1-1-2000

A Decision Technology System To Advance the Diagnosis and Treatment of Breast Cancer

Guisseppi A. Forgionne
University of Maryland - Baltimore County

Aryya Gangopadhyay
University of Maryland - Baltimore County

Monica Adya
Marquette University, monica.adya@marquette.edu

Published version. "A Decision Technology System To Advance the Diagnosis and Treatment of Breast Cancer," in *Managing Healthcare Information Systems with Web-Enabled Technologies*. Ed. Lauren B. Eder. Hershey PA: IGI Global, 2000: 141-150. Permalink. © 2000 IGI Global. Used with permission.
Monica Adya was affiliated with the University of Maryland - Baltimore County at the time of publication.

Chapter X

A Decision Technology System to Advance the Diagnosis and Treatment of Breast Cancer

Guisseppi A. Forgionne, Aryya Gangopadhyay and Monica Adya
University of Maryland Baltimore County

INTRODUCTION

Geographical variations in cancer rates have been observed for decades. Described spatial patterns and trends have provided clues for generating hypotheses about the etiology of cancer. For breast cancer, investigators have demonstrated that some variation can be explained by differences in the population distribution of known breast cancer risk factors such as menstrual and reproductive variables (Laden, Spiegelman, and Neas, 1997; Robbins, Bescianini, and Kelsey, 1997; Sturgeon, Schairer, and Gail, 1995). However, regional patterns also may reflect the effects of Workshop on Hormones, Hormone Metabolism, Environment, and Breast Cancer (1995): (a) environmental hazards (such as air and water pollution), (b) demographics and the lifestyle of a mobile population, (c) subgroup susceptibility, (d) changes and advances in medical practice and healthcare management, and (e) other factors. To accurately measure breast cancer risk in individuals and population groups, it is necessary to singly and jointly assess the association between such risk and the hypothesized factors.

Various statistical models will be needed to determine the potential relationships between breast cancer development and estimated exposures to environmental contamination. To apply the models, data must be assembled from a variety of sources, converted into the statistical models' parameters, and delivered effectively to researchers and policy makers. A Web-enabled decision technology system can be developed to provide the needed functionality.

This chapter will present a conceptual architecture for such a decision technology system. First, there will be a brief overview of a typical geographical analysis. Next, the chapter will present the conceptual Web-based decision technology system and illustrate how the system can assist users in diagnosing and treating breast cancer. The chapter will conclude with an examination of the potential benefits from system use and the implications for breast cancer research and practice.

BACKGROUND

Environmental contaminants tend to be clustered geographically (Chen, 1996; Grimson and Oden, 1996; Oden, Jacquez, and Grimson, 1996). Studies that link breast cancer with environmental factors, then, have focused on geographical analysis.

In a typical geographic analysis, the researcher will utilize clustering techniques and other statistical methodologies to identify abnormal concentrations of breast cancer. Abnormality usually is defined as a concentration that deviates significantly from expected patterns. Sometimes regression and other multivariate statistical analyses will be used to isolate the independent and joint effects of environmental and other variables on breast cancer incidence and mortality. Customarily, the statistical analyses will be based largely on metric data and be single-equation in form. Other times, qualitative evaluations will be used to test the hypothesized relationships. Ordinarily, these evaluations will rely to some extent on the judgment and experience of the researcher and/or other experts.

Researchers and policy makers recognized that information technology could be used to facilitate the desired geographic analysis. Such recognition and subsequent lobbying efforts resulted in legislation that mandated the development of this technology. The National Institutes of Health's National Cancer Institute (NCI) recently sought a geographic information system (GIS) to support breast cancer studies mandated by public law (Geographic Information System for the Long Island Breast Cancer Study Project, 1998).

Geographic Information System (GIS)

In the GIS sought by the NCI, researchers and other interested parties would utilize computer technology to process available inputs into desired outputs. Inputs would include a database that captures and stores spatial and attribute data for the geographic areas defined by the breast cancer studies. Spatial data, which includes longitude and latitude coordinates that are used to draw on study area maps, can be obtained from state and local government base map files, U.S. Postal Service ZIP Code files, U. S. Geological Survey hydrology data files, and U.S. Bureau of the Census TIGER (Topologically Integrated Geographic Encoding and Referencing) files (Fischer and Nijkamp, 1993). Attribute data, which, among other things, consists of pollution measures, population characteristics, and healthcare provider and patient statistics, can be obtained the U.S. Bureau of the Census, the U. S. Healthcare Financing Administration, the National Health and Nutrition Examination Surveys, state-specific public health data files, and state and federal survey data files on pollution-source emissions.

There is also a model base that contains statistical procedures and location formulas. Some statistical procedures are used to categorize attribute data within the study areas and to calculate summary statistics for the demographic, environmental, healthcare, and other pertinent variables within the areas. Other procedures analyze the data for spatial and temporal patterns and for space-time interactions. A third set of statistical procedures are used to perform selected health-related analyses, including cluster detection of breast cancer incident and mortality, assessment of disease risk from nearby environmental hazards, and genetic activity profiling. Location formulas are used to convert various geographic coordinate systems into alternative forms and to make corresponding map projections.

The decision maker (a scientist or public health official) uses available computer technology to organize the collected spatial and attribute data, structure the study area maps, locate breast cancer incidents on the maps, and simulate spatial, temporal, and spatial-temporal patterns. Spatial and attribute data created during GIS processing can be captured and stored as inputs for future processing.

Processing automatically generates visual displays of the outputs desired by the scientists or public health officials. Outputs include study area maps and associated breast cancer incident and mortality reports. The maps define the boundaries of the study area, give the road and street patterns, identify important landmarks, locate the patterns within the study area, and color the concentrations on the study area roads and streets. Reports list the breast cancer incident and mortality statistics by user-selected pattern categories. The reports also summarize key patterns for the study area, including high-incidence and high-mortality clusters, temporal trends, and spatial-temporal relationships, including key incidence and mortality indicators.

The user can utilize the outputs to guide further GIS processing before exiting the system. Typically, the feedback will involve sensitivity analyses in which the user extends or modifies the study area boundaries and observes the effects on breast cancer concentrations and on key spatial and temporal indicators.

MAIN THRUST OF THE CHAPTER

There are several issues, controversies, and problems associated with the NCI GIS concept. An analysis of these issues recommends a Web-based technology solution.

Issues, Controversies, and Problems

Much of the relevant spatial and attribute data needed for the geographic studies are captured and stored on a variety of distributed data sets. In addition, these data are not organized into the variables needed to perform the breast cancer studies. Moreover, required data are collected, captured, and recorded in various formats, and often in an incomplete manner. There is little (if any) sharing of information between the collection sources and the NCI-supported geographic breast cancer study process. To be useful for this process, then, captured data must be located, accessed, and converted into the appropriate format for the analyses and evaluations.

After the relevant data formats have been created, researchers and other analysts will use data mining methods to manually identify spatial and temporal data patterns (Abraham and Roddick, 1998; Davidson, Henrickson, Johnson, Myers, and Wylie, 1999). These methods, however, will not explain the specific environmental and other potential causes of the detected breast cancer patterns. A statistical model will be needed to determine the pattern determinants and their relationship to mortality and morbidity (Borok, 1997; Burn-Thornton and Edenbrandt, 1998). Researchers and analysts must use their insights, analytical and creative abilities, knowledge and experience to formulate the model, test hypotheses, and derive explanations. As new data become available, the models should be updated in terms of parameters and variables. A high level of expertise will be required to perform the model updating and derive the pattern explanations (Hornung, Deddens, and Roscoe, 1998; Makino, Suda, Ono, and Ibaraki, 1999).

All potential users of the prototype GIS are unlikely to possess the necessary analytical and information technology expertise. Even then, the proposed GIS will not have the functionality to perform the required data management, modeling, and presentation tasks.

To achieve the desired functionality, additional information systems will have to be developed and linked to the prototype GIS. Modeling assistance, for example, will require a decision support system (Tan and Sheps, 1998). Another important requirement of such a system is the ease of accessibility of data to users, which would require a wide area intranet or Web-based Internet environment.

Solutions and Recommendations

The NCI can have several stand-alone systems to provide the decision analyses and evaluations (Tan, 1995; Tan and Sheps, 1998). Integrating the stand-alone functions, however, can enhance the quality and efficiency of the segmented support, create synergistic effects, and augment decision-making performance and value (Forgionne and Kohli, 1996). A Web-based information system, called the *Breast Cancer Analysis System (BCAS)*, can deliver such integrated and complete decision support**.** The BCAS will consist of the GIS, an executive information system (EIS), and a decision support system (DSS), with the architecture shown in Figure 1.

Inputs

The BCAS has a database that captures and stores geographic, demographic, environmental, health outcome, and healthcare data. There is also a modelbase, or organized repository of models and algorithms, captures and stores a multiple-equation statistical model along with the GIS, health, and other statistical models required and desired in the National Cancer Institute Specified GIS-H Functions table. As a package, these models compute and estimate breast-cancer-relevant variables, describe and explain the spatial, temporal, and space-time relationships between the variables and breast cancer development, and use the relationships to simulate breast cancer incidence and mortality.

In addition, there is a knowledgebase that captures and stores spatial and temporal profiles linked to breast cancer incidence, development, and mortality patterns. Profile data will be available from the sources identified in the NCI Specified Initial Datasets.

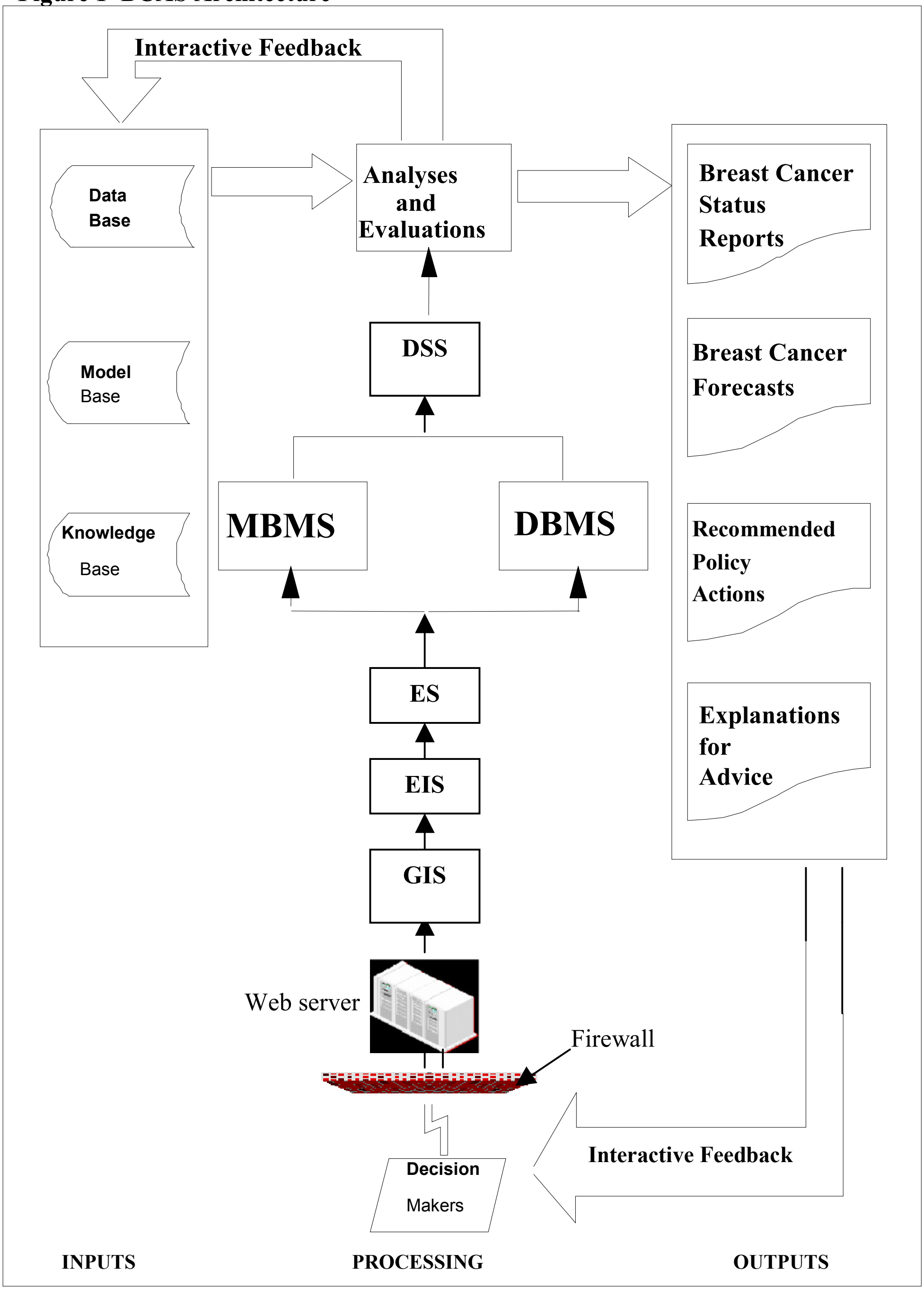
Processing

Interested investigators and officials use the BCAS computer technology interactively to perform analyses and evaluations that include: (a) organizing data into parameters needed for the spatial, temporal, and space-time beast cancer analyses, (b) structuring models that represent and simulate breast cancer development patterns in an integrated and complete manner, and (c) simulating breast cancer incidence and mortality under specified healthcare, health outcome, demographic, and environmental profiles.

Pertinent hardware includes a standard personal computer, Internet (or network) connectivity devices, storage equipment sufficient to accommodate the large volume of spatial and attribute data likely to be processed by the system, and compatible display devices sufficient to show cartographic, interactive visual, map, tabular, plotted, text, and multimedia processing results in hard and soft copy form. Software includes a windowed operating system, a Web-browser, and an integrated application software suite that will accommodate data warehousing, spatial and temporal data analysis, and decision support.

The embedded GIS serves as a front-end to the processing of the profile data. This system extracts the pertinent data from an Internet (or intranet) based data warehouse, links the data to spatial dimensions, and transfers the linked data to the embedded executive

Figure 1 BCAS Architecture



information system (EIS).

An implanted executive information system is used to (a) filter the profile data, (b) form BCAS's database, (c) focus the filtered information, and (d) communicate deviations from expected breast cancer development patterns among affected parties (Keegan and Baldwin, 1992). The EIS also will provide the focused data needed for the DSS analyses and evaluations.

The DSS and Expert Systems (ES) guide the investigator through the intelligent modeling and the modelbase, database, and knowledgebase management needed for the GIS, EIS, and DSS analyses and evaluations. Such embedded expertise assists the user in a virtual manner in performing the epidemiology, statistical, information system, and other tasks required to perform the analyses and evaluations.

BCAS supports Web access to clients using the components shown in Figure 1. Clients request static or dynamic files using a Web browser, which are transmitted through the World Wide Web to a Web listener (an HTTP daemon that runs continuously). The Web listener passes on the client request to the Web server, which sends the results back to the client. Transaction security is supported using secure socket layers and network security is provided through a firewall, shown in Figure 1, using a proxy server.

Outputs

By controlling processing tasks in the desired way, the user can generate: (a) visually attractive tabular and graphic status reports that describe the study area's existing breast cancer incidence and mortality summaries, track meaningful trends, and display important patterns, (b) claims condition and other electronic commerce forecasts, (c) provider policy and payer program simulation results, and (d) recommended claims and other electronic commerce actions. The system also depicts, in a graphic manner, the reasoning (explanations and supporting knowledge) that leads to the suggested actions.

Feedback Loops

Feedback from the processing provides additional data, knowledge, and enhanced decision models that may be useful for future decision activities and tasks (Kalakota, and Whinston, 1997). Output feedback (often in the form of sensitivity analyses) is used to extend or modify the original analyses and evaluations. All processing (including each feedback loop) is done in a user-friendly manner, with artificial intelligence (mainly expert system) technology, that meets the decision styles and requirements of participating parties.

Web-Based Technology

The BCAS will be a Web-based technology. A Web site will be established to collect the pertinent data from the various sources. Utilizing the electronic commerce concept, data suppliers will access the BCAS Web site and select appropriate screen icons (Gerull and Wientzen, 1997). These selections will automatically obtain the data from the supply source and transfer the elements to a data warehouse (Tsvetovatyy, Gini, and Wiecckowski, 1997). A component called *integrator* is responsible for extracting data from the GIS and storing it in the warehouse.

The BCAS's EIS will extract the pertinent data, capture the extractions in user-oriented data marts, and make the marked data available for ad hoc queries by: (a) principal investigators and other scientists; (b) public health officials; (c) private health practitioners;

(d) individual investigators; (e) public interest groups, and (f) the public at large. Ad hoc queries can be made at the users' sites in an easy-to-use, convenient, and interactive manner, utilizing the Web-based BCAS. Results will be displayed in formats (tabular, graphic, and maps) anticipated by the requesting parties.

The system will provide an open and scalable on-line analytical processing (OLAP) solution using the thin-client architecture. The clients generate and view reports and graphs of data stored in the multidimensional database server (MDDB), which derives the raw data from the data warehouse located at the database (DB) server. Users can run queries and generate reports using with any HTML-based browser without having an application session being run on the client. The multidimensional database (MDDB) server is used to enable IT personnel to build, maintain, and optimize multidimensional data. This architecture would reduce the cost of client-server management by centralizing the data and applications on the server, and make it simple to distribute and upgrade applications and add new OLAP users.

FUTURE TRENDS

Web-based environments that provide universal access, via the Internet or an intranet, to integrated health information from a variety of sources is an emerging trend that can benefit the healthcare community and the nation in a variety of ways. Such access enables the healthcare organization to be proactive rather than reactive. For example, through this access, healthcare organizations can process expected outcomes and compare them against actual patient outcomes. Such comparisons can help predict: (a) patient problems; (b) required healthcare interventions; (c) time required for implementation of healthcare services; (d) accessibility of healthcare services; (e) quality of healthcare services, and (f) cost of healthcare services.

The BCAS concept provides a vehicle to utilize the emerging Web-based access for proactive breast cancer diagnosis and treatment. Without the BCAS, breast cancer detection involves a very complex process that requires extensive training for researchers and policy analysts. By decreasing the volume of documentation, by simplifying the educational process, and by simplifying and automating much of the process, the BCAS can be expected to save the medical community millions of dollars per year in breast cancer diagnosis and treatment costs. Based on the experience in a previous, and similar, project, it is anticipated that the BCAS can be developed and implemented for a small fraction of the potential cost savings (Forgionne, 1997).

From a diagnosis perspective, the manual search for breast cancer patterns, morbidities, and mortalities is a tedious process that often results in inaccurate, incomplete, and redundant data. Such data problems can leave cancer victims inadequately diagnosed and treated. With the BCAS, the researcher or analyst identifies all data relevant to the breast cancer analysis process, and the system provides a mechanism that facilitates data entry while reducing errors and eliminating redundant inputs. Reports from the BCAS also offer focused guidance that can be used to help researchers and policy analysts in performing breast cancer searches, diagnoses, and treatments.

Challenges

Realizing the strategic potential will present significant challenges to the traditional healthcare organization. Tasks, events, and processes must be redesigned and reengineered to accommodate the Web-based universal access. Clinicians and administrators must be convinced that the access will be personally as well as organizationally beneficial, and they must agree to participate in the effort. Finally, the organizational changes will compel substantial informational technology support.

When implemented fully, the innovation will alter the work design for, and supervision of, breast cancer diagnosis and treatment. Requisite operations and computations will be simplified, automated, and made error-free. Training requirements will be reduced to a minimum. Processing efficiency will be dramatically increased. User-inspired creative study-area and public health policy experimentation will be facilitated and nurtured. Management learning will be promoted. Knowledge capture will be expedited.

In short, BCAS's usage would substantially reshape the organizational culture. Faced with significant time pressures and limited staff, healthcare leadership may be reluctant to take on this burden at the present time. In addition, public health officials have developed and cultivated strong and enduring relationships with practitioners and vendors. These practitioners and vendors also have important contacts and allies within the government agencies that oversee healthcare programs. For these reasons, it may be politically wise for public health officials to preserve these practitioner and vendor relationships.

Future Research Opportunities

There are a number of future research opportunities presented by the BCAS. To ensure that the information system accurately replicates the inputs, the final version of the BCAS should be tested against Web-collected data from existing institutions study areas. In the testing, warehoused data should be compared against actual values. Statistical tests should be conducted on the estimated models. There should be evaluations of user satisfaction with: (a) the speed, relevance, and quality of ad hoc query results, (b) the system interface, (c) model appropriateness, and (d) the quality of the system explanation. Simulations should be statistically tested for accuracy, and confidence intervals should be established for the results. Tests should also be conducted on the system's ability to improve the decision making maturity of the user.

The BCAS concept can also be adapted for a variety of adjunct healthcare applications. Similar systems can be applied to the diagnosis and treatment of other forms of cancer, mental disorders, infectious diseases, and additional illnesses. Effectiveness studies can be done to measure the economic, management, and health impacts of the additional applications.

Another potential area for future research is the use of artificial intelligence techniques for model calibration and optimization. In a domain such as breast cancer diagnosis and treatment, there will often be a need to optimize models, especially as newer models emerge through BCAS processing. These emerging models should undergo a rigorous process of optimization and calibration. Empirical evidence suggests that genetic algorithms are superior to traditional search and optimization methods (such as hill climbing), and even neural networks, for such tasks (Kao et al., 1995; Kim, 1998).

CONCLUSIONS

The BCAS architecture is based on a combination of database, modeling, data mining, and mapping techniques. Its deployment can enable the NIH to realize significant economic and political benefits. Future enhancements promise to increase the power of the BCAS to further improve the nation's ability to manage its cancer resources.

The BCAS delivers the information and knowledge needed to support breast cancer detection in a comprehensive, integrated, and continuous fashion. In theory, the comprehensive, integrated, and continuous support from the BCAS should yield more decision value than the non-synthesized and partial support offered by any single autonomous system. Improvements should be observed in both the outcomes from, and the process of, strategic claims and other electronic commerce decision making (Lederer, Merchandani, and Sims, 1997). Outcome improvements can include advancements in the level of the users' decision-making maturity and gains in organization performance (Whinston, Stahl, and Choi, 1997). Process improvements can involve enhancements in the users' ability to perform the phases and steps of decision making.

Regardless of BCAS's legacy, the application offers useful lessons for Web-based healthcare decision technology systems development and management. The system is effectively delivering to the user, in a virtual manner, embedded statistical, medical, and information systems expertise specifically focused on the healthcare problem.

The virtual team characteristics and the organizational culture impacts also suggest that a hybrid project-technology organization may work well for Web-based decision technology system design, development, and implementation in a healthcare environment. The organization would be virtual rather than physical. A project team would be established and administered by the practicing healthcare professional. Team technology specialists would be drawn from within and outside the organization to match the expertise needed for the specific project. Telecommuting and distributed collaborative work would be allowed and possibly encouraged.

REFERENCES

Abraham, T., & Roddick, J. F. (1998). Opportunities for knowledge discovery in spatio-temporal information systems. *Australian Journal of Information Systems*. 5(2), 3-12.

Borok, L. S. (1997). Data mining: Sophisticated forms of managed care modeling through artificial intelligence. *Journal of Healthcare Finance*. 23(3), 20.

Burn-Thornton, K. E., & Edenbrandt, L. (1998). Myrocardial infarction-pinpointing the key indicators in the 12-lead ECG using data mining. *Computers and Biomedical Research*, 31(4), 293-303.

Chen, R. (1996). Exploratory analysis as a sequel to suspected increased rate of cancer in a small residential or workplace community. *Statistics in Medicine* 15, 807-816.

Davidson, G. S., Hendickson, B., Johnson, D. K., Meyers, C. E., & Wylie, B. N. (1999). Knowledge mining with VxInsight: Discovery through interaction. *Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*. 11(3), 259-285.

Fischer, M. M. and P. Nijkamp (Eds.) (1993). *Geographic Information Systems, Spatial Modeling, and Policy Evaluation*. New York: Springer-Verlag.

Forgionne, G. A. (1997). HADTS: A decision technology system to support army housing management. *European Journal of Operational Research*, 97, 1120-1135.

Forgionne, G. A. and Kohli, R. (1996). HMSS: A management support system for concurrent hospital decision making. *Decision Support Systems*. 16, 209-223.

Grimson, R. C. and N. Oden (1996). Disease clusters in structured environments. *Statistics in Medicine* 15, 851-871.

Geographic Information System for the Long Island Breast Cancer Study Project (LIBCSP) (1998). National Cancer Institute's Electronic RFP Number NO2-PC-85074-39. Bethesda: National Cancer Institute.

Gerull, D. B. and Wientzen, R. (1997). Electronic commerce: The future of image delivery. *International Journal of Geographical Information Systems*, 7(7) 38-51.

Hornung, R. W., Deddens, J. A., & Roscoe, R. J. (1998). Modifiers of lung cancer risk in uranium miners from the Colorado Plateau. *Health Physics*, 74(1), 12-21.

Huxhold, W. E. (1991). *An Introduction to Urban Geographic Information Systems*. Oxford: Oxford University Press.

Kao, C., Chen, L.-H., Wang, T.-Y., Kuo, S., & Horng, S.-D. (1995). Productivity improvement: efficiency approach vs effectiveness approach. *Omega, International Journal of Managment Science,* 23(2), 197-204.

Kalakota, R. and Whinston, A. B. (1997). *Electronic Commerce: A Manager's Guide*, Reading, Massachusetts: Addison-Wesley.

Keegan, A. J. and Baldwin, B. (1992). EIS: A better way to view hospital trends. *Healthcare Financial Management,* 46(11), 58-64.

Kim, W. (1998). Data mining: Promises, reality, and future. *Journal of Object-Oriented Programming*, 11(4), 61-62+.

Laden, F., Spiegelman, D., and Neas, L. M. (1997). Geographic variation in breast cancer incidence rates in a cohort of U. S. women. *Journal of the National Cancer Institute* 89, 1373-1378.

Lederer, A. L., Merchandani, D. F., and Sims, K. (1997). The link between information strategy and electronic commerce. *Journal of Organizational Computing and Electronic Commerce*, 7(1), 17-25.

Makino, K., Suda, T., Ono, H., & Ibaraki, T. (1999). Data analysis by positive decision trees. *IEICE Transactions on Information and Systems*, E82-D(1), 76-88.

Oden, N., Jacquez, G., and Grimson, R. (1996). Realistic power simulations compare point- and area-based disease cluster tests. *Statistics in Medicine* 15, 783-806.

Regional Variation in Breast Cancer Rates in the U. S. – NIH. National Cancer Institute's Electronic RFA Number CA-98-017. Bethesda: National Cancer Institute, 1998.

Robbins, A. S., Brescianini, S., and Kelsey, J. L. (1997). Regional differences in known risk factors and the higher incidence of breast cancer in San Francisco. *Journal of the National Cancer Institute* 89, 960-965.

Sturgeon, S. R., Schairer, C., and Gail, M. (1995). Geographic variation in mortality rates from breast cancer among white women in the United States. *Journal of the National Cancer Institute* 87, 1846-1853.

Tan, J. K. H. (1995). *Health Management Information Systems.* Gaithersburg, Maryland: Aspen.

Tan, J. K.H., and Sheps, S., Editors (1998). *Health Decision Support Systems.* Gaithersburg, Maryland: Aspen.

Tsvetovatyy, N., Gini, M., and Wieckowski, Z. (1997). Magma: An agent-based virtual market for electronic commerce. *Applied Artificial Intelligence.* 11(6), 501-509.

Whinston, A. B., Stahl, D. O., and Choi, S. (1997). *The Economics of Electronic Commerce.* Indianapolis, Indiana: Macmillan Technical Publishing.

Workshop on Hormones, Hormone Metabolism, Environment, and Breast Cancer, New Orleans, Louisiana, September 28-29, 1995. *Monographs in Environmental Health Perspectives* supplement 1997, 105(3), 557-688.