Breast Cancer Research and Treatment **22**: 263-272, 1992.

A comparison of all-subset Cox and accelerated failure time models with Cox step-wise regression for node-positive breast cancer

Judy-Anne W. Chapman[1], Maureen E. Trudeau[1], Kathleen I. Pritchard[2], Carol A. Sawka[2], Betty G. Mobbs[1], Wedad M. Hanna[1], Harriette Kahn[1], David R. McCready[1], Lavina A. Lickley[1]
[1]*Henrietta Banting Breast Centre, Women's College Hospital, University of Toronto, and* [2]*Toronto-Bayview Regional Cancer Centre and Sunnybrook Health Science Centre, University of Toronto; Toronto, Ontario, Canada*

Key words: breast cancer, prognostic factors, Cox regression, all-subset regression, accelerated failure time models

Summary

Clinical studies usually employ Cox step-wise regression for multivariate investigations of prognostic factors. However, commercial packages now allow the consideration of accelerated failure time models (exponential, Weibull, log logistic, and log normal), if the underlying Cox assumption of proportional hazards is inappropriate. All-subset regressions are feasible for all these models.

We studied a group of 378 node positive primary breast cancer patients accrued at the Henrietta Banting Breast Centre of Women's College Hospital, University of Toronto, between January 1, 1977, and December 31, 1986. 85% of these patients had complete prognostic factor data for multivariate analysis, and 96% of the patients were followed to 1990. There was evidence of marked departures from the proportional hazards assumption with two prognostic factors, number of positive nodes and adjuvant systemic therapy. The data strongly supported the log normal model. The all-subset regressions indicated that three models were similarly good. The variables 1) number of positive nodes, 2) tumour size, and 3) adjuvant systemic therapy were included in all three models along with one of three biochemical receptor variables 1) ER, 2) combined receptor (ER- PgR-; ER+ PgR-; ER- PgR+; ER+ PgR+; or 3) PgR.

Better multivariate modeling was achieved by using quantitative prognostic factors, a check for appropriate underlying model-type, and all-subset variable selection. All-subset regressions should be considered for routine use with the many new prognostic factors currently under evaluation; it is very possible that there may not be a single model that is substantially better than others with the same number of variables.

Introduction

Clinical studies of disease-free period or survival usually report the results of multivariate analyses from Cox step-wise regressions. A single model usually summarizes the results of a whole study.

Address for correspondence and offprints: Dr. J.W. Chapman, Henrietta Banting Breast Centre, Women's College Hospital, 60 Grosvenor Street (5th floor), Toronto, Ontario, Canada M5S 1B6; [Tel 416-323-7732, Fax 415-921-4275].

However, that one model may not be substantially better than several others with the same number of variables, and the step-wise procedure may miss other important combinations of variables. Likewise, there may be a variable combination, with the same number of variables, that is actually substantially better than the best step-wise model. For such reasons, step-wise selection of factors is being replaced by all-subset regressions in other areas of statistics; the latest version of SAS and BMDP contain all-subset regression options. SAS versions 6.04 and later have an all-subset Cox regression procedure, and it is feasible with a moderate number of covariates to manipulate the step-wise Cox procedure in BMDP to consider all-subset Cox regressions. The ISMOD package [1,2] permits all-subset regression with a larger number of covariates.

The merits of non-parametric Cox regressions versus parametric alternatives have been reviewed and discussed extensively [3,4], so this paper will just summarize the considerations involved with model choice. The Cox model has been routinely used for clinical investigations. It does not require a parametric specification of the hazard function (for example, the risk of recurrence or death from breast cancer), and the underlying assumption of proportional hazards is often appropriate; stratification may alleviate problems associated with departures from this assumption. As well, it is fairly robust for moderate departures from proportional hazards.

An alternative to Cox regression involves parametric modelling of the hazard function, and consideration of accelerated failure time models (exponential, Weibull, log logistic, and log normal) where the covariates act additively on log (disease-free) survival time or multiplicatively on (disease-free) survival time. Clinical experience with node-positive breast cancer suggests that the risk of recurrence actually does change with time after primary treatment [5].

Until lately, parameter estimation and parametric model differentiation have been difficult for routine use of accelerated failure time models. However, the recognition that the exponential, Weibull, log logistic, and log normal are all special cases of the generalized gamma, along with a re-parameterization of the generalized gamma model, has allowed better parameter estimation and differentiation between types of accelerated failure time models. These options are now more tractable. The 1990 version of BMDP has a step-wise regression program for accelerated failure time models with censored data that may be manipulated to produce all-subset selections for a moderate number of variables; ISMOD is available for a larger number of covariates. Residuals are generated by these packages so that modelling assumptions may be checked; this is especially important for the non-robust exponential model. SAS will fit these model types, but the variable selection strategies would have to be supplied by the user. SAS does not have a step-wise variable selection option for accelerated failure time models.

Inference from a study should be with an appropriate model. It should also be with a single model only if it is substantially better than others with the same number of variables. If there is no substantially best single model, inference should be based on only those variables in common to similarly good models. Inconsistent prognostic factor results reported in the literature could easily be due to insufficient statistical work-up. With the advent of many new potentially interrelated prognostic factors, it becomes increasingly important that the statistical investigations be thorough.

We compared the results of all subset Cox regressions with step-wise Cox regression, for traditional prognostic factors in a group of patients with primary, invasive, node-positive breast cancer. The appropriateness of the underlying assumption of proportional hazards was examined, and alternative parametric models were employed in all-subset and step-wise regressions on the data.

Methods

Patients

The Henrietta Banting Breast Centre (HBBC) at Women's College Hospital (WCH), University of Toronto, was established in 1977. All patients who presented to the HBBC for consultation from January 1, 1977, had a standard form completed about medical and treatment histories and family history of cancer. Blocks and slides have been archived for all WCH breast cancer patients. Tumour size was determined using pathologic, radiological, and clinical information. Receptor assays were carried out using the dextran-coated charcoal methods [6]. Patients were followed prospectively, with recurrence data being added regularly in a consistent fashion. Relapses have been defined according to the criteria described by Meakin and Hayward [7]. We studied all node-positive patients accrued between January 1, 1977, and December 31, 1986, except those with known metastases, stage 3B, or inflammatory cancer. In addition, patients with previous cancers (excluding skin or in situ cervix) were excluded from analysis. 378 patients will be considered in this study; the reasons for eliminating patients are outlined in Table 1. The study patients were followed to 1990, and complete follow-up was available in 96% of these patients.

Table 1. Henrietta Banting Breast Centre node-positive patients

Patient total		448
Patients eliminated		
Stage 4	27	
Previous cancer (except skin, in situ cervix)	18	
Stage 3B	16	
Inflammatory breast cancer	8	
Recurrence of undetermined source (pt developed a 2d primary, gastric cancer)	1	
Total		70
Patients studied		378

Prognostic factors

The primary outcome variable that was used to assess prognosis was disease-free interval (DFI). The factors considered were 1) age (≤50, >50 years; inferred menopausal status); 2) tumour size (T1, T2, T3; size in cm); 3) the number of positive nodes (1-3, 4-6, 7-10, ≥11; # positive nodes); 4) biochemical estrogen receptor, and 5) progesterone receptor (ER and PgR) values (<10, ≥10; quantitative fmol/mg protein); 6) combined biochemical receptors (ER–/PgR–, ER+/PgR–, ER–/PgR+, ER+/PgR+); 7) weight (kg); and 8) adjuvant systemic therapy (according to menopausal category, see Table 2). These data were subjected to extensive multidisciplinary quality control measures (oncologically by MET, KIP, CAS; biochemically by BGM; and statistically by JWC). There were complete data for 323 patients (85%), and these patients are therefore included in the multivariate analyses.

Statistical method

All the analyses were on an IBM-compatible 286 microcomputer with a math co-processor using

Table 2. Adjuvant systemic therapy

A.	AGE ≤ 50	AGE > 50	CODE
	Chemotherapy	Hormones	2
	Hormones	Chemotherapy	1
	Nothing	Nothing	0
B.	AGE ≤ 50	AGE > 50	CODE
	Chemotherapy + ovarian ablation ± hormones	Hormones	4
	Chemotherapy + hormones	Chemotherapy + ovarian ablation + hormones	3
	Chemotherapy	Chemotherapy + hormones	2
	Hormones	Chemotherapy	1
	Nothing	Nothing	0

BMDP PC90. Preliminary analyses of the prognostic factors included descriptive statistics, Kaplan-Meier plots, and Wilcoxon (Peto-Prentice) statistics. Maximum likelihood estimation and likelihood ratio tests (–2logR) were used in all the regressions.

The following Cox regressions were performed:

1. Step-wise forward with:
   a) grouped variables: age (≤50, >50), T (1, 2, 3), # positive nodes (1-3, 4-6, 7-10, ≥11), ER (<10, ≥10 fmol/mg protein), PgR (<10, ≥10 fmol/mg protein), combined receptors (ER–/PgR–, ER+/PgR–, ER–/PgR+, and ER+/PgR+), adjuvant treatment (3 and 5 categories, in separate runs), weight (kg).
   b) grouped variables as in (a), but quantitative ER and PgR (fmol/mg protein).
   c) age (≤50, >50), tumour size (in cm), # positive nodes (# positive), quantitative ER and PgR (fmol/mg), adjuvant treatment (3 and 5 categories, in separate runs), weight (kg).
2. Models with all the variables, for each of 1 a-c (full models).
3. All variable subsets with the number of variables determined by the step-wise forward procedure and an inability to make a significant improvement to the model by adding a variable.

The Cox assumption of proportional hazards was checked with plots of the log (cumulative hazard) for stratifications of the covariates in the best models. Finally, Cox-Snell residuals were plotted versus each covariate to check the adequacy of the model.

The modelling process of step-wise forward, full model, and all variable subset selection for the appropriate number of variables was then repeated for the accelerated failure time models (exponential, Weibull, log logistic, and log normal). Cox-Snell and standardized residuals were plotted versus each covariate to check model adequacy and the relationships of residuals and covariates.

The plausibility and support for the parametric models may be inferred with likelihood ratio methods. Maximized relative likelihood [4] can

Table 3. Univariate analysis of disease-free interval

Prognostic factor		N	Wilcoxon statistic P-value
Age			
≤ 50		150	
> 50		228	0.67
Tumor size			
T1		125	
T2		177	
T3		52	< 0.001
Number positive nodes			
1-3		248	
4-6		54	
7-10		40	
≥ 11		36	< 0.001
ER			
< 10 fmol/mg protein)		77	
≥ 10 fmol/mg protein		286	0.01
PgR			
< 10 fmol/mg protein		60	
≥ 10 fmol/mg protein		288	0.14
Combined receptor			
ER < 10, PgR < 10		68	
ER ≥ 10, PgR < 10		131	
ER < 10, PgR ≥ 10		6	
ER ≥ 10, PgR ≥ 10		143	0.01
Weight			
≤ 65 kg		230	
> 65 kg		140	0.39
Adjuvant treatment			
Age ≤ 50 yrs	Age > 50 yrs		
A. Chemo	Hormones	191	
Hormones	Chemo	65	
Nothing	Nothing	122	0.51
B. Chemo + ovar. ablation + hormones	Hormones	99	
Chemo + hormones	Chemo + ovar. ablation + hormones	14	
Chemo	Chemo + hormones	96	
Hormones	Chemo	47	
Nothing	Nothing	122	0.09

be used to assess a set of maximum likelihood estimates under a particular model-type compared with the estimates from the best model-type considered (see Appendix 1). A significant value of –2logR for a model-type may be considered as evidence against that model.

Results

Univariate analyses of disease-free interval for the node-positive patient group studied here are given in Table 3 for each of the prognostic factors. Tumor size (T), the grouped number of positive nodes, ER (±), and the combined biochemical receptor variable all had significant univariate DFI results.

Table 4 summarizes the Cox regression results from the variety of step-wise and all-subset procedures. There is a substantial increase in –2logR values with models D-F, where quantitative variables were used, versus model A with grouped variable values; similar –2logR values were obtained with one fewer variable by the quantitative variable models of D-F than models B and C, which had a mixture of grouped and quantitative variables. Model F ($-2\log R = 76.75 \sim \chi^2_{(4)}$) was the best Cox step-wise forward model with quantitative variables; the corresponding full model had a –2logR value of 80.54. The difference of 3.79 is less than what would be required for one variable to significantly (P = 0.05) improve the 4-variable model, so the all-subset selections focused on other 4-variable models. Models D and E were identified by the all-subset process and have fairly similar –2logR values to the best model. The three 4-variable models have number of positive nodes, tumour size, and the five-category adjuvant treatment variable in common, and each has a biochemical receptor variable (ER, combined receptor, or PgR). The correlations for the three biochemical variables were -0.147, -0.237, and -0.269, respectively, for (ER, PgR), (ER, combined receptor), and (PgR, combined receptor).

Table 4. Cox multivariate models

Models	-2logR
A. Nodes, T, adjuv, comb receptors	59.49
B. Nodes, ER, T, PgR, adjuv	69.62
C. Nodes, size, PgR, ER, adjuv	75.05
D. Nodes, size, adjuv, ER	75.24
E. Nodes, size, adjuv, ER	75.67
F. Nodes, size, adjuv, PgR	76.75

A. Best standard step-wise with grouped variables:
Nodes positive (1-3, 4-6, 7-10, ≥11)
T (1, 2, 3)
Combined receptors (–/–, +/–, –/+, +/+)
Adjuvant therapy (3 categories)

B. Best step-wise with quantitative ER, PgR (fmol/mg):
Adjuvant therapy (3 categories)

C. Best step-wise and 5-variable all-subset model:
Nodes positive (# positive)
Tumour size (cm)
PgR (fmol/mg)
Adjuvant therapy (3 categories)

D-F. Best step-wise and 4-variables all-subset models:
Nodes positive (# positive)
Tumour size (cm)
ER, PgR (fmol/mg)
Combined receptors (–/–, +/–, –/+, +/+)
Adjuvant therapy (5 categories)

The log (cumulative hazard) plots against time were made to examine the Cox assumption of proportional hazards for the covariates: number of positive nodes, tumour size, adjuvant therapy, and PgR (Figures 1-4). Under the assumption of proportional hazards, plots of log (cumulative hazard) for variable strata should be approximately parallel. There is evidence against proportional hazards for two variables, number of positive nodes and adjuvant treatment; these departures could not be readily handled by stratification.

The results from step-wise, full model, and all-subset variable fits for the exponential, Weibull, log logistic, and log normal models are compared with the Cox results in Table 5. Four variables, number of positive nodes, tumour size, adjuvant treatment, and PgR, were included in all the best models although the –2logR values varied greatly, and the log logistic model had an addi-

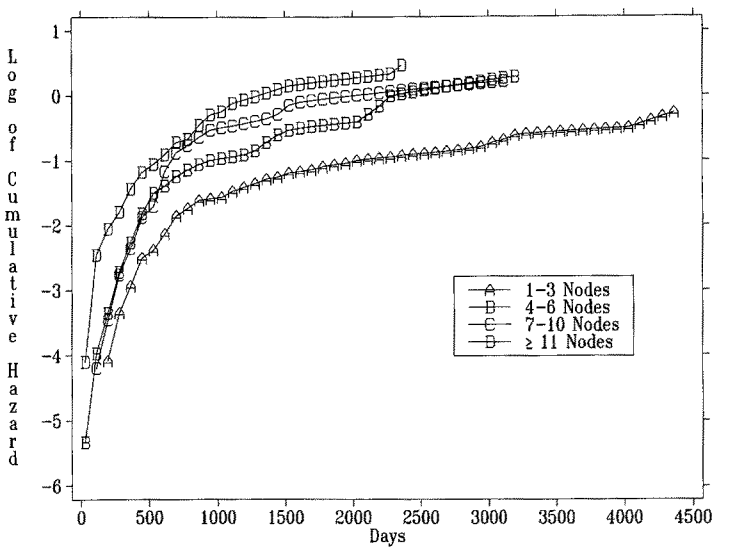

Figure 1. Log cumulative hazard for number of positive nodes

tional variable, combined receptor. The underlying modelling is compared in Table 6, with the data strongly supporting the log normal over other non-parametric and parametric model-types.

The relative merits of the best log normal models are shown in Table 7. The step-wise for-

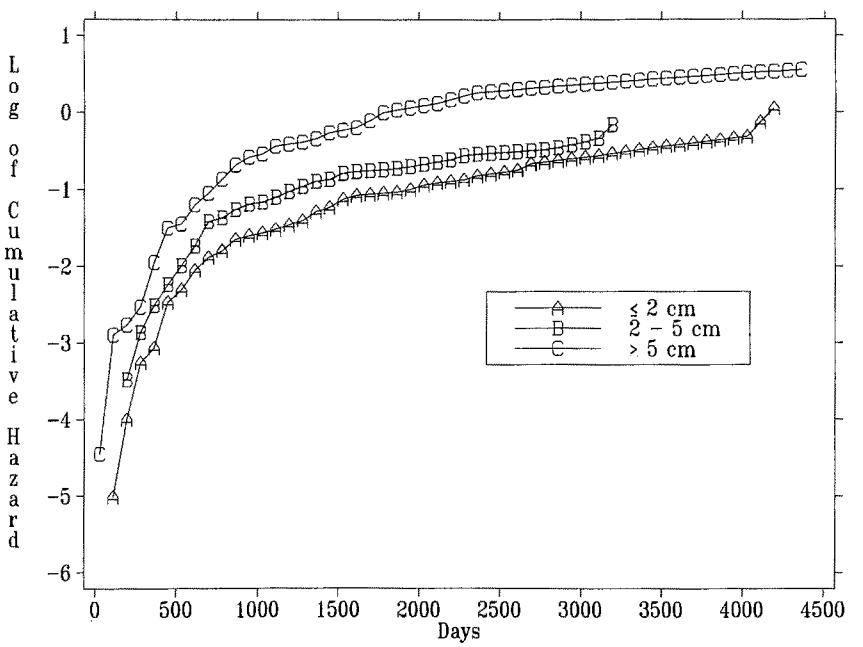

Figure 2. Log cumulative hazard for tumor size

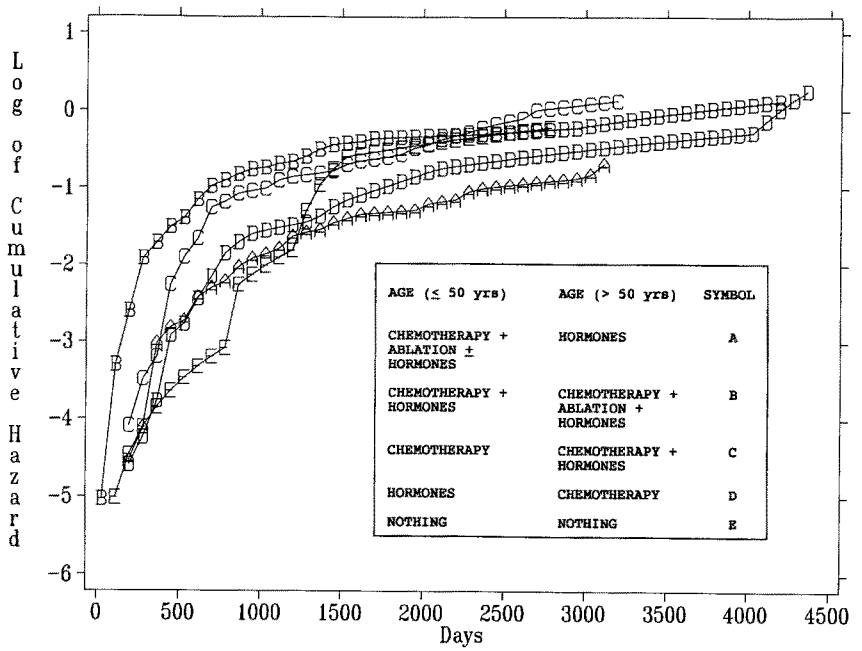

Figure 3. Log cumulative hazard for adjuvant therapy

ward process yielded the 4-variable model C with a $-2\log R \sim \chi^2_{(4)}$ value of 86.81, while the full model had a $-2\log R \sim \chi^2_{(8)}$ value of 92.10. The difference of 5.29 was sufficiently large that all 5-variable subsets were considered. The best 5-variable models were D-G, but none of D-G

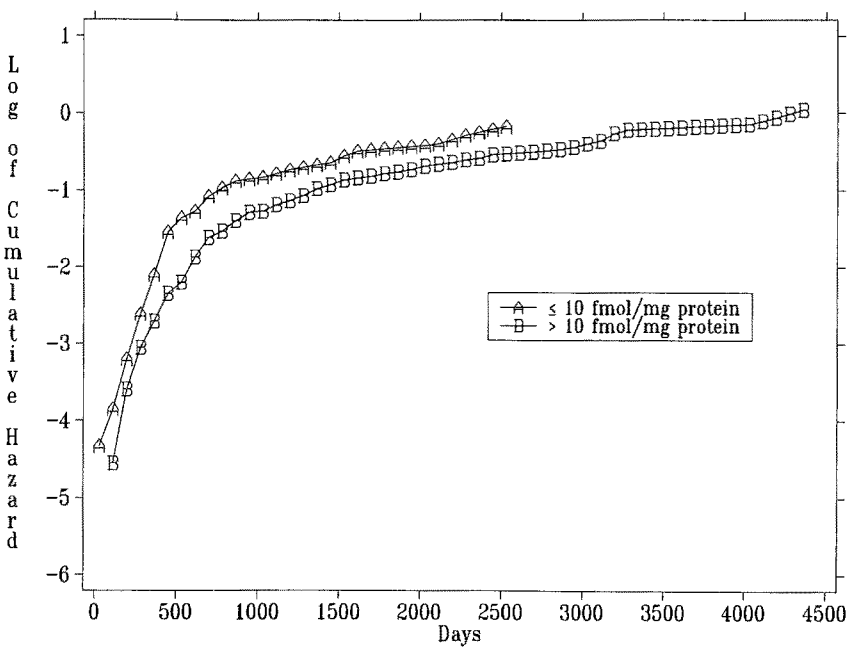

Figure 4. Log cumulative hazard for progesterone receptor

were significantly better than C. (The addition of the combined receptor variable in Model G led to a model improvement with a P = 0.08 value). The 4-variable models A-C had fairly similar –2logR values, three variables (number of positive nodes, tumour size, and adjuvant treatment) in common, and one biochemical receptor variable (ER, combined receptors, or PgR). Table 8 has the standardized parameter estimates for the best 4-variable model. The standardized residuals did not suggest alternative or additional relationships with the covariates.

Discussion

We have compared standard Cox step-wise regressions with all-subset Cox regressions for analyses with standard grouped variable values and quantitative values. The use of quantitative variables improved the step-wise modelling by yielding substantially better models with one less variable.

The all-subset approach was important with the prognostic factors studied here, as there were generally several similarly good models regardless of underlying model-type. The number of positive nodes, tumour size, and adjuvant systemic therapy were significant factors in all models. Others [8-10] have found PgR to be an important prognostic factor singly or in combination with ER, depending on stage of disease and adjuvant treatment. The consistency with which ER, the combined receptors, and PgR resulted in similarly important models in this work suggests caution in adopting the marginally better variable, PgR, as the variable of choice for our data.

The evidence against the Cox assumption of proportional hazards suggested that it might be more appropriate to use an accelerated failure time model, and the data strongly supported the log normal model. The hazard for a log normal model will increase to some point in time after which it will begin to decrease; this is a reasonable underlying description of the clinical expectation for node-positive breast cancer. After 5-10 years, the risk of recurrence may decrease [5]. While the inference with the Cox and log normal models was similar with the prognostic factors investigated here, a difference might have emerged had there been other prognostic factors available to consider adding to the model. The –2logR value with the same 4 variables was only

Table 5. Comparison of multivariate models

Models	–2logR
A. Cox	
Nodes, size, adjuv, PgR	76.75
B. Exponential	
Nodes, size, adjuv, PgR	82.32
C. Weibull	
Nodes, size, adjuv, PgR	82.39
D. Log logistic	
Nodes, size, adjuv, PgR, comb receptors	91.19
E. Log normal	
Nodes, size, adjuv, PgR	86.81

Notes about models:

1. Models are best step-wise or all-subset models for 4 or 5 variables.
2. For all models, the variables are:
   Nodes positive (# nodes positive)
   Tumour size (cm)
   Adjuvant therapy (5 categories)
   PgR (fmol/mg protein)
   Combined receptors (–/–, +/–, –/+, +/+).

Table 6. Comparison of Cox and parametric model-types

Model-type (MT)	Log likelihood	–2logR′	P-value
Cox	–747.69	812.02	< 0.001
Exponential	–351.16	18.96	<0.001
Weibull	–350.77	18.18	<0.001
Log logistic	–344.00	4.64	<0.05
Log normal	–341.68	1.00	—

Notes:

1. Log likelihood (L_{MT}) values are for models with the maximum likelihood parameter estimate for Nodes (# positive), Adjuvant treatment (5 categories), and PgR (fmol/mg protein).
2. For the Cox model, $-2\log R' = -2\log(L_{Cox}/L_{Log\ Normal}) \sim \chi^2_{(2)}$ tests the hypothesis that the estimated intercept and scale parameters of the Log Normal model are zero.
3. For parametric models, $-2\log R' = -2\log(L_{MT}/L_{Log\ Normal}) \leq -2\log R = -2\log(L_{MT}/L_{unrestricted\ best\ model}) \sim \chi^2_{(1)}$.

76.75 with the Cox as compared to 86.81 with the log normal. Conceptually, the best Cox model, with the addition of one or more new prognostic variables, might have yielded a similar overall –2logR model improvement to that achieved with the 4-variable log normal.

The use of all-subset modelling with both the non-parametric Cox and parametric accelerated failure time models (exponential, Weibull, log logistic, and log normal) with multiplicative effects of the covariates on survival time is feasible with the number of variables considered here and using commercial statistical packages. The package ISMOD, which is available for UNIX and S, and is being produced for PCs, would be the preferred mechanism for investigations with a larger number of variables.

Conclusions

Where it is possible to do so, quantitative variable values should be used for multivariate investigations. The underlying assumption of proportional hazards for Cox regressions should be examined, as alternative models for censored data are now readily available. Our data strongly support a log normal model for censored data. Meanwhile, all-subset regressions should be used more routinely in clinical investigations. With the increase in prognostic factors, it becomes even more likely that a single model may not be substantially better than others with the same number of variables, and that combinations of variables not considered in a step-wise procedure could lead to a model that is actually better than the best step-wise.

Table 7. Log normal multivariate models

Models	–2logR
A. Nodes, size, adjuv, ER	84.97
B. Nodes, size, adjuv, comb receptors	86.30
C. Nodes, size, adjuv, PgR	86.81
D. Nodes, size, adjuv, PgR, weight	84.72
E. Nodes, size, adjuv, ER, comb receptors	86.47
F. Nodes, size, adjuv, PgR, ER	87.72
G. Nodes, size, adjuv, PgR, comb receptors	89.95

For all models, the variables are:
Nodes positive (# positive)
Tumour size (cm)
Adjuvant therapy (5 categories)
ER, PgR (fmol/mg protein)
Combined receptors (–/–, +,–, –/+, +/+)
Weight (kg)

Acknowledgements

This work was supported by a Program grant from the Women's College Hospital Research Fund and a grant from the Henrietta Banting Memorial Fund. We gratefully acknowledge the technical assistance of Geneitha Allen, Gladys Oldfield, and Stephanie Oldfield.

Appendix 1

Maximized relative likelihood [4] can be used to assess a set of $\tilde{\alpha},\tilde{\beta},\tilde{\sigma}$ maximum likelihood estimates under a particular parametric model-type compared with the unrestricted $\hat{\alpha},\hat{\beta},\hat{\sigma}$. The likelihood ratio test $-2\log R = -2\log\{L[\tilde{\alpha},\tilde{\beta},\tilde{\sigma}]/L[\hat{\alpha},\hat{\beta},\hat{\sigma}]\} = \chi^2_{(1)}$ can be used for this purpose. Unfortunately, we do not obtain a completely unrestricted set of $\hat{\alpha},\hat{\beta},\hat{\sigma}$ by the process described in this paper. However, we will have $L[\hat{\alpha}',\hat{\beta}',\hat{\sigma}']$ for the best model-type considered which will be $\leq L[\hat{\alpha},\hat{\beta},\hat{\sigma}]$. Therefore, $-2\log R' = -2\log\{L[\tilde{\alpha},\tilde{\beta},\tilde{\sigma}]/L[\hat{\alpha}',\hat{\beta}',\hat{\sigma}']\} \leq \chi^2_{(1)}$. A significant value of –2logR′ for a model-type, may be considered as

Table 8. Standardized parameter estimates for variables from Log Normal model

Variable	$\hat{\beta}$/Standard error	P-value
Number positive nodes	–5.62	< 0.01
Adjuvant treatment	3.71	< 0.01
Tumour size	–4.01	< 0.01
PgR	2.32	< 0.05

evidence against that model.

References

1. Lawless JF, Singhal K: ISMOD: an all-subsets regression program for generalized linear models. I. Statistical and computational background. Computer Methods and Programs in Biomedicine 24:117-124, 1987.
2. Lawless JF, Singhal K: ISMOD: an all-subsets regression program for generalized linear models. II. Program guide and examples. Computer Methods and Programs in Biomedicine 24:125-134, 1987.
3. Kalbfleisch JD, Prentice RL: The Statistical Analysis of Failure Time Data. John Wiley and Sons, New York, 1980.
4. Lawless JF: Statistical Models and Methods for Lifetime Data. John Wiley and Sons, New York, 1982.
5. Harris JR, Henderson IC: Natural history and staging of breast cancer. Harris JR, Hellman S, Henderson IC Kinne DW (eds) Breast Diseases. JB Lippincott, Philadelphia, 1987, pp 233-258.
6. Mobbs BG: Close correlation between progesterone receptor concentration and hormonal sensitivity in DMBA-induced mammary tumours of the rat. Eur J Cancer Clin Oncol 19:835-842, 1983.
7. Meakin JW, Hayward JL, Stewart HJ: Assessment of response and recurrence in breast cancer. Semin Oncol 5:445-449, 1978.
8. Clark GM, McGuire WL: Steroid receptors and other prognostic factors in primary breast cancer. Semin Oncol 15(Suppl 1):20-25, 1988.
9. Fisher B, Redmond CK, Wickerham DL, Rockette HE, Brown A: Relation of estrogen and/or progesterone receptor content of breast cancer to patient outcome following adjuvant chemotherapy. Breast Cancer Res Treat 3:355-364, 1983.
10. Saez S, Cheix F, Asselain B: Prognostic value of estrogen and progesterone receptors in primary breast cancer. Breast Cancer Res Treat 3:345-354, 1983.