2001

BCAS: A Web-enabled and GIS-based Decision Support System for 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

Joseph K. Tan
University of British Columbia Faculty of Medicine

BCAS: a web-enabled and GIS-based decision support system for the diagnosis and treatment of breast cancer

Guisseppi A. Forgionne,
Aryya Gangopadhyay and Monica Adya

Information Systems Department, University of Maryland Baltimore County, 1000 Hilltop Circle, Catonsville, MD 21250, USA
Fax: 410-455-1073 E-mail: forgionn@umbc7.umbc.edu

Joseph K. Tan

Department of HealthCare & Epidemiology, University of British Columbia, Faculty of Medicine, 5804 Fairview Avenue, Vancouver BC V6T 1Z3, Canada
Fax: (604) 822-4994 E-mail: josepht@interchange.ubc.ca

Abstract: For decades, geographical variations in cancer rates have been observed but the precise determinants of such geographic differences in breast cancer development are unclear. Various statistical models have been proposed. Applications of these models, however, require that the data be assembled from a variety of sources, converted into the statistical models' parameters and delivered effectively to researchers and policy makers. A web-enabled and GIS-based system can be developed to provide the needed functionality. This article overviews the conceptual web-enabled and GIS-based system (BCAS), illustrates the system's use in diagnosing and treating breast cancer and examines the potential benefits and implications for breast cancer research and practice.

Keywords: Geographic information systems (GIS); breast cancer detection; decision support; web-enabled technology.

Reference to this paper should be made as follows: Forgionne, G.A., Gangopadhyay, A., Adya, M. and Tan, J.K. (2001) 'BCAS: a web-enabled and GIS-based decision support system for the diagnosis and treatment of breast cancer', *Int. J. Healthcare Technology and Management*, Vol. 3, Nos. 5/6, pp.424–443.

Biographical notes: Guisseppi A. Forgionne is Professor of Information Systems at the University of Maryland Baltimore County (UMBC). Professor Forgionne holds a BS in Commerce and Finance, an MA in Econometrics, an MBA and a PhD in Management Science and Econometrics. He has published 24 books and approximately 125 research articles and consulted for a variety of public and private organisations on decision support systems theory and applications. Dr. Forgionne has also served as department chair at UMBC, Mount Vernon College and Cal Poly Pomona. He has received several national and international awards for his work.

Aryya Gangopadhyay is an Assistant Professor of Information Systems at the University of Maryland Baltimore County (USA). His research interests include Electronic Commerce, multimedia databases, data warehousing and mining and geographic information systems. He has authored and co-authored two books, book chapters and numerous papers in journals.

Monica Adya received her PhD in Information Systems from Case Western Reserve University. Her research interests relate to the application of Artificial Intelligence to various business functions. In particular, her work has focused on the use of such systems for forecasting and prediction. Recently she has also examined the use of intelligent DSS in supporting healthcare issues, defence contracting and digital libraries. Monica has published in *Information Systems Research, The International Journal of Forecasting* (IJF) and *The Journal of Forecasting*. She also serves on the editorial board of IJF.

Joseph K.H. Tan, PhD is Associate Professor of Health Informatics, Department of HealthCare and Epidemiology, Faculty of Medicine, the University of British Columbia, Canada. He has served as an editor and co-editor of special journal issues, as a member of editorial boards for a number of health information and business journals and as a conference coordinator, track and session chair and reviewer of various academic and professional associations at local, national and international levels. He is well published in the field of health management information systems and health decision support systems. His most recent book is the 2nd edition of *Health Management Information Systems: Methods and Practical Applications* published by Aspen Publishers, Inc.

1 Introduction

For decades, geographical variations in cancer rates have been observed. Detected spatial patterns and trends have provided clues for generating hypotheses about the etiology of cancer. Regional patterns, for example, may reflect the effects of environmental hazards (such as air and water pollution), demographics and the lifestyle of a mobile population, subgroup susceptibility, changes and advances in medical practice and healthcare management and other factors. Specifically for breast cancer, investigators have demonstrated that some variation can be explained by differences in the population distribution of known breast cancer risk factors, for example, menstrual and reproductive variables [1-3]. Yet, the precise determinants of geographic differences in breast cancer development are unclear. Consequently, it is not yet possible to determine conclusively whether spatial variation in the environment contributes to elevated breast cancer incidence and mortality.

Biological data relating environmental pollutants to breast cancer are sparse and epidemiological studies have been challenged by methodological limitations (most often in determining past exposure levels). To date, the scientific literature on the association of measurable exposures (for example, organochlorine pesticides) and breast cancer is conflicting [4]. Endogenous hormones are known to be important in breast cancer development. A leading question is whether environmental factors (such as xenoestrogens and other hormone-mimicking pollutants) also exert an effect on such development [5]. The power of molecular and bioinformatic technology could potentially provide biologic probes and sensitive methods for epidemiologic studies to gain insights into the relationship between the environment, the individual and breast cancer. To measure accurately breast cancer risk in individuals and population groups, it appears necessary therefore singly and jointly to assess the association between such risk and the hypothesised factors.

Various statistical models are needed and have been suggested to determine the potential relationships between breast cancer development and estimated exposures to environmental contamination. Applications of these models, however, require that 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 article presents the conceptual architecture for such a web-based decision technology system. First, we will discuss the background information on a typical geographical analysis. Next, we present the conceptual web-based decision technology system and illustrate how the system can assist users in diagnosing and treating breast cancer. Finally, we will conclude with an examination of the potential benefits and challenges from the system user perspective and the implications for breast cancer research and practice.

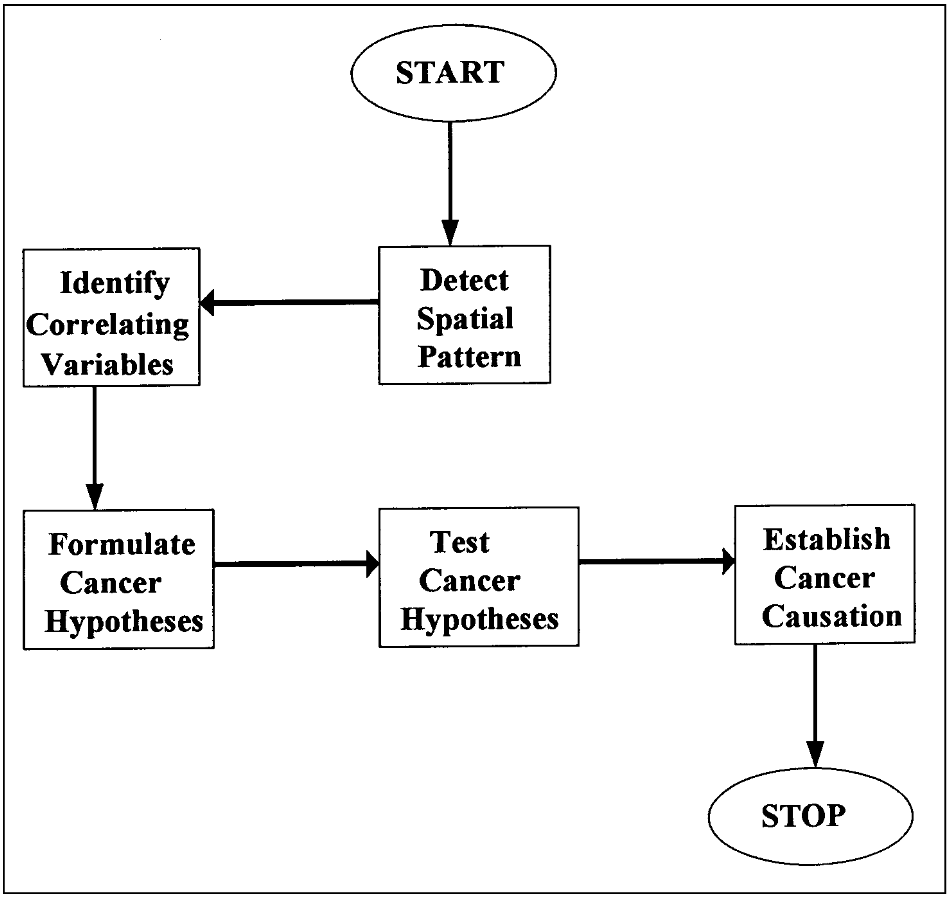
2 Background

Environmental contamination is geographic in nature. Studies that link breast cancer with environmental factors, then, have focused on geographical analysis, that is, clustering contaminants geographically [6-8]. The analytical process is summarised in Figure 1.

In a typical geographic analysis, clustering techniques and other statistical methodologies are used to identify abnormal concentrations of breast cancer. Abnormality is defined by a concentration that deviates significantly from expected patterns. Often, the statistical analyses will be based largely on metric data and be single-equation in form. 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. At other times, qualitative evaluations will be used to test the hypothesised relationships. Regardless, these evaluations will rely to some extent on the judgement and experience of the researcher and/or other experts.

Researchers and policy makers recognised the need for advancing information technology (IT), for example, the internet and related decision support technologies, to facilitate the desired geographic analysis. Such recognition has resulted in legislation that mandated the development and application of associated technologies. The National Institute of Health's National Cancer Institute (NCI), for example, recently sought a geographic information system (GIS) to support breast cancer studies mandated by public law [9]. Although this GIS has been described only in general terms through a Request for Proposal (RFP), it is possible from the RFP to deduce the desired system requirements.

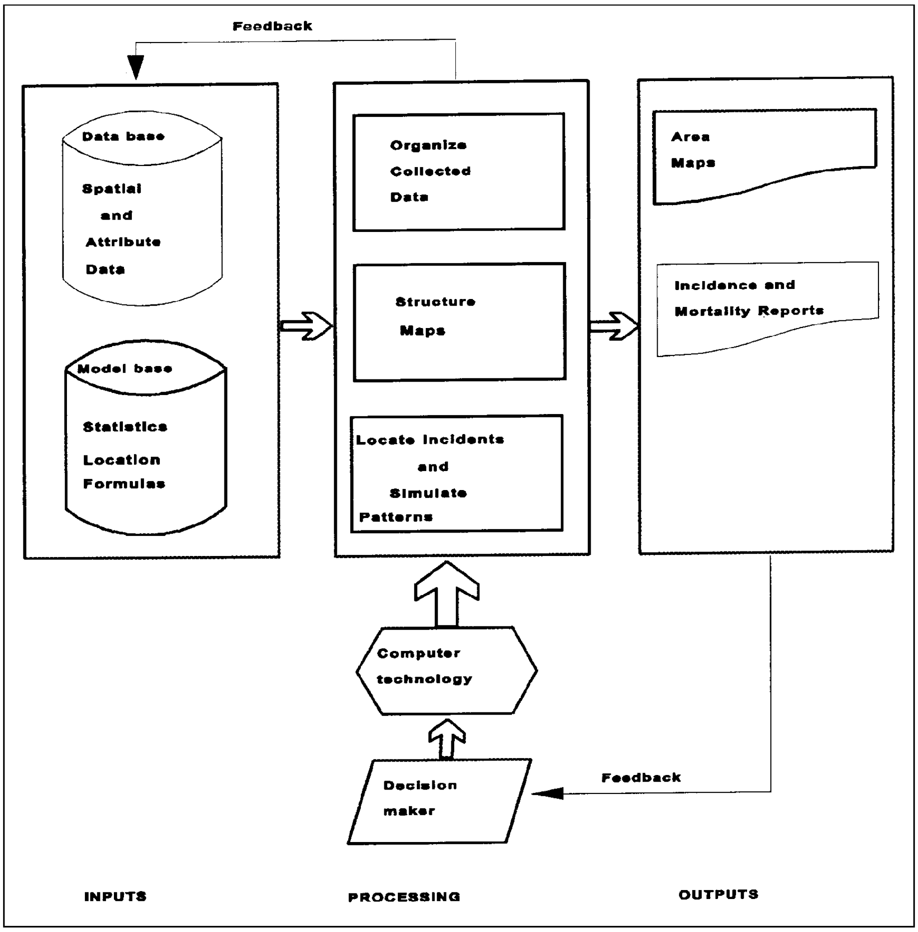
Figure 1 Breast cancer geographic analysis process



2.1 Geographic Information System (GIS)

The GIS sought by the NCI can be conceptualised architecturally. As shown in Figure 2, researchers and other interested parties would utilise computer technology to process available inputs into desired outputs.

Figure 2 NCI GIS



2.1.1 Inputs

The prototype GIS would have a database that captures and stores spatial and attribute data for the geographic areas defined by the breast cancer studies. Spatial data would include longitude and latitude coordinates that are used to draw the following features on study area maps:

- county and census block boundaries;
- ZIP code polygons;
- bodies of water;
- streets and highways;
- base-maps and other geographic frameworks; and
- geo-coded population-based cancer incidence and mortality.

These spatial data can be obtained from state and local government base map files, US Postal Service ZIP Code files, US Geological Survey hydrology data files and US Bureau of the Census TIGER (Topologically Integrated Geographic Encoding and Referencing) files [10].

Attribute data consist of variables such as contaminated drinking water, sources of indoor and ambient air pollution, including emissions from aircraft, electromagnetic fields (EMF), pesticides and other toxic chemicals, hazardous and municipal waste, statistical characteristics of the population, hospital, physician and other healthcare provider statistics and consumer demand and access to healthcare services. Population data can be drawn from census blocks from the US Bureau of the Census. Health-related demographic data can be acquired from the US Health Care Financing Administration (state-specific Medicare data) and the National Health and Nutrition Examination Surveys (prevalence of disease, distribution of physiological and psychological measurements and nutritional status in the general population). Health outcome and care data can be obtained from state-specific public health data files. Environmental and other data can be acquired from state and federal survey data files on water quality, pollutants, toxic waste, ambient air and source emission, air quality, radiation, powerline and chemical usage and waste generation.

The NCI GIS would have a model base that contains statistical procedures and location formulas. Some statistical procedures would be used to categorise 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 would analyse the data for spatial and temporal patterns and for space-time interactions. A third set of statistical procedures would be used to perform selected health-related analyses, including cluster detection of breast cancer incidence and mortality, assessment of disease risk from nearby environmental hazards and genetic activity profiling.

Location formulas would be used to convert various geographic coordinate systems into alternative forms and to make corresponding map projections. These formulas can involve conversions from State Plane, Universal Transverse Mercator and Lambert Conformal Conic Projection coordinates into the latitude/longitude global reference system for various units of measurement and storage formats.

2.1.2 Processing

The decision maker (a scientist or public health official) would use available IT to organise the collected spatial and attribute data, structure the study area maps, locate breast cancer incidence on the maps and simulate spatial, temporal and spatial-temporal patterns.

Among other activities, GIS processing tasks will therefore entail address matching with TIGER files, sensitivity analyses for aggregation effects, area interpolation, buffering, computing confidence intervals for age-adjusted disease rates, generating density equalising map projections, analysing for spatial disease diffusion, estimating small area population demographics, disease cluster investigations, disease choropleth and isopleth mapping, integrating geographical and epidemiologic analyses, geographic scale analyses, geographic masking, region building, grid point and case geographic analyses, proximity analyses, space searching for boundary creation, spatial data, time series and space-time pattern analyses, creating general, point pattern and other smoothing maps, spatial filtering, spatial and temporal clustering and geographical analysis matching, spatial interpolation, significance mapping, uncertainty specification, GIS computational functions (such as angle determination between three points and perimeter determination for a polygon), conversion of geo-referenced input datasets into the latitude/longitude global reference system,and map overlaying and modification of area units [11].

The GIS can be designed to execute on a standard personal computer (PC) with graphics capabilities. It can run under any GIS tool through a windowed operating system. Such a configuration offers a consistent, less time-consuming, inexpensive and more flexible development and implementation environment than the available alternatives.

As indicated by the top feedback loop in Figure 2, while preserving the original data elements, spatial and attribute data created during GIS processing can be captured and stored as inputs (in the form of additional or derived fields and records) for future processing, thereby updating the database dynamically. The user would execute the functions by selecting pushbuttons, pull-down menu options, list box items, radio buttons choices and checkbox options on attractive visual displays that make the computer processing virtually invisible (transparent) to the user.

2.1.3 Outputs

Processing automatically would generate visual displays of the outputs desired by the scientists or public health officials. Outputs would include study area maps and associated breast cancer incidence and mortality reports. The maps would define the boundaries of the study area, give the road and street patterns, identify important landmarks, locate the patterns within the study area and colour the concentrations on the study area roads and streets. Reports would list the breast cancer incidence and mortality statistics by user-selected pattern categories. The reports would also summarise 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.

As indicated by the bottom feedback loop in Figure 2, the user can utilise 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.

3 Web-based technology

Several issues, controversies and problems can be found with the NCI GIS concept we have just highlighted. A careful analysis of these issues indicates the need for a web-based technological solution.

3.1 Issues, controversies and problems

Much of the relevant spatial and attribute data needed for the geographic studies are captured and stored on an assortment of distributed data sets. In addition, these data are not organised 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, data mining methods have to be used to identify spatial and temporal data patterns manually[12,13]. 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 [14,15]. 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 [16,17].

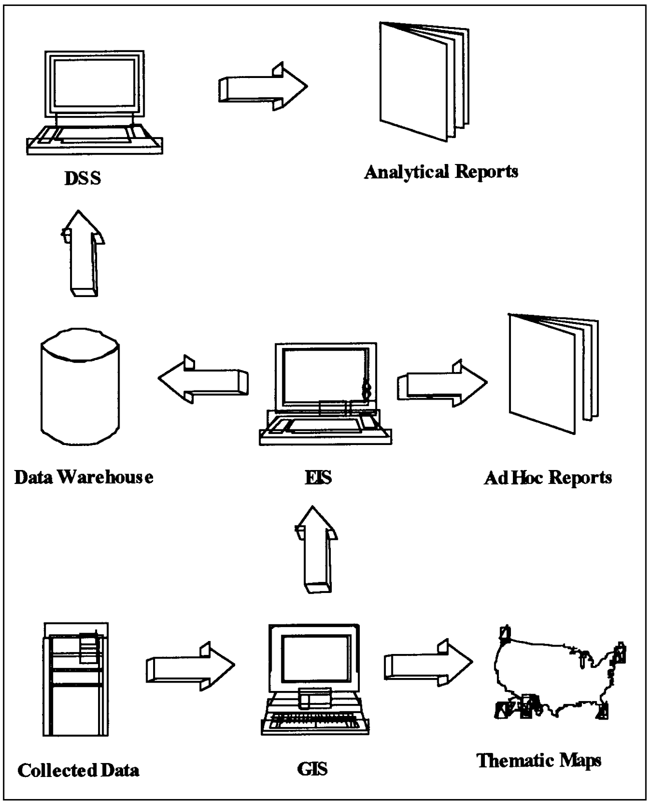
All potential users of the prototype GIS are unlikely to possess the necessary analytical and information technology expertise. Even then, the proposed GIS will not have the functionality to perform the required data management, modelling and presentation tasks. To achieve the desired functionality, additional IT capabilities will have to be developed and linked to the prototype GIS. Modelling assistance, for example, will require a decision support system [18]. Another important requirement of such a system is the ease of accessibility of data to users, which would require a wide area network with sufficient bandwidth.

3.2 Solutions and recommendations

The NCI can have several stand-alone systems to provide the decision analyses and evaluations [19]. 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 [20]. A conceptual 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). Figure 3 shows the relationship between the BCAS components.

As shown, the BCAS concept utilises decision support principles to address the information sharing and integration deficiencies of the NCI's GIS concept. Besides extracting data and creating thematic maps, the GIS provides inputs for the EIS. These inputs consist of the geographic, health outcome, environmental and natural resource, demographic and healthcare variables associated with the desired geographic area. The EIS provides a data warehouse with online analytical processing (OLAP) tools and an intelligent decision support system processor (IHP). A component called *integrator* is responsible for extracting data from the GIS and storing it in the warehouse. A multidimensional database (MDDB) server is used to enable IT personnel to build, maintain and optimise multidimensional data. The OLAP tools take ad hoc queries from the users in an interactive manner and display query results in various formats. The IHP captures the data from the MDDB server, updates the decision support system (DSS) models, performs DSS analyses and evaluations and generates detailed reports of the results automatically without human (manual) intervention.

Figure 3 Breast Cancer Analysis System (BCAS)



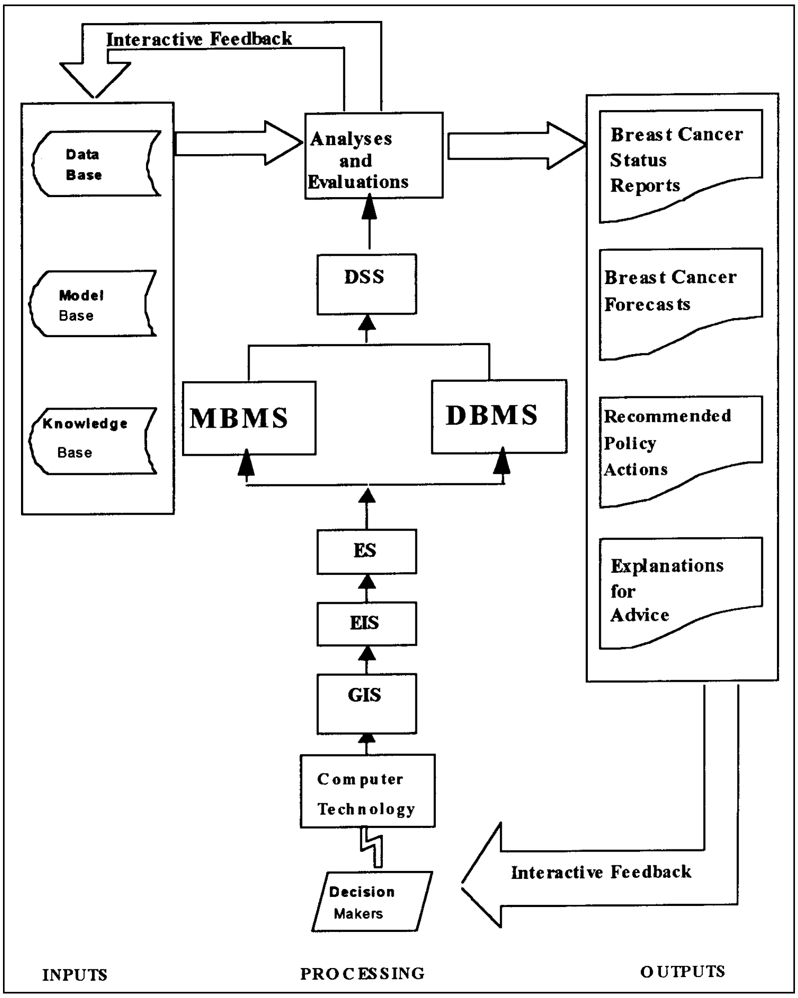
The BCAS will be web-enabled. A website will be established to collect the pertinent data from the various sources. The system will provide an open and scalable OLAP solution using the thin-client architecture. The clients will generate and view reports and graphs of data stored in the MDDB server, which in turn derives the raw data from the data warehouse located at the DB server. Users can run queries and generate reports using any HTML-based browser without having an application session being run on the client. This architecture would reduce the cost of client-server management by centralising the data and applications on the server and make it simple to distribute and upgrade applications and add new OLAP users.

As discussed previously, geographic data, general population characteristics, health-related demographic data and health outcome and care data can all be obtained from relevant authoritative sources. Similarly, environmental data would be acquired from state and federal survey data files on water quality, pollutant, toxic waste, ambient air and source emission, air quality, radiation, powerline and chemical usage and waste generation.

Utilising the electronic commerce (e-commerce) concept, data suppliers will access the BCAS website and select appropriate screen icons [21]. These selections will automatically obtain the data from the supply source and transfer the elements to a data warehouse [22]. The BCAS's EIS will extract the pertinent data, capture the extractions in user-oriented data marts and make the 'marted' data available for ad hoc queries by various stakeholders, including principal investigators and other scientists, public health officials, private health practitioners, individual investigators, public interest groups and the public at large.

Utilising the web-based BCAS, ad hoc queries can be made at the users' sites in an easy-to-use, convenient and interactive manner. Results can be displayed variously (e.g., tabular, graphic and maps). Developers would identify the collected spatial and spatial-temporal variables that the literature suggests predict breast cancer incidence. Data mining and statistical testing would determine the collected variables most relevant to such predictions. The EIS would extract only these pertinent variables. Figure 4 depicts the proposed BCAS architecture.

Figure 4 BCAS architecture



3.2.1 Inputs

The BCAS database captures and stores EIS-extracted geographic, demographic, environmental, health outcome and healthcare data. A multiple equation statistical model can be used to establish independent and joint causation among these data. The model would predict why spatial patterns are occurring for specific demographic, spatial and other profiles. It would also determine if there are deviations from the profile and therefore if there is an epidemic at hand and identify the specific causes of the epidemic.

Using the literature to identify potential methodologies, developers would use these methodologies to establish the precise form of the multiple-equation statistical model. A model base, or organised repository of models and algorithms, would capture and store this model along with the GIS, health and other statistical models required and desired in the NCI Specified GIS-H Functions table. As a package, these models would 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. As well, a knowledgebase is used to capture and store 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.

3.2.2 Processing

BCAS users can interactively perform analyses and evaluations that include organising data into parameters needed for the spatial, temporal and space-time breast cancer analyses, structuring models that represent and simulate breast cancer development patterns in an integrated and complete manner and profiling breast cancer incidence and mortality under specified healthcare, health outcome, demographic and environmental scenarios.

By making desired BCAS screen selections, these users will be able to access automatically the data warehouse, extract pertinent data, form the variables needed for the system's models, use the variables to estimate the models' parameters, simulate expected breast cancer incidences and mortalities, compare actual results with the expected occurrences and obtain a brief explanation for the deviations.

For hardware, a standard PC with sufficient storage devices to accommodate the massive spatial and attribute data to be processed and compatible display devices sufficient to show cartographic, interactive visual, map, tabular, plotted, text and multimedia processing results in hard and soft copy form will be needed. Software would include a windowed operating system and an integrated application software suite that will accommodate data warehousing, spatial and temporal data analysis and decision support. The embedded GIS would serve as a front-end to the processing of the profile data. This system would extract the pertinent data from the data warehouse, link the data to spatial dimensions and transfer the linked data to the embedded EIS.

The EIS would assist in filtering the profile data, form BCAS's database, focus the filtered information and communicate deviations from expected breast cancer development patterns among affected parties [23]. Such processing will involve masking individual identities from publicly released health data, performing disease surveillance, identifying stabilised rates, assessing exposure, profiling genetic activity, analysing interactive spatial data, kriging, scanning statistic drill downs, analysing disease rate in small area variations, clustering space-time groupings and computing standard mortality rates. The EIS will also provide the focused data needed for the DSS analyses and evaluations.

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

3.2.3 Outputs

By controlling specific processing, the users can generate visually attractive status reports that describe the study area's existing breast cancer incidence and mortality summaries, track meaningful trends and discover data patterns. They can also perform claims condition and other e-commerce forecasts, simulate provider policy and payer program results and evaluate recommended claims and other e-commerce actions. The system also would depict, in a graphic manner, the reasoning (explanations and supporting knowledge) that leads to the suggested actions.

3.2.4 Feedback loops

Feedback from the processing would provide additional data, knowledge and enhanced decision models that may be useful for future e-commerce activities and tasks [24]. Output feedback (often in the form of sensitivity analyses) would be used to extend or modify the original analyses and evaluations. All processing (including each feedback loop) would be done in a user-friendly manner, with artificial intelligence (mainly expert system) technology, that meets the decision styles and requirements of participating parties.

4 Simulated system operations

At this time, there is insufficient data available to develop, test and implement the BCAS concept. However, there is sufficient preliminary information to simulate system operations, examine the results from the simulation and measure the potential benefits from these results. The simulation discussed here is based on the use of the SAS System for Information Delivery for design, development and testing.

4.1 Data warehousing

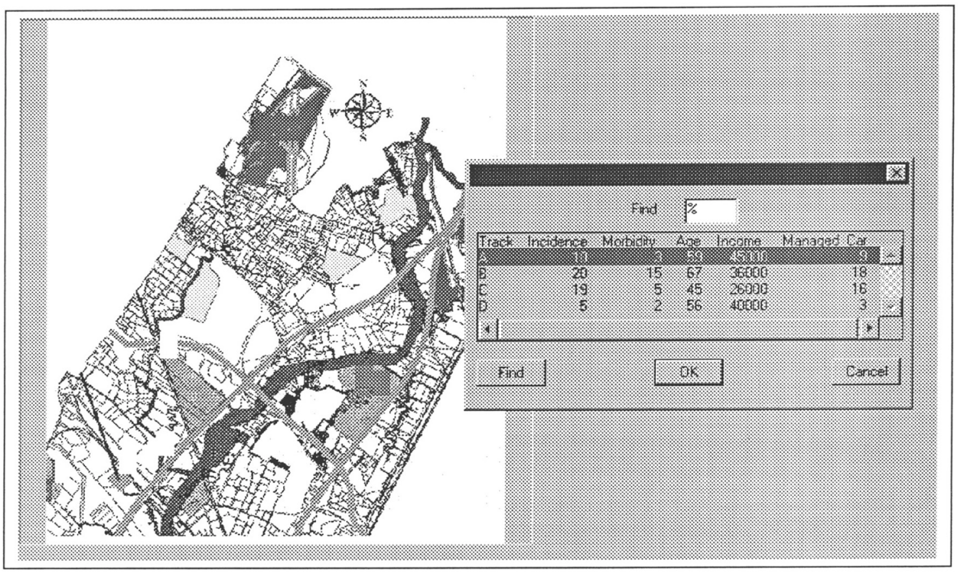
Users would enter the system through a 'Welcome' screen. This screen, like others in the application, would be created with SAS/AF. From here, users would be directed to a MAP SELECTION screen that would enable them to identify the study area map by selecting the pertinent state, county, city and region from screen displayed list boxes.

The selections would call SAS/ACCESS procedures to view the data sources and extract relevant variables for the BCAS analysis. These variables would be converted into a common format using SAS/BASE procedures and warehoused using SAS/BASE and SAS/EIS procedures. Such access and warehousing functions would require manual processing offline in the NCI GIS concept, but the functions would be automated and performed in real time for the user in the BCAS.

4.2 Mapping

After the data warehouse has been created, the BCAS would display a map depicting what is occurring in the study area and what would be expected to occur in the absence of an epidemic. This displayed map would highlight abnormal cancer concentrations as well as demographic and other correlating factors in pop-up tables. Figure 5 gives an illustration of the initial display map.

Figure 5 BCAS map display



Cancer concentrations would be displayed as hyper-regions on the displayed maps. Clicking on these regions, users could drill down to a very concentrated study area to determine if the deviations are focused in one area or general to the entire region. Map displays and drill down capabilities would be similar in the NCI GIS and BCAS.

4.3 Data mining

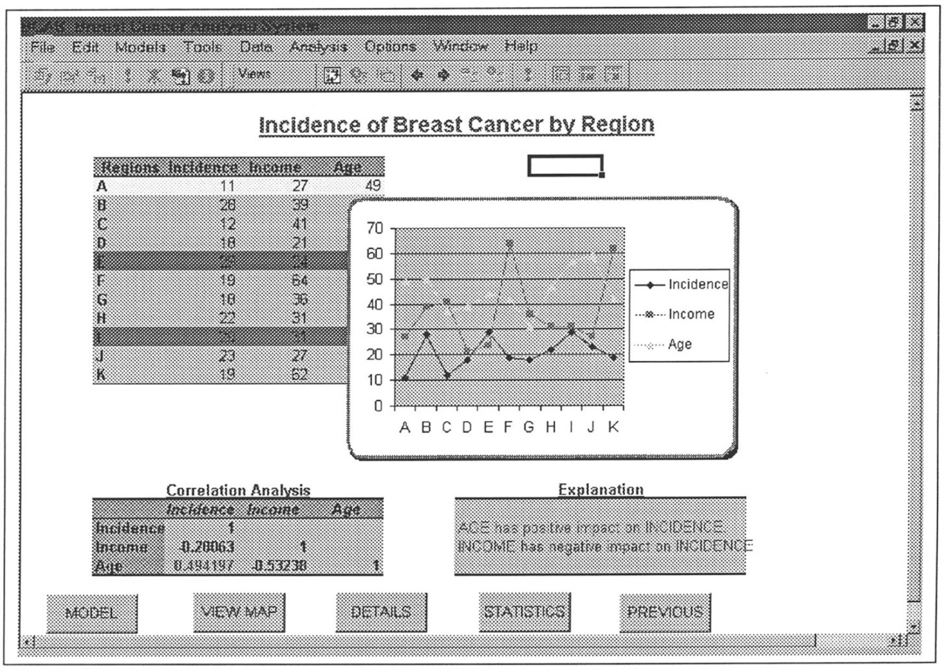
Whereas the NCI GIS concept requires that the cancer correlates and any subsequent analyses, be determined offline and manually, the BCAS allows data mining tools, delivered through SAS/BASE and SAS/STAT procedures, to perform these correlation analyses automatically and in an artificially intelligent manner for the user.

Furthermore, the AI-supported data mining would enable the user to request an explanation for the detected correlation pattern. Such a request would call on SAS/BASE and SAS/STAT procedures that display the supporting statistics graphically and present an accompanying verbal explanation for the results. From this display, users could select a pushbutton that would automatically call the multiple equation statistical model, estimate the model's parameters with the latest data from the data warehouse and display the Policy Analysis screen. This model would utilise SAS/ETS procedures to analyse whether the cancer pattern is caused by a specific factor or a series of interrelated factors specific to the region.

4.4 Explanatory policy analysis

The BCAS Policy Analysis screen would elicit the values for the model's controllable (policy) and uncontrollable (environmental and other) inputs from the user. With this information, the system would call SAS/ETS procedures to generate forecasted study area cancer incidences both as point and confidence interval estimates. In this way, users could perform experiments, in real time and before implementation and sensitivity analyses to find out what would happen if particular conditions or policies changed. All results could be displayed on maps with or without tabular reports, thereby justifying the analyses and evaluations as shown in Figure 6.

Figure 6 BCAS results display



5 Potential benefits and challenges

Use of the BCAS has potential economic benefits. Without the BCAS, the NCI must contract for the access and warehousing functions needed to provide the data necessary for the GIS analyses. Based on figures specified in the 1998 solicitation, such contracting may easily cost 1.3 million dollars in actual expenditures and solicitation expenses. Detecting cancer correlates, forming a multiple equation explanatory model and simulating policies with the developed model will require additional contracting with different specialists. This additional contracting cost will be at least another half a million dollars. These costs will recur as new data becomes available, as the model is revised to accommodate the data changes and as new policies become popular. In fact, it has been suggested that the updating will occur every three years. Over a six-year system life span, we estimate that use of the BCAS can be expected to save up to 3.5 million dollars.

Moreover, use of the BCAS can result in management benefits. Acting as an electronic counsellor, the web-based DSS sequentially guides the user through an effective analysis and evaluation. System operations, which are performed in an intuitive, timely and error-free fashion, liberate the user to focus on the creative aspects of breast cancer diagnosis and treatment. Based on the experience in a previous and similar project, it is anticipated that a typical system session will last about five minutes [25]. This speed is in sharp contrast to the months of effort that would be involved in the contracting process for manual data access and warehousing, explanatory modelling and policy analyses.

More generally, the medical community can be expected to benefit from BCAS usage. Web-based e-commerce is an emerging trend that can benefit the healthcare community and the nation in a variety of ways. Such commerce enables the healthcare organisation to be proactive rather than reactive. For example, through this commerce, healthcare organisations can process expected outcomes and compare them against actual patient outcomes. Such comparisons can help predict:

- patient problems,
- required healthcare interventions,
- time required for implementation of healthcare services,
- accessibility of healthcare services,
- quality of healthcare services and
- cost of healthcare services.

The BCAS concept provides a vehicle to utilise the emerging web-based e-commerce technology for proactive breast cancer diagnosis and treatment. Without it, 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.

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.

Finally, the BCAS enables the researcher and policy analyst to improve the decision making required in the breast cancer analysis process. The BCAS provides quicker analysis of the breast cancer environment, rapid sensitivity analyses of geographic changes and additional flexibility in diagnosing and treating breast cancer. These capabilities will enhance the ability of healthcare providers to manage efficiently and effectively the nation's cancer resources.

5.1 Challenges

Realising the strategic potential of BCAS will present significant challenges to the traditional healthcare organisation. Tasks, events and processes must be redesigned and reengineered to accommodate the concurrent e-commerce. Clinicians and administrators must be convinced that the e-commerce will be personally as well as organisationally beneficial and they must agree to participate in the effort. In the end, these organisational changes will compel substantial IT support.

With the target population of users being so large and diverse, a typical challenge would be to make the system accessible and easy to use for various user profiles. The use of web-based technologies can offer a mechanism to facilitate this task. The wider accessibility of BCAS over the internet can further enhance the usability of the system.

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 minimised while processing efficiency can be dramatically increased. User-inspired creative study-area and public health policy experimentation will be facilitated and nurtured. Knowledge capture will be expedited and new forms of management learning promoted.

Apparently, then, use of BCAS technology may substantially reshape the organisational culture. Faced with significant time pressures and limited staff, healthcare leadership may be reluctant to take on this burden for several reasons. For instance, 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, there may be political pressures on public health officials to preserve established practitioner and vendor relationships than to take up the challenges.

6 Conclusions

The BCAS conceptual architecture is based on a combination of database, modelling, data mining and mapping techniques. It brings together the information and knowledge needed to support breast cancer detection in a comprehensive, integrated and continuous fashion. Its deployment would enable the healthcare community to realise significant economic and political benefits. Future enhancements, suggested by the future research directions to be discussed later, can increase the power of the conceptual BCAS to improve further the nation's ability to manage its scarce cancer resources.

In theory, the BCAS concept points to a comprehensive, integrated and continuous support, thereby yielding more decision value than the non-synthesised and partial support offered by any single autonomous system. Improvements should be observed both in the outcomes from and the process of, strategic claims and other e-commerce decision making [26]. Outcome improvements can include advancements in the level of the users' decision-making maturity and gains in organisational performance [27] whereas process improvements can involve enhancements in the users' ability to perform the phases and steps of decision making.

Regardless of the eventual BCAS legacy, the conceptual system application offers useful lessons for web-based healthcare decision technology systems development and management. The system would effectively deliver to the user, in a virtual manner, embedded statistical, medical and information systems expertise specifically focused on the healthcare problem. Any single human technical specialist typically will not be proficient with, or even aware of, all pertinent tools, or possess sufficient domain knowledge to understand fully the medical situation, propose trials, or interpret outcomes. While practitioners will have the domain knowledge, they will not usually have the technical expertise to develop and implement relevant technology effectively.

Breast cancer diagnosis and treatment is inherently a semi-structured (or even ill structured) problem. When initially confronted with such situations, researchers and analysts have a partial understanding of the problem elements and relationships. Typically, their understanding evolves as they acquire more information, knowledge and wisdom about the problem. Decision technology systems are designed to support such decision situations. Relying on the information centre, or other traditional IT organisation, to design and develop a web-based healthcare decision technology system is likely to be ineffective as these types of organisations are typically staffed by personnel with general skills, limited technological expertise and restricted problem-specific knowledge. The BCAS concept suggests that decision technology system design, development and implementation should be a team effort. In addition, the team should involve the affected practitioners, IT personnel and technological specialists proficient with the tools needed to address the healthcare problem.

A hybrid project-technology organisation may work well for web-based decision technology system design, development and implementation in a healthcare environment. The organisation could be 'virtual' rather than 'physical.' A project team would be established and administered by the practising healthcare professional. Team technology specialists would be drawn from within and outside the organisation to match the expertise needed for the specific project. Telecommuting and distributed collaborative work would be allowed and possibly encouraged.

6.1 Future research opportunities

The BCAS concept, nonetheless, introduces a number of future research opportunities. To ensure that the system accurately replicates the inputs, the final version of the BCAS should be tested against web-collected data from existing institutions' study areas. In such a testing, warehoused data should be compared against actual values. Statistical tests can be conducted on the estimated models. Evaluations of user satisfaction with BCAS in terms of:

- the speed, relevance and quality of ad hoc query results,
- the system interface,
- model appropriateness and
- the quality of the system explanation

are also key to its success and acceptance. Simulations can also be statistically tested for accuracy and confidence intervals established for the results. Tests can also be conducted on the system's ability to improve the decision-making maturity of the user.

Enhancements can be made to the BCAS architecture. Machine learning techniques can be developed to improve the intelligent modelling, database management and user interface operations of the system. Communication links can be created to disseminate system results to affected parties more effectively.

Finally, 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 AI techniques for model calibration and optimisation. In a domain such as breast cancer diagnosis and treatment, there will often be a need to optimise models, especially as newer models emerge through BCAS processing. These emerging models should undergo a rigorous process of optimisation and calibration. Empirical evidence suggests that genetic algorithms are superior to traditional search and optimisation methods (such as hill climbing) and even neural networks, for such tasks [28, 29]. Thus, the road ahead for deploying web-based technologies is paved with challenges and exciting new opportunities both for research and practice.

References

1 Laden, F., Spiegelman, D. and Neas, L.M. (1997) 'Geographic variation in breast cancer incidence rates in a cohort of US women', *Journal of the National Cancer Institute,* Vol. 89, pp.1373–1378.

2 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,* Vol. 89, pp.960–965.

3 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,* Vol. 87, pp.1846–1853.

4 'Regional variation in breast cancer rates in the US – NIH'. (1998) National Cancer Institute's Electronic RFA Number CA–98–017. Bethesda: National Cancer Institute.

5 *Workshop on Hormones, Hormone Metabolism, Environment and Breast Cancer,* New Orleans, Louisiana, 28–29 September, (1995) *Monographs in Environmental Health Perspectives* supplement 1997, Vol. 105, No. 3, pp.557–688.

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

7 Grimson, R.C. and Oden, N. (1996) 'Disease clusters in structured environments', *Statistics in Medicine,* Vol. 15, pp.851–871.

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

9 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.

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

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

12 Abraham, T. and Roddick, J.F. (1998) 'Opportunities for knowledge discovery in spatio-temporal information systems', *Australian Journal of Information Systems*, Vol. 5, No. 2, pp.3–12.

13 Davidson, G.S., Hendickson, B., Johnson, D.K., Meyers, C.E. and Wylie, B.N. (1999) 'Knowledge mining with VxInsight: discovery through interaction', *Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*, Vol. 11, No. 3, pp.259–285.

14 Borok, L.S. (1997) 'Data mining: sophisticated forms of managed care modelling through artificial intelligence', *Journal of Health Care Finance*, Vol. 23, No. 3, p.20.

15 Burn-Thornton, K.E. and Edenbrandt, L. (1998) 'Myrocardial infarction-pinpointing the key indicators in the 12-lead ECG using data mining', *Computers and Biomedical Research*, Vol. 31, No. 4, pp.293–303.

16 Hornung, R.W., Deddens, J.A. and Roscoe, R.J. (1998) 'Modifiers of lung cancer risk in uranium miners from the Colorado Plateau', *Health Physics*, Vol. 74, No. 1, pp.12–21.

17 Makino, K., Suda, T., Ono, H. and Ibaraki, T. (1999) 'Data analysis by positive decision trees', *IEICE Transactions on Information and Systems*, E82–D(1), pp.76–88.

18 Tan, J. with Sheps, S. (1998) *Health Decision Support Systems*, Gaithersburg, Maryland: Aspen.

19 Tan, J. (2001) *Health Management Information Systems: Methods and Practical Applications,* 2nd Edition, Gaithersburg, Maryland: Aspen.

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

21 Gerull, D.B. and Wientzen, R. (1997) 'Electronic commerce: the future of image delivery' *International Journal of Geographical Information Systems*, Vol. 7, No. 7, pp.38–51.

22 Tsvetovatyy, N., Gini, M. and Wieckowski, Z. (1997) 'Magma: an agent-based virtual market for electronic commerce', *Applied Artificial Intelligence,* Vol. 11, No. 6, pp.501–509.

23 Keegan, A.J. and Baldwin, B. (1992) 'EIS: a better way to view hospital trends', *Healthcare Financial Management,* Vol. 46, No. 11, pp.58–64.

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

25 Forgionne, G.A. (1997) 'HADTS: a decision technology system to support army housing management', *European Journal of Operational Research*, Vol. 97, pp.1120–1135.

26 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,* Vol. 7, No. 1, pp.17–25.

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

28 Kao, C., Chen, L-H., Wang, T-Y., Kuo, S. and Horng, S-D. (1995) 'Productivity improvement: efficiency approach vs effectiveness approach', *Omega, International Journal of Management Science*, Vol. 23, No. 2, pp.197–204.

29 Kim, W. (1998) 'Data mining: promises, reality and future', *Journal of Object-Oriented Programming*, Vol. 11, No. 4, pp.61–62+.