2000

Community Healh Assessments: A Data Warehousing Approach

Donald J. Berndt
University of South Florida, dbermdt@coba.usf.edu

Alan R. Hevner
University of South Florida, ahevner@usf.edu

James Studnicki
University of South Florida, jstudnic@hsc.usf.edu

Community Health Assessments: A Data Warehousing Approach[*]

Donald J. Berndt
Alan R. Hevner
College of Business Administration
University of South Florida
Tampa, FL 33620
{dberndt, ahevner}@coba.usf.edu

James Studnicki
College of Public Health
University of South Florida
Tampa, FL 33620
jstudnic@hsc.usf.edu

Abstract - **The measurement and assessment of health status in communities throughout the world is a massive information technology challenge. The Comprehensive Assessment for Tracking Community Health (CATCH) methodology provides a systematic framework for community-level assessment that can be a valuable tool for resource allocation and health care policy formulation. CATCH utilizes health status indicators from multiple data sources, using an innovative comparative framework and weighted evaluation process to produce a rank-ordered list of critical community health care challenges. The community-level focus is intended to empower local decision-makers and provide a clear methodology for organizing and interpreting relevant health care data. The effectiveness of the CATCH methodology is based on a data warehousing approach. The data warehouse allows a core set of reports to be produced at a reasonable cost for community use. In addition, online analytic processing (OLAP) functionality can be used to gain a deeper understanding of the health care issues. The data warehouse in conjunction with Internet-enabled dissemination methods will allow the information to be presented in a variety of formats and be distributed more widely in the decision-making community. On-going research directions in community health care decision making conclude the paper.**

I. COMMUNITY HEALTH ORGANIZATIONS

It is well documented that considerable variation exists in the health status of defined populations. This variation is evident when we compare large population groups, such as separate nations, states, or regions within a single country. Surprisingly, variation often persists within smaller population groups, such as census tracts or zip codes inside United States counties. These variations exist not only for what would be considered epidemiological health status outcomes (i.e., morbidity and mortality rates), but also for indicators which could be considered other dimensions or domains of population health such as socioeconomic and demographic characteristics, the availability of health resources, patterns of health behaviors, and many other factors. In order to improve the health status of populations, a continuous monitoring and improvement system must be implemented. Such a system would require a comprehensive, objective, and uniform methodology for defining and characterizing the many dimensions that comprise the health status of a community.

In the U.S., the Institute of Medicine (IOM) of the National Academy of Sciences, in its influential 1988 report on the Future of Public Health, emphasized that assessment was one of the core functions of public health and recommended that there should be a regular and systematic collection, assemblage, and analysis of information on the health status and needs of communities [7]. In 1997, the IOM Committee on Using Performance Monitoring to Improve Community Health outlined a community health improvement process through which communities can assess health needs and priorities, formulate a health improvement strategy, and use performance indicators as part of a continuing and accountable process [9]. The report called for a *community health profile* made up of socio-demographic characteristics, health status indicators, quality of life indicators, health risk factors, health resource indicators, and other measures which can be used to support priority setting, resource allocation decisions, and the evaluation of health program impacts.

As part of the on-going clarification of the public health role at the community level and the transition from a *disease* to a *health* focus and from a *treatment* to a *prevention* strategy, there has been recognition that partnerships and collaboration are necessary to support effective action [8, 12]. Health organizations, public sector agencies, medical care providers, businesses, the religious community, educational institutions, and other community organizations are interdependent components of a multi-sectoral community health organization. The overall community must be empowered to make the necessary, and sometimes difficult, resource allocation choices to improve health through information, education, behavior change, and social support [3]. Such collaborative action at the community level must be informed by *unbiased data* describing the community's health status, needs, and resources. The ability is also needed to track progress over time to meet the community's health care goals.

The gap between current practice in community health care spending and the above goals of collaborative community health care decision making is vast. The availability and quality of data on health indicators are problematic. There is little empirical evidence on the use, sharing, or strategies to integrate health data into decision-making to provide guidance to community health organizations. While most of the literature on collaborative leadership and community engagement focus on the process [1, 2], little attention has been focussed on the effect of the availability of a common set of data, such as the community health profile, on the quality and inclusiveness of decision-making. There is also scant information about the use of data and information technology to support and monitor the process.

The purpose of this paper is to present an outline of the Comprehensive Assessment for Tracking Community Health (CATCH) methodology and its implementation in a data warehouse. The various modes of data dissemination from the data warehouse to the community are explored and examples of current CATCH interfaces are demonstrated. We conclude by examining important issues of community decision support on health care.

II. THE CATCH METHODOLOGY

The University of South Florida's Center for Health Outcomes Research has developed the CATCH methodology to provide comprehensive, objective health status data for community health planning purposes. CATCH collects, organizes, analyzes, prioritizes, and reports data on 225 health and social indicators on a local community basis. The CATCH methodology has been tested, refined, and validated over the past nine years. Reports have been prepared for 15 U.S. counties both within and outside of Florida.

The CATCH methodology can be briefly described as shown in Fig. 1. Community health indicator data are gathered from a variety of sources. Secondary data sources include health care data reported by hospitals, local, state, and federal health agencies, and national health care groups. Primary data sources would involve data gathered from door-to-door or mail-in surveys. All health care data are normalized into common formats and organized into a community health care report card listing values for each important community indicator.

Each indicator value is then compared against the state average, a peer group of communities average, and other interesting values (e.g., a national goal for that indicator). The results of these comparisons are organized into an n-dimensional matrix based on favorable or unfavorable comparisons against each comparison dimension. Fig. 1 shows a 2-dimensional comparison matrix based on state averages and peer averages. Community indicators that demonstrate unfavorable comparisons on all dimensions are highlighted as community health challenges.
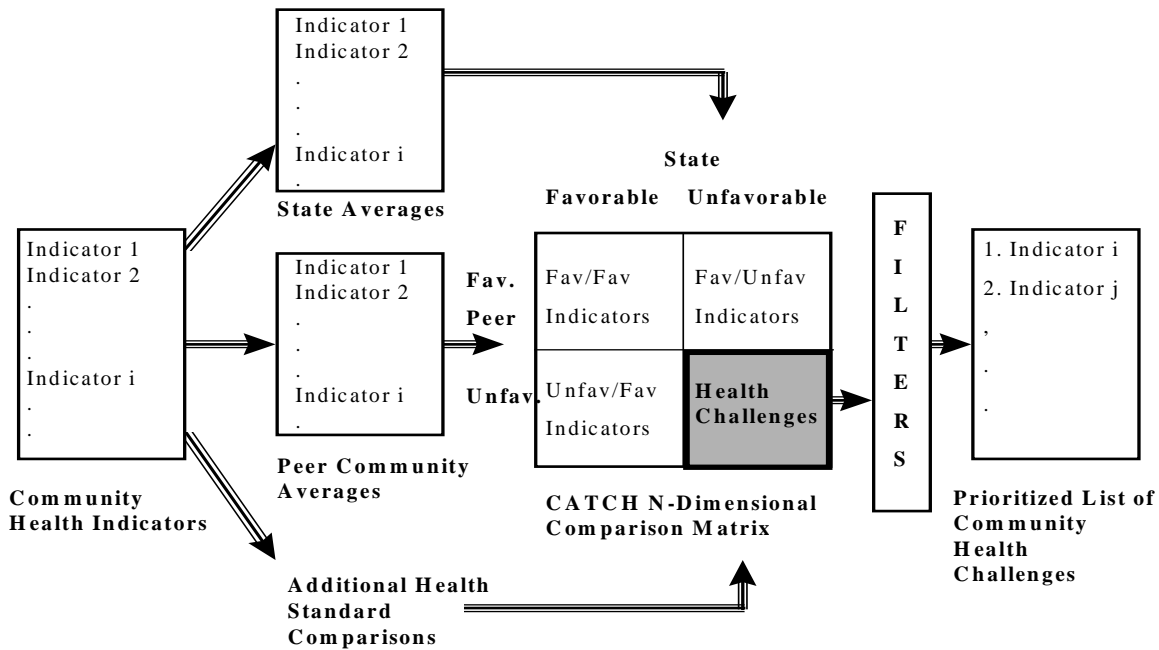

Fig. 1. The CATCH Methodology

This set of health challenges are prioritized by passing each indicator through a set of ranking filters:

➢ Number Affected – Number of persons in the community affected by the indicator.
➢ Economic Impact – An estimate of the direct cost per case for individuals affected by the indicator.
➢ Availability of Efficacious Intervention – An estimate of the relative degree to which treatment or prevention is likely to be effective.
➢ Magnitude of Difference – The degree to which the community indicator is worse than the dimensional comparisons.
➢ Trend Analysis – For a five-year period is the trend favorable or unfavorable and what is the magnitude of change in the trend direction.

The community stakeholders are given an opportunity to weight the importance of each of the above filters. The final product of the CATCH methodology is a comprehensive, prioritized listing of community health care challenges. A more detailed description of the CATCH methodology with a complete listing of health care indicators can be found in [13].

III. CATCH DATA WAREHOUSE

A. Limitations of Manual CATCH

While the value of CATCH is incontrovertible, the ultimate deployment of CATCH throughout Florida and the nation has been constrained by several serious limitations:

➢ The handcrafted process is labor-intensive and slow. Hundreds of individual sources of data must be identified and contacted. Data are often provided in hard copy formats and must be manually checked, validated, and entered into spreadsheets. With current methods, it takes 3 to 4 months to complete a CATCH report for a single county.
➢ Longitudinal trend analyses over many years are cost prohibitive for most communities. Since each application is expensive and time-consuming, the capability to fund and produce annual assessments in a single community is limited.
➢ Most public health funding comes from state and federal governments. A statewide CATCH assessment would help to prioritize funding and serve to enable effective program evaluation based on quantifiable outcomes assessment. Since nearly all data elements available in Florida are available in most other states, there is reason to be confident that CATCH might be expanded nationally and even internationally.
➢ With the massive amount of health care data involved, many interesting relationships and correlations of health status indicators can be found

and investigated. Currently, in the manual system such discovery is not being done.

B. CATCH Data Warehouse Challenges

A CATCH data warehouse has been constructed to overcome these limitations, enabling both cost-effective report generation and ad-hoc analyses of critical health care issues. The construction of a data warehouse for public health care data poses major challenges beyond that required for the construction of a commercial data warehouse (e.g., retail sales). Such challenges include:

➢ Data come from a diverse set of sources. Health care data are published in a wide variety of formats with differing semantics. There are currently few standards in the health care field for data. The data integration task to build the data warehouse requires significant effort.
➢ CATCH reports are disseminated to a diverse and geographically distributed set of stakeholders. The next section discusses the different dissemination modes that must be accommodated by the data warehouse.
➢ The data warehouse is required to support the activities of public policy formulation. The socio-political issues of health care policy impact design features such as security, availability, data quality, and performance.

C. Data Warehouse Design

Important missions of a data warehouse include the support of decision-making activities and the creation of an infrastructure for ad-hoc exploration of very large collections of data. Decision-makers should be able to pursue many of their investigations using browsing tools, without relying on database programmers to construct queries. The emphasis on end-user data access places a premium on an understandable database design that provides an intuitive basis for navigating through the data. The star schema or dimensional model has been recognized as an effective structure for organizing many data warehouse components [6]. The star schema is characterized by a center fact table, which contains numeric information that can be used in summary reports. Radiating from the fact table are dimension tables that provide a rich query environment. This structure provides a logical data cube, with dimensions such as time and location identifying a set of numeric measurements within the cube. To illustrate, Fig. 2 contains a fragment from the hospital discharge star schema that is part of the CATCH data warehouse.

The most appropriate facts are *additive* numeric data items that can be summed, averaged, or combined in other ways to form summary statistics. The only way to
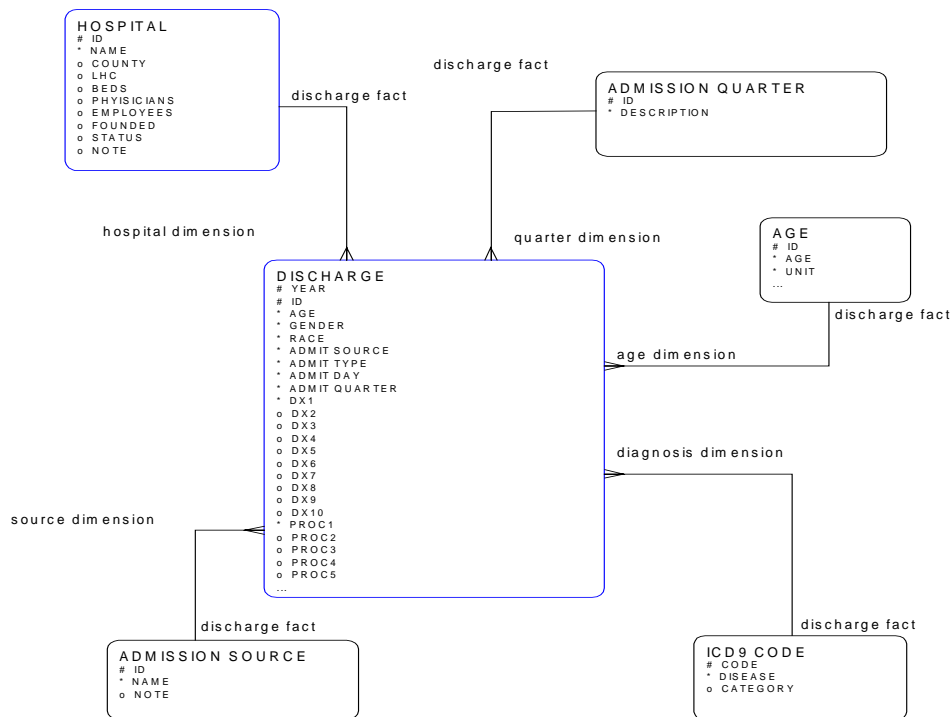
Fig. 2. Hospital Discharge Star Schema

"compress" the millions of transaction items is to present some mathematical summarization. No human will want thousands, let alone millions, of items in answer to their queries. As Kimball points out, "The best and most useful facts are *numeric*, *continuously valued*, and *additive*" [10].

The mission of the CATCH data warehouse is to support the automated and cost-effective application of the CATCH methodology, as well as to enable more detailed analyses that were not possible using the coarse-grained data that typified past CATCH reports. In order to meet these goals, the data warehouse design includes several levels of data granularity, from the coarse-grained data used in generic report production to actual event-level data, such as hospital discharges. The data warehouse includes major components at three levels of granularity.

1. Reporting tables with highly aggregated data are used to support the core CATCH reports, including comparisons between the target county and peer counties. These tables also provide fast interactive response for interactive access via data browsing tools and can provide the foundation for simple community-wide Internet access.

2. There are families of star schemas that provide true dimensional data warehouse capabilities, such as

interactive roll-up and drill-down operations. These components have carefully designed dimensions that can be utilized by more sophisticated data browsing tools. The star schemas are populated using thorough data staging and quality procedures that usually involve processing detailed data sets extracted by various health care agencies and organizations. Typically, the data is aggregated and transformed for loading into a family of related star schemas that share important dimensions and support interactive online analytic processing (OLAP) techniques.

3. For certain types of information, the design calls for retaining very fine-grained or even event level data. An example is the hospital discharge data that includes each hospital discharge event for the more than 200 hospitals that are mandated to report such information in Florida. This data is retained at the transaction level because of the rich set of facts and dimensions available for analysis and the density of potential aggregations that result in negligible space savings.

These three levels of aggregation within the data warehouse combine to meet a wide range of reporting requirements and performance goals. Thus providing a flexible basis for disseminating health care information to community decision-makers.

IV. Modes of CATCH Data Dissemination

The human-computer interface is of paramount importance in the data warehouse environment and the primary determinant of success from the end-user perspective. In order to support analysis and reporting tasks, the data warehouse must have high quality data and make that data accessible through intuitive interface technologies. The act of releasing data in a warehouse is in a very real sense the same as publishing that data in printed form, retractions in both media can be very painful. Once the data becomes accessible, it may be included in reports, forecasts, and analyses that form the basis of decision-making activities within an organization or community. Therefore, data staging and quality procedures within the data warehouse are often among the most expensive and critical ingredients in providing a successful end-user experience.

The types of access in a data warehouse can be broadly categorized as *navigation* or *summarization* tasks. Navigation activities include data browsing, ad-hoc queries, and traditional report generation. These tasks require human guidance and design to produce the appropriate queries, often presenting the results in tabular or graphical form. Though online analytic processing (OLAP) usually incorporates roll-up/drill-down features, the navigation style is highly interactive and driven by previous steps in data exploration.

Summarization tasks are algorithmic in nature, applying techniques that summarize patterns in the data and usually produce models, often with some notion of reliability, which can be used to predict as well as describe the underlying data. Traditional statistics and data mining techniques are often used as summarization tools. A distinction is drawn between their use as *exploratory* or *confirmatory* methods, but the results are a model or set of abstract patterns that can be applied to other data sets. For example, connectivity to statistical packages is an important interface component that allows analysts to use statistical techniques to confirm or more fully investigate interesting properties discovered through browsing in the CATCH data warehouse. While these techniques are clearly important and applicable to health care data warehousing, the following discussion focuses on the navigation tools and more traditional database access technologies being utilized in the project.

A. Data Browsing

Data warehouse browsing tools provide star query-like access through a flexible menu-based interface, with pull-down menus representing important data dimensions. These types of tools are easy to use and support some ad-hoc exploration, but are usually controlled through some sort of administrative layer that determines the data available to end-users. In developing a flexible interface, there is a tradeoff between the ability to express ad-hoc queries and the ease-of-use that results from pre-defined constructs implemented by data warehouse designers and administrators. The interface discussed in the following sections is based on Oracle Discoverer, which includes both an administrative tool for controlling data access and an end-user tool that allows interactive creation and customization of the reports. Of course, many reports are implemented through the normal development process and made available for use.

As noted in the data warehouse design discussion, the CATCH data warehouse consists of several levels of granularity from transaction-oriented data, such as hospital discharges, to summary data at the CATCH report level. Therefore, the interface requirements differ for each of the major components, especially with regard to the role of browsing tools. For instance, the browsing tools provide a convenient method for CATCH analysts to view the preliminary report results with more detailed information than most community planners would want to sift through. Fig. 3 shows some of the indicators for a given county over a two year period together with a preliminary view of a target-to-peer comparison for those indicators. Final report components may be generated using the browsing tools, or more likely be implemented as part of a reporting function that more fully automates the process.

A second and in some ways more important role for the browsing tools is to provide a flexible interface for more customized analysis. Health care issues highlighted by the CATCH methodology can be investigated more fully using the finer levels of detail maintained in the data warehouse. These tasks might entail querying the true dimensional star schemas that include age, gender, race, and other dimensions, or even the event-oriented data, such as hospital discharges. Thus the data warehouse allows the user to focus on issues such as differences in age or race with regard to specific health status indicators. Once decision-makers review the CATCH report, they may have specific issues that relate to the diverse communities that inevitably fall outside of arbitrary political boundaries. Fig. 4 illustrates a detailed browsing screen in which specific volume, length of stay, and cost data are presented for a specific hospital for specific diseases. It is clear how a hospital could effectively use this data for in-depth analyses of utilization and cost/profit ratios.

The browsing tools also offer graphing capabilities that provide simple visualization capabilities. Fig. 5 graphs the length of stay data for a group of infectious diseases for a particular hospital.

Fig. 3. Browsing Screen for Community Indicators

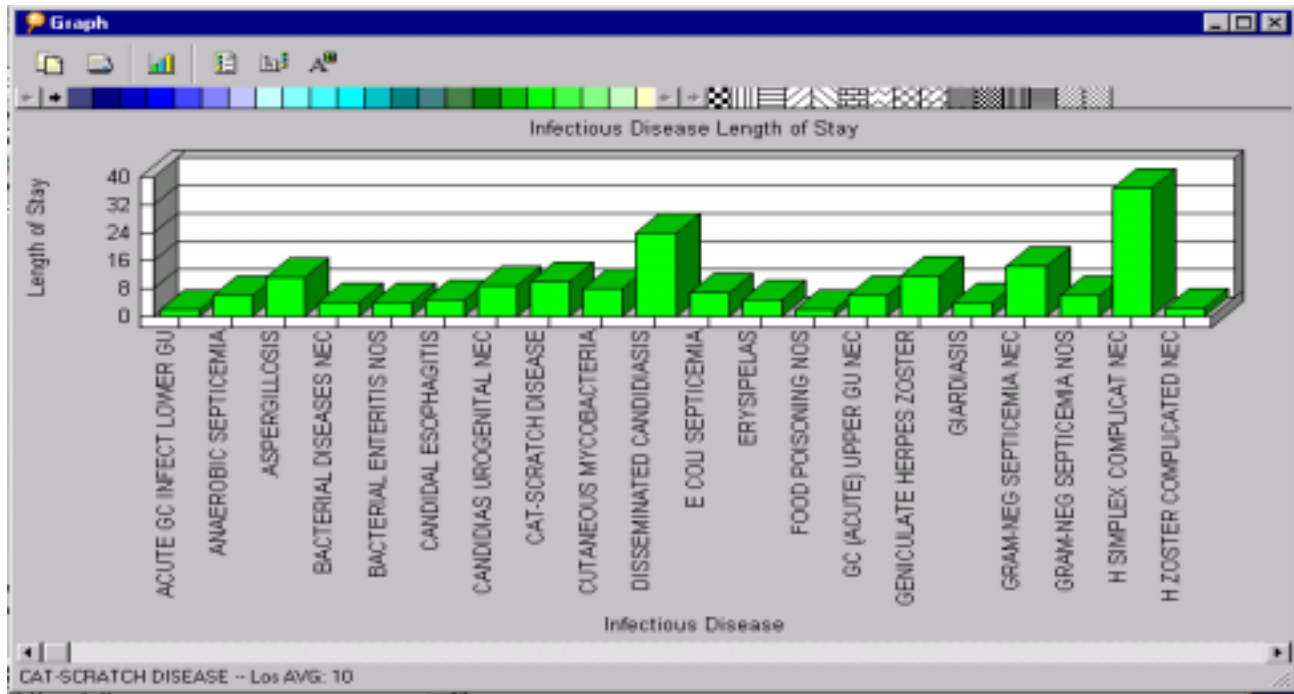Fig. 4. Browsing Screen for Hospital Disease Indicators

Fig. 5. Infectious Diseases vs. Length of Stay in a Hospital

B. Report Generation

It is clear how the data tables and graphs from the browsing tools can be incorporated into comprehensive community health assessment reports. Reports allow quick and easy access to comprehensive summaries and more detailed collections of information from the data warehouse. This type of pre-defined and thorough reporting is critical for implementing a more automated CATCH methodology. For example, the comparison of target counties to peer counties, as well as the state, are fundamental components of the original CATCH reports and important tools for community health care planners.

C. Ad-Hoc Queries

Free-form queries formulated using Structured Query Language (SQL) provide a flexible ad-hoc query capability for the more advanced user. This basic access mechanism is a standard relational database access path, but requires some care in the data warehouse environment. Very large tables and ill-formed queries can conspire to produce some truly awful performance. Currently, the development team has been the most prevalent user of SQL in writing the procedures for constructing the data warehouse, as well as providing queries and views for use by end-user tools.

D. Internet Access

Security issues, as well as a primary focus on research and development, have led to a conservative policy with regard to Internet access to the CATCH data warehouse. However, the use of the Internet to deliver information to local health planners is an important capability for the future. There are two ways to incorporate this capability within the data warehouse.

1. The first method for using the Internet is to save artifacts created by the research team in a format that allows delivery via the Internet. Many of the current tools have embedded support for this approach. The CATCH methodology has traditionally been centered on a large hardcopy report, so much of this content could be re-created in web-friendly form and easily disseminated to local health planners. The advantage in this approach is the continued role of a strong methodology, rather than simply distributing raw data with no guidance in how to apply analytic methods.

2. A second approach is to provide dynamic access to the data warehouse via the Internet and allow direct queries by a larger community of end users. This approach will almost certainly have a role in the future, but the project will move cautiously in this direction. Most data warehouse vendors are moving to support web-enabled data warehouses, so these

types of tools will easily integrate into the current framework. Reliance on a web-enabled tool set will minimize the need for customized web development and allow the focus of the project to remain on the content and evolution of a comprehensive community assessment methodology.

V. COMMUNITY DECISION MAKING WITH CATCH DATA

The CATCH data warehouse will result in widespread distribution of data previously unavailable to most communities, as well as, on-line access for specialized inquiry. Many issues arise as to how the communities will make most effective use of the CATCH data for health care decision making. This is an area with considerable research potential.

There is a rich literature on the decision making process both with and without information technology. The study of group decision support systems and environments has a strong tradition in the management information systems field [4]. In many ways, this important body of work is appropriate to health care decision making which is usually group oriented. For example, the research in [5] studies the effects of minority influence on decision making and finds that the presence or absence of technology has very different effects. Another important contributing area would be the political process and its ramifications to decision making [11]. Certainly, policy making in health care is very much a political process.

The use of the CATCH methodology and the state-of-the-art data warehousing technology across many Florida communities will provide a rich research opportunity for studying many interesting issues on group decision making in community health care organizations. Some of the issues we plan to study include:
➤ The presence of a champion for specific actions.
➤ The size and make-up of the decision making group.
➤ The speed of the decision making process
➤ The stakeholders in the process and their influence in the decision-making.
➤ Resource constraints that are faced by the community.
➤ The political nature of the process.
➤ The differential accesses to data among the communities.
➤ Information exchange patterns and practices.
➤ The ease of access and usefulness of the data.
➤ The presence of more thorough and structured data via the CATCH methodology.
➤ The ability to produce customized analyses via the CATCH data warehouse.

The complexities of each issue and the interrelationships among these issues make the design of research studies both a challenge and an opportunity. We plan to disseminate the CATCH information to communities in Florida during 2000. Research on health care decision making will study the communities' use of the CATCH information for health care planning.

REFERENCES

[1] Center for Disease Control and Prevention, *Principles of Community Engagement*, Atlanta, 1997.
[2] D. Chrislip and C. Larson, *Collaborative Leadership: How Citizens and Civic Leaders Can Make a Difference*, San Francisco, Jossey-Bass, 1994.
[3] S. Cropper, "Collaborative Working and the Issue of Sustainability," In: C. Huxham, editor, *Creating Collaborative Advantage*, London: SAGE Pub., 1996.
[4] A. Dennis, "Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information But You Can't Make It Think," *MIS Quarterly*, 20(4), 1996.
[5] A. Dennis, K. Hilmer, and N. Taylor, "Information Exchange and Use in GSS and Verbal Group Decision Making," *Journal of MIS*, 14(3), Winter 1998.
[6] P. Gray and H. Watson, *Decision Support in the Data Warehouse*, Prentice Hall, Englewood Cliffs, New Jersey, 1998.
[7] Institute of Medicine, "Summary of Recommendations," In: W. Waterfall, editor, *The Future of Public Health*, Washington DC, National Academy Press, 1988.
[8] Institute of Medicine, "Healthy Communities: New Partnerships for the Future of Public Health," Washington DC, National Academy Press, 1996.
[9] Institute of Medicine, "Measurement Tools for a Community Health Improvement Process," In: J. Durch, L. Bailey, M. Stoto, editors, *Improving Health in the Community, a Role for Performance Monitoring*, Washington (DC): National Academy Press; 1997
[10] R. Kimball, *The Data Warehouse Toolkit*, John Wiley & Sons, New York, 1996.
[11] H. Mintzberg, *The Nature of Managerial Work*, New York, Harper and Row, 1973.
[12] H. Nakajima, "Editorial: New Players for a New Era," *World Health*, 50(3): 3, 1997.
[13] J. Studnicki, B. Steverson, B. Myers, A. Hevner, and D. Berndt, "Comprehensive Assessment for Tracking Community Health (CATCH)," *Best Practices and Benchmarking in Healthcare*, September/October: 196-207, 1997.