the Healthy People 2000 (and later 2010) protocol for addressing a series of defined national objectives.[7] At about this time, the IOM of the National Academies proposed a model for community health improvement that involved an iterative process that cycled through assessment, evaluation, and action phases. The IOM of the National Academies proposed a set of 25 indicators which expanded on the 18 recommended by the original Healthy People 2000 objectives.[8] The National Association of County and City Health Officials and the CDC developed the Assessment Protocol for Excellence in Public Health (APEXPH) which focused on process indicators to determine the internal capacity of public health organizations, and this protocol was later expanded to include other community health organizations.[9] There have been other community health assessment models developed by the hospital industry,[10] state and local government agencies, university research groups, and collaborations of multiple partners. Although these methods vary in their approach to assessment, they all share a common requirement: data. The data required to populate these various methodologies typically include existing secondary sources such as vital statistics (birth and death registration) and the extensive menu of secondary data from the surveys maintained by the National Center for Health Statistics (NCHS). Some of the methods also require

"Access to a wide range of data sources is a prerequisite to effective community health status assessment."

James Studnicki, ScD, is Irwin Belk Endowed Chair in Health Services Research, College of Health and Human Services, at the University of North Carolina at Charlotte. He can be reached at jstudnic (at) uncc.edu or 9201 University Boulevard, Charlotte NC, 28223-0001.

John W. Fisher, PhD, is an assistant professor in the College of Health and Human Services at the University of North Carolina at Charlotte.

Christopher N. Eichelberger, MS, is director of the Software Solutions Laboratory in the College of Computing and Informatics at the University of North Carolina at Charlotte.

primary data collection efforts to provide a perspective on local populations or providers not available in existing secondary data sources.

From Data to Knowledge: The Challenge of Analytics

Access to a wide range of data sources is a prerequisite to effective community health status assessment. Bringing together data from multiple sources, linking and integrating them, and continually updating and maintaining them for useful analysis can be a daunting challenge to even the most capable and resourceful organization. For the typical local community organization, such a task is often beyond their financial and technical capabilities.

The potential of the Internet as a means to disseminate data for health assessment has been recognized, and there has been a rapid increase in the number of states which maintain at least some form of access to health data through the World Wide Web. Although only 10 states reported a Web-based query tool in 2000, by 2006 the number had grown to 27 states.[11] In most cases, the state systems are limited to a static interface which provides prestructured statistics such as death counts or rates organized into preformatted reports. In a few instances, the query menu itself is dynamic and enables the user to make a selection of data sets, statistics, tables, and maps. These systems, however, do not have the kind of multidimensional navigation and analytical capabilities that can be made available with modern online analytical processing software. Therefore, when viewed from the perspective of what is really state-of-the-art given the rapid advances in computational software and Web-enabled applications, currently existing systems demonstrate major limitations. These include unidimensional filtering queries; single grain data without real time "roll up" or "drill down" capability; data set "silos" which are unlinked and nonintegrated; inconsistent definitions across data sets; lack of metadata (ie, data about the data necessary to facilitate the understanding, use, and management of the system); and lack of an imbedded, organized framework for extracting decision support knowledge from the underlying data.

The Community Health Assessment Portal

North Carolina has been one of the most active states nationally in promoting community health status assessment, and counties are required to produce a formal report at least every 4 years. The State Center for Health Statistics maintains a number of databases useful to this process and works closely with Healthy Carolinians in performing assessments and mobilizing multisectored community action. With assistance from a health services research and technical development team from the University of North Carolina at Charlotte, the Division of Public Health (with additional funding from the Kate B. Reynolds Charitable Trust) initiated the development and deployment of a system that will address many of the weaknesses of current systems, thus bringing modern Web-enabled software technology to public health. To this end, the Division of Public Health decided to move beyond merely producing data to actually enhancing the capability to analyze the data more effectively.

The North Carolina Community Health Assessment Portal (NC CHAP) can best be understood as a means to bring the capabilities of business intelligence to public health surveillance, particularly community health assessment. The original innovators in this field used the word "business" in a broad sense as a collection of activities carried on for some purpose such as commerce, science, government, or public health. "Intelligence" is also defined in a general sense as the ability to "apprehend the interrelationships of presented facts in such a way as to guide actions toward a desired goal."[12] Modern business intelligence systems use data that has been gathered into a data warehouse or data mart and also occasionally use transactional or operating data. Using various types of software to support reporting, interactive pivot-table analyses, visualization, statistical data mining, and other technologies and applications, business intelligence systems provide historical, current, and predictive fact-level views of the enterprise in order to support better decision making.

NC-CATCH

As a business intelligence system, the Comprehensive Assessment for Tracking Community Health (NC-CATCH) has the following components:

Data Warehousing. Extant data organized into the data warehouse will include demographic/population data at the census track level (updated annually by a commercial firm); mortality data; birth data; pregnancy data; hospital discharge data; emergency department visit data; behavioral risk factor survey data (regional and county level only); cancer incidence and treatment data; and other miscellaneous social, economic, and health-related data available at least at the county level. The data will be geocoded to the census tract where possible. Access to the various levels of warehouse capability will be through a portal on the Web site of the State Center for Health Statistics.

Methodology. Health is a multidimensional concept, and there is no single perfect measure for summarizing the health status of a defined population. CATCH is a comparative framework that provides a consistent, objective, multidimensional method for organizing and interpreting community health data. Key elements of CATCH include multiple indicators organized into categories (eg, infectious disease, maternal and child); comparisons with peer counties; state values; Healthy People 2010 values and other benchmarking standards; trend analysis for the most recent 3-5 year time period; an objectively derived rank ordered list of community health challenges; and a concise assessment of racial/ethnic health status disparities.[13] The CATCH report, uniformly aggregated for each North Carolina county, will provide a comprehensive starting point from which the relative health challenges and achievements can be understood, and areas requiring further analyses can be identified. The CATCH

report includes embedded hyperlinks into the appropriate queries that allow full investigation of the underlying data.

Prestructured Queries. Prestructured queries allow the user more flexibility to choose the variables and dimensions of interest, but the choices are necessarily limited. A user, for example, may want to view infant mortality by race or by county, or hospitalization rates for diabetes by county. Some of these views are precalculated and made available through the portal, but they too are static like the CATCH report. These types of queries are the typical method for supplying information in most states. Another type of query is more active in that the user can select from a few number of dimensions to provide a limited form of query customization. These systems are interactive only to the extent that the user is given a menu of queries to select from. Since the number of possible queries is limited only by available data, the challenge is to provide a menu of queries that satisfy the needs of most users or provide information about the most important problems. Structured queries, however, can never maximize the use of information since they cannot anticipate nor perform all the multidimensional data navigation and analysis necessary to model a complex system such as population health.

Online Analytical Processing. Online analytical processing is a category of software technology that enables users to gain insight into data through fast interactive access to a wide variety of possible views of information transformed from raw data. The NC CATCH user utilizing online analytical processing functionality will be able to calculate and model across dimensions, through hierarchies, and across members; slice subsets for onscreen viewing (eg, organize hospital discharge data by pay source); drill down to deeper levels of consolidation (eg, identify all census tracts with a certain percentage of population in poverty); reach through to underlying detail data (eg, compare infant mortality in different counties by the specific causes of infant death); or rotate to new dimensional comparisons in the viewing area (eg, move from Black/White comparisons to Hispanic/non-Hispanic comparisons of age-banded chronic disease mortality). Online analytical processing truly brings the power of the data warehouse to the user's desktop. Multidimensional analysis makes it easier to navigate the database, and because the data are physically stored in a multidimensional structure, the speed of these operations is faster and more consistent than is possible in other database structures.

A Brief Explanation of the Technical Approach

The principal challenge to presenting consolidated health data consists of identifying the appropriate levels at which to enforce data standards. Raw health data exist in many different forms and formats: plain text documents, images, printed tables, spreadsheets, and databases. Each of these forms is well suited for its particular audience.

Each of these levels is treated separately within the following sections and related to the left-hand side of Figure 1:

(1) The raw health data sources (the small cylinders at the bottom of the figure) were identified at the beginning of the project. Each was defined in terms of format, layout, and meaning. Staff used this information to build small software routines that load a given raw data source into a central database (the large cylinder in the figure). This transaction-level database is the first level of standardization in the project.

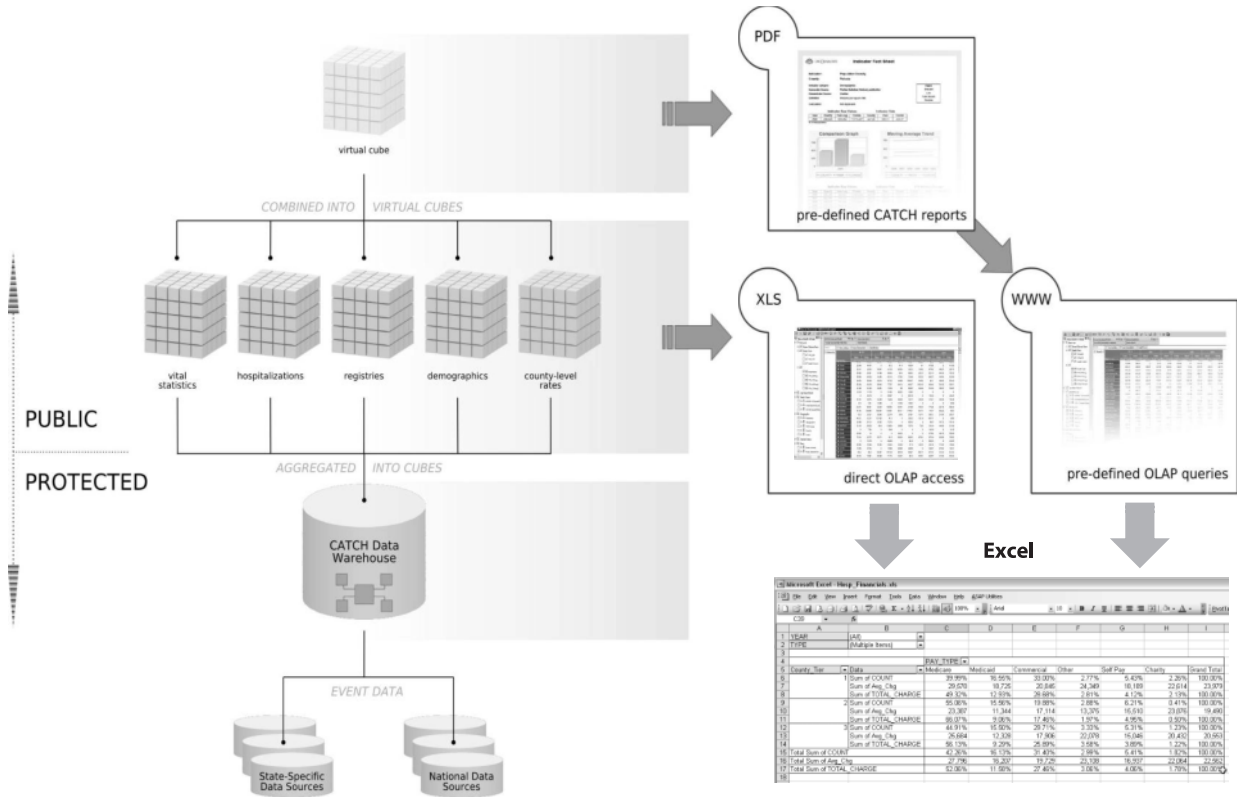
(2) Once the raw data have been loaded into the transaction-level database, a second set of standards is applied to bridge separate sources and construct a series of unified multidimensional data cubes. A data cube is merely a collection of data that has been preaggregated along a given set of dimensions so that every possible combination of dimension values can provide a set of output measures efficiently. In the figure, the data cubes are depicted as gridded boxes in which each small component corresponds to one intersection of dimensions. Typically, we interpret dimensions as X, Y, and Z, but in data cubes the dimensions are more likely to be time, age band, race, sex, cause of death, etc. Identifying the standards that allow for the bridges among data sets, as well as defining the common dimensions and measures that typify the logical cubes, is the second level of standardization in the project.

(3) The logical data cubes themselves can be recombined into virtual cubes (illustrated as translucent, gridded boxes at the top of the figure) providing glimpses into the data that were not previously available. For example, one logical cube is dedicated to the detailed demographics of North Carolina; another cube is dedicated to the deaths that have occurred over time. Combining these 2 cubes into a virtual cube allows us to explore not only the base demographics and the deaths, but rates and age-adjustments across all geographies, years, demographic splits, and causes of death. Identifying the meaningful combinations of logical cubes into virtual cubes is the third level of standardization in the project.

This project seeks to provide 3 levels of access targeted to meet the needs of 3 distinct user communities, differentiated by the trade-off between ease-of-use and analytical flexibility/complexity. These users are served with different views as shown in the right-hand side of Figure 1:

(1) The general public and many senior policy staff are likely interested in the predefined county profiles that are part of the CATCH report (available as a Web page). This use of the commonly used report format provides rapid access to the highest level of aggregation and comparative benchmarks. However, since the report is not a static document but actually contains active hyperlinks, it also

Figure 1.
North Carolina Health Assessment Portal Construction and Interface



provides immediate access to the more detailed underlying data.

(2) These data, of significant concern for county health departments, are accessed through a large set of predefined online analytical processing queries or exploratory reports using both the logical and virtual cubes. Once again, a report-style interface is employed to minimize the technical expertise required to quickly access the desired information. This simplified drill-down capability provides a powerful tool for investigating multidimensional relationships.

(3) At the high end of complexity and flexibility, researchers have access to the lowest level data in the free-form online analytical processing queries. The full range of dimensions is available in a drag-and-drop visual interface for creating ad hoc queries across the entire data warehouse.

Each of these levels serves a specific role in the reporting of health data, and each has an independently maintained set of standards for how data are allowed to participate. These metadata themselves constitute one of the most significant new capabilities of the project. **NCMJ**

REFERENCES

1 World Health Organization. *The World Health Report 2000. Health Systems: Improving Performance*. Geneva, Switzerland: World Health Organization; 2000: 143-145.

2 Bloche MG. Health care disparities: science, politics and race. *N Engl J Med*. 2004;350(15):1568-1570.

3 Committee for the Study of the Future of Public Health, Division of Health Care Services, Institute of Medicine of the National Academies. *The Future of Public Health*. Washington, DC: National Academies Press; 1988.

4 Durch JS, Bailey LA, Stoto MA, eds, Committee on Using Performance Monitoring to Improve Community Health, Division of Health Promotion and Disease Prevention. *Improving Health in the Community: A Role for Performance Monitoring*. Washington, DC: National Academies Press; 1997.

5 Murry CJL, Kulkarni SC, Michaud C, et al. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PloS Med*. 2006 September;3(9):e260. doi:10.1371/journal.pmed.0030260.

6 Green LW. PATCH: CDC's planned approach to community health, an application of PRECEED and an inspiration for PROCEED. *J Health Educ*. 1992;23(3):140-147.

7 US Department of Health and Human Services. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives*. Washington, DC: US Government Printing Office; 1991.

8 Perrin EB, Koshel JJ, eds, Panel on Performance Measures and Data for Public Health Performance Partnership Grants, National Research Council. *Assessment of Performance Measures for Public Health, Substance Abuse, and Mental Health*. Washington, DC: National Academies Press; 1997.
9 National Association of County and City Health Officials. *Assessment Protocol for Excellence in Public Health*. Washington, DC: National Association of County and City Health Officials; 1991.
10 Gordon RL, Baker EL, Roper WL, Omenn GS. Prevention and the reforming US health care system: changing roles and responsibilities for public health. *Annu Rev Public Health*. 1996;17:489-509.
11 Friedman DJ, Gibson Parish R. Characteristics, desired functionalities, and datasets of state web-based data query systems. *J Public Health Manag Pract*. 2006;12(2):119-129.
12 Luhn HP. A business intelligence system. *IBM Journal*. 1958;October:314-319.
13 Studnicki J, Steverson B, Meyers B, Hevner AR, Berndt DJ. A community health report card: Comprehensive Assessment for Tracking Community Health (CATCH). *Best Pract Benchmarking Healthc*. 1997;2(5):196-207.