GRANT NUMBER DAMD17-94-J-4383

TITLE: Developing and Implementing the AJCC Prognostic System for Breast Cancer

PRINCIPAL INVESTIGATOR: Philip H. Goodman, M.D.

CONTRACTING ORGANIZATION: University of Nevada
Reno, Nevada 89520

REPORT DATE: August 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
**Title and Subtitle**
Developing and Implementing the AJCC Prognostic System for Breast Cancer

**Author(s)**
Philip H. Goodman, M.D.

**Performing Organization Name(s) and Address(es)**
University of Nevada
Reno, Nevada 89520

**Sponsoring/Monitoring Agency Name(s) and Address(es)**
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

**Abstract**
The goal of this project is the creation of a computer-based prognostic system for breast cancer that: (1) is significantly more accurate than the TNM staging system, (2) that predicts survival over time based on therapy, and (3) presents its predictions in a manner that physicians can understand and use.
Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

Where copyrighted material is quoted, permission has been obtained to use such material.

Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

[Signature]

PI - Signature    Date
<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3-4</td>
</tr>
<tr>
<td>Detailed Report by Task</td>
<td>5-8</td>
</tr>
<tr>
<td>Conclusions</td>
<td>9</td>
</tr>
<tr>
<td>References</td>
<td>10-11</td>
</tr>
<tr>
<td>Attachment # 1 - Accuracy methods</td>
<td>12</td>
</tr>
<tr>
<td>Attachment # 2 - Graphical output</td>
<td>13</td>
</tr>
</tbody>
</table>
Introduction

The goal of this project is the creation of a computer-based prognostic system for breast cancer that is significantly more accurate than the TNM staging system, predicts survival over time based on therapy, and presents its predictions in a manner that physicians can understand. This project can be viewed as consisting of three components: (1) Data analysis and prognostic factor evaluation, (2) developing the prognostic model, and (3) implementing a clinically useful system, i.e., breast cancer prognostic factors, the artificial neural network statistical model, and the clinician user interface.

The first year of research was characterized by work on the artificial neural network statistical model and related statistical models, specifically tasks 2.1.2 (artificial neural network generating survival curves), 2.1.3 (determining the accuracy of the survival curves), 1.03, 2.1.4 (comparing the accuracy of the artificial neural network to other statistical models), 2.2 (implementing an effective solution for missing data in training and performance), and 2.3 (dealing with censored data).

In addition, during the first year we started work related to data analysis and prognostic factors including 1.02 and 1.08.1 (recurrence as an endpoint), 1.04.2 (creating a taxonomy of prognostic factors in breast cancer), 1.04.3 (writing a book on prognostic factors in breast cancer, in preparation), 1.06.3 (determining minimum data set size), 1.11 (examining physician breast cancer survival estimates). We also began work on 3.1 (the code) and 3.2 (the physician interface).

Also during the first year we added three tasks, (1) a comparison of the two main American cancer data bases, namely, the Surveillance, Epidemiology, and End Results and the National Cancer Data Base data bases. (2) An examination of the issue of what to do when confronted with cases not lost completely at random and competing risks. (3) Computerization of the TNM staging system.

The second year of research was characterized by the continuation of work begun in the first year and by data analysis and prognostic factor development, specifically tasks 1.01 and 2.1.1 (extending the survival
endpoint from five to ten years), 1.02 and 1.03 (see first year), 1.05 (the identification of high risk node negative women), 1.06 (clinical trials), and 1.07 (therapy). Work continued on 2.1.2 (artificial neural network generating survival curves), 2.1.3 (creating a new method for assessing prediction accuracy), 1.03 and 2.1.4 (model comparisons). In addition, during the second year work began on 1.04.1 (new molecular-genetic prognostic factors). Work was completed on the computerization of the TNM staging system and the comparison of the two national cancer data bases.
Detailed Report: by Task

Task 1. Data analysis and prognostic factor evaluation.

1.01) Extend analysis of binary survival endpoint to 10 year survival.

We have analyzed SEER 10 year survival data. We found that the predictors collected at disease discovery are less accurate in predicting 10 year survival than 5 year survival.

SEER 1977 - 1982 Breast Cancer Data: 10 year Survival Prediction Accuracy, TNM Variables

<table>
<thead>
<tr>
<th>PREDICTION MODEL</th>
<th>ACCURACY*</th>
<th>SPECIFICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNM Stages</td>
<td>.692</td>
<td>Ø,IIA,IIIB,IIIA,IIIB,IV</td>
</tr>
<tr>
<td>Artificial Neural Network</td>
<td>.730§</td>
<td>3-5-1</td>
</tr>
</tbody>
</table>

* The receiver operating characteristic area.
§ p < .01

Five year survival accuracy for the TNM staging system was .720 and for the artificial neural network, .784.

1.02) Extend the analysis to recurrence as an endpoint.

Completed, in last year's report.

1.03) Comparison of prognostic models.

Completed, in last year's report.

1.04.1) New prognostic factors

We have obtained from Duke University a data set that contains, in addition to the TNM variables age, estrogen and progesterone receptor status, histology, p53, and erbB-2. The accuracy results are shown below. These new prognostic factors produce major increases in prognostic accuracy. This is a very encouraging result.

Duke Breast Cancer Data: 5 year Survival Prediction Accuracy

<table>
<thead>
<tr>
<th>PREDICTION MODEL</th>
<th>ACCURACY*</th>
<th>SPECIFICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTNM Stages</td>
<td>.567</td>
<td>Ø,IIA,IIIB</td>
</tr>
<tr>
<td>Stepwise Logistic Regression</td>
<td>.865</td>
<td>no interactions</td>
</tr>
<tr>
<td>Backpropagation Neural Network</td>
<td>.926</td>
<td>9-5-1</td>
</tr>
</tbody>
</table>

* The area under the curve of the receiver operating characteristic.

These results are based on slightly more than 100 cases, so the results must be interpreted cautiously. We will be obtaining several hundred additional cases from Duke in the near future. We will use the additional cases to confirm these results.

1.04.2) Create a taxonomy of prognostic factors in breast cancer.

Completed, in last year's report.


1.05) High risk node negative women

Based on our initial results with the new prognostic factors contained in the Duke data set (1.04.1) we believe that some of these factors will be very useful in identifying high risk node negative women. This work is ongoing.

1.06.1, 1.06.2) Clinical trials

The increases in prognostic accuracy we described last year and this year suggest that we are finding more homogeneous patient populations. The Duke data set contains treatment information and we are currently using it to discover which new prognostic factors predict response to specific therapies. This work is ongoing.

1.06.3) Determining minimum data set size.

Completed, in last year's report.

1.08.1) Recurrence analysis.

Completed, in last year's report.

1.11) Patient information and physician credibility.

Completed, in last year's report.

Task 2. Developing the prognostic model.

2.1.1) Generate survival curves for 10 year data.

Refer to 1.01 for details.

2.1.2) Generating survival curves.

Completed, in last year's report.

2.1.3) Determining the accuracy of the survival curves.

We are currently working on a new measure of prediction accuracy that we call "A". It includes the area under the receiver operating characteristic as a special case. This work is ongoing. We have implemented several accuracy methods and derived the asymptotic variance for each in order to assess the adequacy of each method. (See Attachment).

2.1.4) Comparison of artificial neural networks with Cox proportional hazards model.

We began our comparison of the Cox by examining whether breast cancer violates the proportional hazards assumption of the model. Proportional hazards methods include the Cox (1972), and less commonly the Weibull or exponential distributions (Evans, 1993). Proportional hazards methods assume that the hazard
of each patient is proportional to the hazards of all the other patients and that a individual patient's hazard is related to that patient's relative risk. The Cox model does not create survival curves. For Cox-related survival curves a baseline hazard must be introduced (Breslow-Cox estimates; Breslow, 1974). Some researchers incorrectly believe that only regression methods that assume proportional hazards can deal with censoring, but a multiinterval regression model that drops patients during the interval in which they are censored is capable of dealing with censoring. It is always vital to test the proportional hazards assumption when using a regression method that relies on it. There are several methods for assessing proportional hazards violation, including Schoenfeld's partial residuals (Schoenfeld, 1982) and the log hazard ratio as a function of time (Gore, 1986). We have created a method somewhat similar to Gore. We construct a Cox model, divide the time into sub-intervals, and assess the accuracy of the model for each sub-interval. If proportional hazards holds, accuracy should be constant across sub-intervals. Results for breast cancer are shown below.

**TABLE.** Area under the receiver operating characteristic (Az) for two Cox models; breast cancer (five one-year intervals) N = 1,222 and melanoma (three six-month intervals) N = 60.

<table>
<thead>
<tr>
<th>Model/Interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast (SE)</td>
<td>.734</td>
<td>.735</td>
<td>.758</td>
<td>.773</td>
<td>.693</td>
</tr>
</tbody>
</table>

[Graph showing area under the curve for Cox model evaluated at five time intervals]

Because the Az values are not constant across the sub-intervals, proportional hazards does not hold for breast cancer.

Artificial neural networks are a general regression method that do not assume proportional hazards and can capture nonlinearity and complex interactions (Burke, 1994, 1995a). Multiinterval artificial neural networks can handle censoring in the same way that multiinterval logistic regression models handle censoring. It seems clear that proportional hazards is probably not appropriate for breast cancer or lung cancer (results not reported).

2.2) Missing data (see discussion in last year's report also)

We have completed work on an efficient missing data mechanism. It is available for use by interested researchers.

*Rosen DB, Burke HB. Applying a gaussian-bernoulli mixture model network to binary and continuous missing data in medicine. Submitted for publication.*

2.3) Censored cases

Discussed in detail in last year's report (2.1.2).

Task 3. Implementation of a clinically useful prognostic system


All our work is written in either C, C++, or XLISP-STAT.

3.2) Physician interface.

It is very important that physicians find the new prognostic system easy to use and useful. To this end we have implemented the prognostic system on a DOS platform with a Windows interface. We are presenting the system to clinicians and receiving feedback regarding what is important to them in terms of information and the graphical display of the information. (see 2.1.2) See Attachment for screen output.

Tasks added to the project

(1) The computerization of the TNM staging system for breast cancer has been completed and has been integrated into the prognostic system.

(2) Comparison of the NCDB and the SEER data sets in terms of breast cancer.

We have completed our comparison of the two national breast cancer data bases, the National Cancer Data Base (NCDB) and its associated Patient Care Evaluation (PCE), and the Surveillance, Epidemiology, and End Results (SEER) data sets. We evaluated them in terms of: (i) representativeness, is the data set an unbiased representation of the breast cancer population. (ii) Incidence/prevalence, how good is the data set in capturing the incidence and prevalence of breast cancer. (iii) Prognosis/outcome, how good is the data set in providing information that is useful for predicting outcome. An overview of the results are shown below.

<table>
<thead>
<tr>
<th></th>
<th>SEER</th>
<th>NCDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Incidence/prevalence</td>
<td>good</td>
<td>adequate</td>
</tr>
<tr>
<td>Prognosis/outcome</td>
<td>not acceptable</td>
<td>adequate</td>
</tr>
</tbody>
</table>

Both are representative of the breast cancer population. SEER does a better job at incidence and prevalence because it ascertains all cases in a catchment area, regardless of whether the hospital belongs to the American College of Surgeons accreditation program, and it contains relatively little missing data. NCDB contains a great deal of missing data. SEER can not be used for prognosis because it does not provide therapy data due to the unreliability of their data. NCDB suffers from a lack of follow-up, resulting in high censoring rates.

Burke HB, Hoang A, Visintainer, P. Comparison of the two national cancer data sets: SEER and NCDB. In preparation.
Conclusions

The second year of research was characterized by the continuation of work on the artificial neural network statistical model and related statistical models and by data analysis and prognostic factor development. In summary, the research is going well we are ahead of our time schedule. We feel that we will be able to successfully meet our goal of providing a computer-based prognostic system that is more accurate than the TNM staging system and that is easy to use and understand within the four year time frame of this grant. In addition, we have created several new systems that we believe will advance the domain of cancer prognosis, e.g., artificial neural network survival-over-multi-interval-time models, an effective missing data method for training and performance, and a new approach to the assessment of prediction accuracy.
References

Publications and Presentations Related to this Grant During The Second Funding Year

Books


Peer Reviewed Book Chapters


Peer Reviewed Journals and Proceedings

Rosen DB, Burke HB. Applying a gaussian-bernoulli mixture model network to binary and continuous missing data in medicine. Submitted for publication.
Burke HB. The TNM staging system. In preparation.
Burke HB, Henson DE, Bostwick DG, Hoang A. Backward and forward relationships between prior events and disease. In preparation.
Hoang A, Burke HB. Methods for dealing with cases nonrandomly lost-to-follow-up. In preparation.
Burke HB, Hoang A. Assessment of multi-interval event models. In preparation.
Burke HB, Hoang A, Visintainer P. Comparison of the two national cancer data sets: SEER and NCDB. In preparation.

Invited papers


Presented papers

Rosen DB, Burke HB, Goodman PH. Improving prediction accuracy using a calibration postprocessor. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.
Burke HB. Measuring classification/prediction accuracy. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.
Burke HB. Defining the computer-aided diagnostic device domain. World Congress on Neural Networks, San Diego CA, September 15 - 20, 1996.

Conference positions

1996 Program Committee and Co-chair, Biomedical Applications Section, World Congress on Neural Networks and 1996 International Neural Network Society Annual Meeting, San Diego CA, September 15 - 20, 1996.
<table>
<thead>
<tr>
<th>MEASURE OF ASSOCIATION</th>
<th>PROBABILITY STATEMENT</th>
<th>INTERPRETATION</th>
<th>ESTIMATION (DISCRETE)</th>
<th>ASYMPTOTIC VARIANCE OF THE ESTIMATE (DISCRETE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodman and Kruskall's Gamma</td>
<td>$P((x_i &gt; x_j, y_i &gt; y_j) \text{ or } (x_i &lt; x_j, y_i &lt; y_j)) - \frac{P((x_i &gt; x_j, y_i &lt; y_j) \text{ or } (x_i &lt; x_j, y_i &gt; y_j))}{P(y_1 \neq y_2)}$</td>
<td>The excess of the degree of the same movement in X and Y over the degree of the opposite movement in X and Y, after removing ties in X and Y</td>
<td>$\frac{P-Q}{P+Q}$</td>
<td>$16 \sum_{i,j} N_{ij}(Q_{ij} - P_{ij})^2$</td>
</tr>
<tr>
<td>Somer's D_{xy}</td>
<td>$P((x_i &gt; x_j, y_i &gt; y_j) \text{ or } (x_i &lt; x_j, y_i &lt; y_j)) - \frac{P((x_i &gt; x_j, y_i &lt; y_j) \text{ or } (x_i &lt; x_j, y_i &gt; y_j))}{P(y_1 \neq y_2)}$</td>
<td>The excess of the degree of the same movement in X and Y over the degree of the opposite movement in X and Y, after removing ties in X and Y</td>
<td>$P-Q \over N^2 - \sum N^2_i$</td>
<td>$4 \sum_{i,j} N_{ij}[w_{ij} - (P-Q)(N-N\cdot)]^2$</td>
</tr>
<tr>
<td>Kendall's Tau-b</td>
<td>$P((x_i &gt; x_j, y_i &gt; y_j) \text{ or } (x_i &lt; x_j, y_i &lt; y_j)) - \sqrt{P(x_i \neq x_j)P(y_1 \neq y_2)}$</td>
<td>Tau-b is the excess of the degree of the same movement in X and Y over the degree of opposite movement in X and Y. Tau-b is Tau-a relative to the geometric average of the chance of no ties in X and Y.</td>
<td>$P-Q \over \sqrt{(N^2 - \sum N^2_i)(N^2 - \sum N^2_j)}$</td>
<td>$1 \over w^2 \sum_{i,j} N_{ij}(2w_{ij} + (\text{tau}<em>b)v</em>{ij})^2 - \overline{N^2(\text{tau}_b)^2(w_i + w_j)^2}$</td>
</tr>
</tbody>
</table>

Area under the receiver operating characteristic (A_{xy})

$P(x_i > x_j | y_i > y_j)$, continuous x

Given the direction of movement of Y, the degree to which X has the same direction of movement.

$S(P+Z) \over L_z$
Patient Name: 
Patient ID: 
Institution: 
Date: 
Physician: 
Cancer Type: Breast 
Prediction Type: 5 Year Survival Curve 

Tumor: 
Tumor Size: 2 
Lymph Nodes Positive: 3 
Lymph Nodes Examined: 10 
Distant Metastasis: n 
Estrogen Receptor: n 
Progesterone Receptor: n 
Menopausal Status: y 
Grade: 1 
Age: 55 
Lymph Nodes pTNM: 1 

Factors: 

tsize = 
Tumor: 
T Size: 2.4 
LN Pos: 4 
LN Exam: 10 
Mets: n 
ER: n 
PR: n 
Menopausal: n 
Grade: 2 
Age: 40 
LN pTNM: 2 

<table>
<thead>
<tr>
<th>Year</th>
<th>ANN Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>0.983</td>
</tr>
<tr>
<td>2</td>
<td>0.898</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>0.775</td>
</tr>
<tr>
<td>5</td>
<td>0.721</td>
</tr>
</tbody>
</table>

Probability: 1.0
Probability: 0.8
Probability: 0.6
Probability: 0.4
Probability: 0.2

TNM Stage: IIIA 
Prediction: 0.595