NC CATCH: Advancing Public Health Analytics

**James Studnicki[1], John W. Fisher[1], Christopher Eichelberger[2],
Colleen Bridger[3], Kim Angelon-Gaetz[4], Debi Nelson[4]**

[1]**University of North Carolina, Charlotte**
  College of Health and Human Services
  Department of Public Health Sciences

[2]**College of Computing and Informatics**
  Software Solutions Laboratory, Charlotte, North Carolina

[3]**Gaston County Health Department[3]**
  Gastonia, North Carolina

[4]**North Carolina Office of Healthy Carolinians/Health Education[4]**
  Raleigh, North Carolina

Abstract

The North Carolina Comprehensive Assessment for Tracking Community Health (NC CATCH) is a Web-based analytical system deployed to local public health units and their community partners. The system has the following characteristics: flexible, powerful online analytic processing (OLAP) interface; multiple sources of multidimensional, event-level data fully conformed to common definitions in a data warehouse structure; enabled utilization of available decision support software tools; analytic capabilities distributed and optimized locally with centralized technical infrastructure; two levels of access differentiated by the user (anonymous versus registered) and by the analytical flexibility (Community Profile versus Design Phase); and, an emphasis on user training and feedback.

The ability of local public health units to engage in outcomes-based performance measurement will be influenced by continuing access to event-level data, developments in evidence-based practice for improving population health, and the application of information technology-based analytic tools and methods.

Introduction

The 1988 Institute of Medicine (IOM) report titled "The Future of Public Health", and other IOM reports since then, have advanced the idea that community health status could be improved by a data-driven continuous iterative cycle of assessment, program implementation, reassessment of results, and further implementation of newly focused programs.[1] These reports emphasized the need for a regular and systematic collection, assemblage, and analysis of information on the health status of communities which would support priority setting and evaluation of the impacts of programs and policies, and stimulate the collaboration and actions necessary to improve community health outcomes.[2,3]

In response to this measurement mandate, there has been a continuing production of frameworks, models, and community health status report cards.[4,5,6,7] Each of these efforts presents a rendition of community health status accompanied by a set of indicators or measures linked to determinants of health (e.g. poverty, race), root causes of adverse variations on health (e.g. smoking, obesity), or key intervention points related to selected health issues (e.g. immunizations, screening). In some cases, these community measures are weighted and mathematically manipulated in order to derive a community score or ranking.[8,9] Static models using a fixed selection of indicators and a similarly static scoring algorithm provide the basis for coarse comparisons, but are not alone sufficient to enable communities to discover their own unique determinant-outcome relationships and practice priorities for subpopulations defined by race, ethnicity, age, poverty, geography, outcomes and other factors.[10]

Brief catch history

The CATCH methodology evolved from a series of comprehensive community health status assessments conducted in Florida in the 1990s. These extensive hardcopy reports were manually cobbled together from multiple data sources using a comparative framework which enabled each community (usually a county or group of counties) to compare itself against sociodemographically similar peer communities.[11] Funding from the U.S. Department of Commerce, Telecommunications and Information Infrastructure Assistance Program (TIIAP) in 1998 enabled the automation of many of the analytic steps and resulted in larger and more complex reports, as well as a vibrant research agenda with studies in racial and ethnic disparities, the impact of special taxing districts on health outcomes, warehouse applications to bioterrorism alert algorithms, and improved methods for community health status assessment.[12,13,14,15] With the realization that the same data and analytical capability required to support these research endeavors was necessary to understanding variations in the health status of defined populations, the CATCH effort in North Carolina evolved away from simply providing data and reports to deploying an operating analytical environment composed of a rich repository of data harnessed to a powerful analytic capability.

NC catch: system elements

In North Carolina, the State Center for Health Statistics (SCHS) maintains an inventory of databases to support the mandated community health status assessment process and works closely with the Office of Healthy Carolinians and Health Education (OHC) and local community partnerships in performing assessments and mobilizing community action. With assistance from a health services research and technical development team from the University of North Carolina at Charlotte (UNCC), the NC Division of Public Health initiated the development and deployment of a system that would address many of the weaknesses of current systems, thus bringing the benefits of modern web-enabled software technology to public health. Key components of the system include:

Data from multiple sources. Extant data from multiple sources with conformed definitions are organized into the warehouse: demographic/population data at the census tract level; mortality; pregnancies; births; hospital discharges; emergency room visits; 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 various levels of granularity. Data are geocoded to the census tract where possible. An important future source of data is the electronic health record (EHR), since the analytical capabilities of the system are congruent with the goal of at least one category of "meaningful use" of EHRs as specified by the Office of the National Coordinator (ONC) for Information Technology[16,17]; i.e. to improve population and public health. The ability to move clinical practice data from health information exchanges (HIEs) into a CATCH data warehouse in a timely manner will enable broader use of that data for management, evaluation and policy purposes.

On-Line Analytical Processing (OLAP). The most prevalent electronic storage system is the relational database, in which data elements are organized into two-dimensional tables of columns (that remain fixed) and rows (that can be added to, deleted from, and modified in place). The following (Table 1) illustrates one such simplified data table.

Table 1. Simplified death record

Death Record I.D.	Age	Race	Cause of Death
2185	65	01	ICD-10-CM codes
7364	85	01	ICD-10-CM codes
1122	21	02	ICD-10-CM codes
7419	53	03	ICD-10-CM codes

This structure facilitates storing transactions which are single (row-based) assertions about each death: patient identity, cause of death, age and race of the deceased, etc. Each different type of data, however, requires a separate data table. These individual tables can be logically joined through common data elements such as the death record ID or cause of death. Though efficient for storing individual facts, this structure is not particularly conducive to open-ended data exploration tasks because the user has to traverse all of the tables to assemble a coherent view of

3

the data that are spread across the entire transactional database. OLAP-based data warehouses address this shortcoming by providing pre-assembled collections of system-wide data into hypercubes (or just "cubes" for brevity). The following (Figure 1) illustrates one such simplified cube:
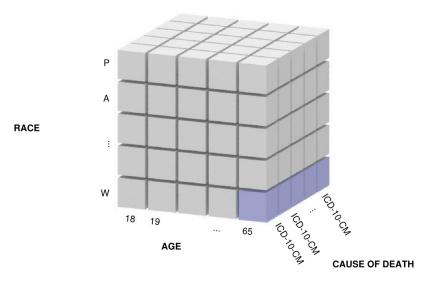

Figure 1. Multidimensional "hypercube"

Even though this example cube contains only some of the columns from the preceding data table, it can contain an arbitrary number of dimensions, typically including geography as well as time. Every intersection of these dimensions represents a cell that can contain one or more pre-computed, aggregate measures such as the total number of deaths, mean mortality rates, total cost of services, etc.

The following (Table 2) contrasts relational databases, pre-computed aggregate indicators, and OLAP cubes:

Table 2. Comparison of database structures

	Relational database	**Pre-computed aggregate (indicator)**	**OLAP cube**
Identity?	All records are identified.	No records are identified.	No records are identified.
Aggregation?	These are event-level (fully disaggregated) data with specific values, such as MRN, DOB, or cause of death.	Data are binned into ranges, but a single indicator typically allows only one column to vary, e.g., death rate by age-band for a fixed location, time period, race(s), cause(s) of death.	Data are binned into ranges (that can be organized into hierarchies), but all dimensions can be explored in any combination, even mixing and matching hierarchy levels.
Big picture?	Must join multiple tables into a single, sparse matrix, but making sense of this is difficult.	Even simple domains require thousands of indicators to express the full nature of the problem.	Each cube *is* the big picture.

A crucial advantage of this cube-like structure is the ability to extract arbitrary subsets very quickly. Asking for everything related to any death record yields a subset (or "slice") that contains all of the pre-computed measures relating to this single death across all other characteristics such as age, race, and cause of death. Asking for the intersection of all deaths belonging to 65-year-old whites produces the aggregates relating to this one specific age-of-death by race (the shaded area in Figure 1). The principal advantage of having loaded the base transactional data into a data warehouse is that it allows the local health departments to sift-and-sort through their data in a much more interactive -- and much more natural -- way than would have been possible through a traditional transaction-oriented data store. OLAP cubes can produce an answer for complex queries much faster than the same query on an online transaction processing (OLTP) system.[18]

Multidimensional, event-level data. For simple, shallow, pre-computed reports, summary data aggregated at the county, region, or state level may suffice. To take full advantage of the exploratory capabilities that are provided by NC CATCH, however, requires having event-level data wherever possible, because the system cannot anticipate what level of analysis the end users wish to conduct. A mature platform for data exploration should allow its users to query data by geography, time, demographics, and data-set-specific properties such as disease, cause of death, birth weight, procedure performed, etc. This is what NC CATCH does, and it works best with data that are fully described; that is, entirely disaggregated.

Consider, for example, the various dimensions and measures which are available for inclusion in the typical hospital discharge (fact) data set: reporting year, reporting quarter, hospital number, type of admission, source of admission, discharge status, patient race, patient sex, patient zip code, principal diagnosis, secondary diagnoses, principal procedure, secondary procedures, principal payer, charges by revenue groups, DRG code, patient age at admission, length of stay, day of week admitted, days from admission to procedure, patient county, facility county, and (in some states) attending and operating physician identification numbers. Each dimension will have a set of hierarchical elements which themselves can be relatively coarse such as patient sex (i.e. male, female, unknown) or fine grained such as diagnosis (i.e. thousands of possibilities based upon the ICD-9-CM coding system). The analytical potential of this extensive information is only available to the user who can access all of the detail and has the infrastructure to enable the analyses, as well as the knowledge and experience to exploit this potential for maximum insight.

Access to fine grained, event-level data, such as hospital discharge datasets, also makes it possible to utilize analytical software which has been developed by third parties (including government agencies) specifically to analyze this available information. NC CATCH, for example, utilizes a series of software tools that are available without cost from the Agency for Healthcare Research and Quality (AHRQ).

The Prevention Quality Indicators (PQIs) are a set of measures to be used with inpatient discharge data to identify ambulatory sensitive conditions (ASC) in discharges; i.e. conditions for which good outpatient care can potentially prevent hospitalization or for which early intervention could prevent complications or more severe disease. Although these indicators are based on hospital inpatient data, they are often used to provide insight into the community health care system or services outside the hospital setting. Other AHRQ indicator sets available in NC CATCH are the Inpatient Quality Indicators (IQIs) and the Patient Safety Indicators (PSIs). The IQIs are a set of measures that reflect quality of care inside hospitals including inpatient mortality for certain procedures and medical conditions; the utilization of procedures for which there are questions of overuse or underuse; and the volume of procedures for which there is evidence that higher volume is associated with lower mortality. A subset of the indicators is recommended for area-level utilization rate analysis. The PSIs are a set of indicators providing information on potential in hospital complications and adverse events following surgeries, procedures and childbirth. Six of the indicators also have area level analogs and can be used to detect patient safety problems on a regional level, or for subpopulations defined in other ways.

Although commonly used in many static report card systems, summarized data that are aggregated from event level data have no analytical flexibility and are, therefore, of limited usefulness in interpreting the various relationships which influence population health status. An example of such an indicator is the hospitalization rate for ambulatory sensitive conditions (ASC) per 1,000 Medicare enrollees. This indicator aggregates all causes for an ASC admission and provides data only for Medicare, thus providing a very restricted view of preventable hospitalizations within any community. By contrast, with access to multiple years of event level hospital discharge data and the AHRQ suite of analytical software, NC CATCH is able to derive the full analytical benefit from the ASC construct – to understand avoidable hospitalizations for subgroups defined in multiple ways, e.g. by diagnosis, age, race, payer source, geographical

location, pathway of hospitalization (scheduled or through the ER), trends in the variables over time, and many other factors. The following screenshot (Figure 2) shows a query which displays the distribution of four specific diabetes related types of avoidable hospitalizations within a single county, by gender and type of admission. With the ability to provide flexible alternative views of preventable hospitalizations, NC CATCH is able to model across dimensions, through hierarchies, and across members inside any population of interest. This flexibility enables the public health analyst to understand the nature of preventable hospitalizations as manifested uniquely in each community.

Figure 2. Screenshot: Diabetes related ASC admissions by type and gender

Two levels of access. NC CATCH supports both anonymous public users and registered users (Figure 3).
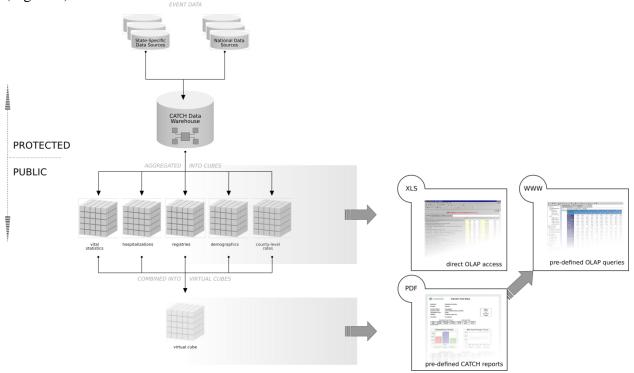

Figure 3. Access architecture for NC CATCH

Anonymous users have access to the Community Profiles that summarize, by category, public health indicators relating to any county of their choice. These indicator groupings were composed by a committee of system users in order to enable the local analyst to select the category or categories of particular interest; e.g. overall mortality (shown), injury and violence, reproductive health and others. Each selected group of indicators opens to a series of gauges which place the subject county value in reference to the state average and highest and lowest county values for each indicator (Figure 4). These indicator values are contrasted with both the values of the county's peers -- chosen specifically for each county on the basis of selected socio-demographic characteristics -- and with the State values. There is some additional detail available to the users of this level of the system, such as thematic mapping for geographical granularity (census tract, community, county). The flexible customized views of the underlying data cubes (i.e. Design Phase) within the warehouse are restricted to registered users, giving them the ability to explore the data for deeper relationships and greater understanding. The process of becoming a registered NC CATCH user requires approval by the local health officer and the county administrator.
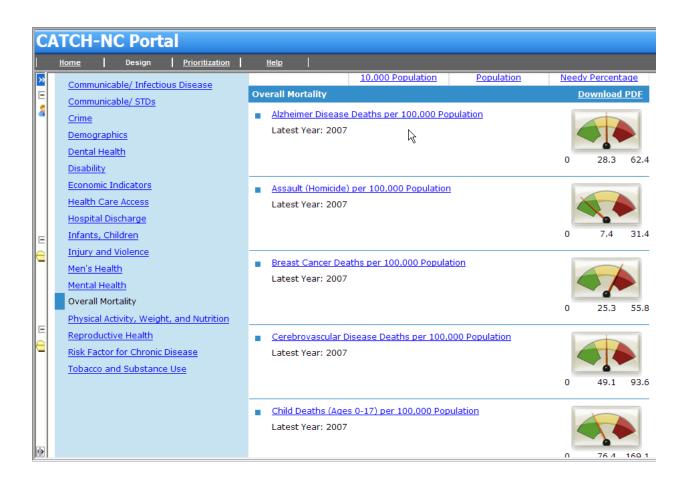
Figure 4. NC CATCH Public access county profiles and indicator groupings

<u>Operational governance and structure</u>. All aspects of the NC CATCH system are directed by the SCHS working with an advisory committee composed of representatives from the SCHS, the Office of Healthy Carolinians and Health Education (OHC), local public health directors and staff, and the UNCC development team. The advisory committee sets the strategy for new development and incorporates modifications, as appropriate, based on user feedback on various aspects such as the look and feel of the interface, the grouping of various health measures into meaningful categories, and the content and conduct of training sessions. The advisory committee is responsible for maintaining a coherent vision of the NC CATCH system as it changes over time, and for determining that the maintenance and enhancement of the system is consistent with that strategic vision.

The technical infrastructure is centralized to minimize development and maintenance costs, but the analytic capabilities are distributed and optimized locally. This enables even the smallest, resource poor local public health unit to have access to this powerful, flexible system. Use of the hypercube aggregation model (OLAP) also addresses privacy concerns by allowing full analysis of event-level data while protecting the data itself. No event-level data is actually deployed; only the precomputed aggregates are populated for every combination of dimension cross sections.

Training. After the launch of Phase 1 (County Health Profiles) in October 2008, the program was introduced to target users through a series of webinars. The webinars exposed the need for instruction and training particularly for the OLAP Design Phase. The OHC staff was tasked with the planning, designing and evaluation of the on-site trainings. Health department staff and their Healthy Carolinians partners from all counties in North Carolina were invited to one of 25 training sessions conveniently located throughout the state. Training groups were limited to 15 or fewer participants.

The five-hour trainings were composed of four modules: Introduction to NC CATCH, Understanding Statistics, Using the County Health Profile, and Tailoring County Reports. The "Understanding Statistics" section reviewed the basic statistics featured in NC CATCH and familiarized users with vocabulary and notations specific to the system. The last two modules focused on learning how to gather and interpret data through the system to meet CHA needs and accreditation standards. During the training, participants completed both instructor-guided and independent exercises to practice creating useful data queries. For example, one exercise asked participants to examine and graph their county's pregnancies by maternal age, allowing them to practice selecting and filtering many fields to find the answer to a relevant question in the Design Phase. OHC developed a training manual as a reference for the new user trainees that reviewed basic statistical concepts, the documentation available in the system (metadata) to aid data interpretation, and highlights of additional features available to the advanced user. Pre and post training evaluations were administered to determine whether participants learned the basic concepts presented. In addition, each participant evaluated both the on-site and webinar trainings, so that the effectiveness of each training method could be compared. Results from the tests and evaluations were reviewed weekly. Trainings were modified when necessary, based on feedback from the training participants.

NC CATCH training sessions were typically held at a computer lab or conference room in the local county health department or community college. Between May and October 2009, over 200 health professionals from 83 out of 100 counties were trained on NC CATCH. Most participants were health educators, although health directors, epidemiologists, program evaluators, and health policy staff also regularly attended the trainings. Participants worked in priority areas including youth tobacco prevention, nutrition, childhood obesity, environmental health, HIV/STD prevention, cancer prevention, women's health, and substance abuse prevention. Most participants had formal education in public health and qualitative data analysis; however, most had not had recent training in statistics and quantitative data interpretation. Anonymous evaluations were used to determine the participants' satisfaction with the training sessions and their reactions to the system itself.

Improvement and expansion of training opportunities for NC CATCH users continues to be a system priority. In person and online (webinar) training is now available. A hypertext help file is available online. Video answers (screen video to frequently asked questions) are in the process of development. A formal user group has been established with regular feedback to the SCHS regarding system enhancements and training needs.

Future development issues

The evolution of a distributed analytical environment for population health measurement and improvement is particularly dependent upon three major issues:

Data availability. A frequent complaint from public health decision makers is the paucity of hard data about the health status and behaviors of vulnerable subpopulations. However, the trend in most states is toward more, not fewer, restrictions on access to health outcome data. Driven primarily by patient privacy concerns and in response to ever-more powerful data aggregation technologies, access to event-level data is becoming increasingly difficult. Even pre-aggregated data is often suppressed. For instance, the CDC WONDER data warehouse suppresses all mortality data where the total death count is less than six in counties of under 100,000 population and the time span is less than three years[19]. In North Carolina, over 75% of the counties are under 100,000 population (2007 estimates).

The desire to use patient encounter data for wider purposes undergirds such efforts as the Agency for Health Care Research and Quality's Provider Based Population Health initiative and the ONC-IT Beacon Communities program. The allure of gaining greater understanding of patient behaviors and the "Meaningful Use" mandate will require some accommodation of privacy concerns if data are to be utilized in anything approaching their true potential.

The current default strategy is selective masking and total suppression. A more useful strategy is the practice of forcing aggregation until sufficient numbers of events and/or populations are covered. For instance, if a particular cause of death for a small geographical area for a single year, specific gender and particular race results in too few events to satisfy data identification concerns, aggregation across either time, race, gender, or years can be forced until sufficient numbers are achieved. For this approach to satisfy the needs of researchers and decision makers, however, the end user must be in control of the aggregation.

Evidence based practice. Current thinking regarding population health status is oriented to the measurement model best typified by the National Quality Forum (NQF) measurement endorsement process, most successfully applied to healthcare structural, process, and outcome measures.[20] Major limitations in this approach are apparent when attempting to apply this process to health status outcomes for geographically defined populations. Evidence for community level interventions (in the form of programs and services) that will produce reliable and valid results across communities of varying sizes, sociodemographic composition, and other characteristics (measured and unmeasured) is sparse. The science of measuring healthcare performance has made progress in the last decade largely through rigorous evidence-based review, the development of risk-adjustment techniques and methods, and access to event-level clinical data. Deployment of electronic health record technology is expected to accelerate this ability to measure healthcare services and outcomes. By contrast, public health practice has been largely bypassed by the advances in modern information technology: event-level data is difficult to access; no model of comprehensive community risk adjustment has been validated; and the local public health unit, with rare big city exceptions, has limited analytical infrastructure with which to determine local priorities or evaluate the impact of programs and services.

Information technology. The existence of the CATCH infrastructure opens up the possible utilization of many methodologies and technologies which can enhance the system, among them data mining and non-linear pattern recognition. One area of particular promise is visual analytics which is the science of analytical reasoning facilitated by visual interactive interfaces. Visual analytics is most useful in situations which are complex and where the need for closely coupled human and computer analytics may make them otherwise infeasible; for example, where one is trying to determine the varying contribution of community racial composition on a large number of multiple outcomes such as many specific causes of mortality. These techniques hold the promise of providing the ability to analyze large and complex datasets rapidly either independently or as a screening precursor to traditional computational analysis.

Conclusions

The shortcomings of the system of local public health units in the United States have been well documented: lack of modern information technology, an aging workforce in need of training, declining public financial support, and the lack of a clear vision about its role. The performance measurement initiative taking place in the healthcare system has not been replicated with similar urgency in public health; program evaluation is rare, the evidence base for public health practices is growing but still sparse, and population outcomes are neglected.[21] Advances in information technology and software development have made it cost-effective to provide powerful and flexible analytic capability to local public health units. This important infrastructure for evolving an analytical culture for public health is also a necessary component for measuring and improving population health.

Acknowledgements

References

1) Institute of Medicine: Summary of recommendations, in Waterfall W (ed): The Future of Public Health. Washington, DC, National Academy Press, 1988; 7-9.

2) Institute of Medicine: Measurement tools for a community health improvement process, in: Durch J. Bailey L, Stoto M (eds): Improving Health in the Community: A Role for Performance Monitoring. Washington, DC, National Academy Press, 1997; 126.

3) Institute of Medicine: Healthy Communities: new partnerships for the future of public health. Washington, DC, National Academy Press, 1996.

4) 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.

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

6) 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, substances abuse and mental health. Washington, DC, National Academy Press; 1997.

7) National Association of County and City Health Officials. Assessment Protocol for Excellence in Public Health. Washington, DC, 1991.

8) Fielding JE, Sutherland CE, Halfon N. Community report cards: results of a national survey. Am J Prev Med. 1999; 17(1): 79-86.

9) www.countyhealthrankings.org.

10) Wolfson, Michael C., Notes on Measurement and Accountability, Presentation to IOM Committee on Public Health Strategies to Improve Health Meeting Two, January 2010, http://iom.edu/~/media/Files/Activity%20Files/PublicHealth/PHStrategies/Meeting%202/Wolfson2.pdf

11) Studnicki J, Steverson B, Meyers B, et. al. A community health report card: Comprehensive Assessment for Tracking Community Health (CATCH). Best Pract Benchmarking Healthc. 1997; 2(5): 196-207.

12) Studnicki J, Berndt D, Luther S. Hispanic health status in Orange County, Florida. J Public Health Manag Pract. 2005; 11(4): 326-332.

13) Studnicki J, Gipson L, Fisher J, et. al. Special healthcare taxing districts: association with population health status. Am J Prev Med. 2007; 32(2): 116-123.

14) Berndt D, Fisher JW, Craighead JG, et. al. The role of data warehousing in bioterrorism surveillance. Decision Support Systems. 2007; 43: 1383-1403.

15) Studnicki J, Luther SL, Kromrey J, et. al. A minimum data set and empirical model for population health status assessment. Am J Prev Med. 2001; 20(1): 40-49.

16) ehealthinitiative.org. "National Progress Report". 7 Dec 2010 <http://www.ehealthinitiative.org/national-progress-report-improving-population-health leveraging-electronic-data.html>

17) Office of the National Coordinator for Health Information Technology. "HIT_Strategic_Framework_2010-05-10.pdf". 7 Dec 2010. <http://healthit.hhs.gov/portal/server.pt/document/911844/hit_strategic_framework051010_pdf>

18) Chaudhuri S, Umeshwar D. An overview of data warehousing and OLAP technology. SIGMOD Rec. 1997. (ACM)26: 65. doi: 10.1145/248603.248616.http://doi.acm.org/10.1145/248603.248616.

19) Center for Disease Control. "Compressed Mortality File 1979-1998 and 1999-2007". 7 Dec 2010. <http://wonder.cdc.gov/wonder/help/cmf.html>

20) National Quality Forum. The ABCs of Measurement. NQF, Washington, DC 20005. www.qualityforum.org.

21) Jacobson PD, Gostin LO. Restoring health to health reform. JAMA. 2010; 304(1): 85-86.

Correspondence:

James Studnicki, Sc.D.
Irwin Belk Endowed Chair and Professor[1]
jstudnic@uncc.edu
Phone: 704-687-8981
Fax: 704-687-6122