Research article

Ratios of involved nodes in early breast cancer

Vincent Vinh-Hung1,2, Claire Verschraegen3, Donald I Promish4, Gábor Cserni5, Jan Van de Steene1, Patricia Tai6, Georges Vlastos7, Mia Voordeckers1, Guy Storme1 and Melanie Royce3

1Oncology Center, Academic Hospital, AZ-VUB, Vrije Universiteit Brussel, Jette, Belgium
2Student Management Healthcare Data, BISI-VUB, Computer Science and Medical Informatics, Vrije Universiteit Brussel, Jette, Belgium
3The University of New Mexico, Cancer Research and Treatment Center, Albuquerque, New Mexico, USA
4Decision Analyst, Burlington, Vermont, USA
5Surgical Pathology, Bács-Kiskun County Teaching Hospital, Kecskemét, Hungary
6Radiation Oncology, Saskatchewan Cancer Agency, Regina, Canada
7Gynecologic Oncology and Senology, Geneva University Hospitals, Geneva, Switzerland

Corresponding author: Claire Verschraegen, cverschraegen@salud.unm.edu

Received: 29 Mar 04 Revisions requested: 7 May 04 Revisions received: 9 Aug 04 Accepted: 26 Aug 04 Published: 6 Oct 2004

Breast Cancer Res 2004, 6:R680-R688 (DOI 10.1186/bcr934)
© 2004 Vinh-Hung et al.; licensee BioMed Central Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Introduction  The number of lymph nodes found to be involved in an axillary dissection is among the most powerful prognostic factors in breast cancer, but it is confounded by the number of lymph nodes that have been examined. We investigate an idea that has surfaced recently in the literature (since 1999), namely that the proportion of node-positive lymph nodes (or a function thereof) is a much better predictor of survival than the number of excised and node-positive lymph nodes, alone or together.

Methods  The data were abstracted from 83,686 cases registered in the Surveillance, Epidemiology, and End Results (SEER) program of women diagnosed with nonmetastatic T1–T2 primary breast carcinoma between 1988 and 1997, in whom axillary node dissection was performed. The end-point was death from breast cancer. Cox models based on different expressions of nodal involvement were compared using the Nagelkerke R² index ($R^2_{N}$). Ratios were modeled as percentage and as log odds of involved nodes. Log odds were estimated in a way that avoids singularities (zero values) by using the empirical logistic transform.

Results  In node-negative cases both the number of nodes excised and the log odds were significant, with hazard ratios of 0.991 (95% confidence interval 0.986–0.997) and 1.150 (1.058–1.249), respectively, but without improving $R^2_{N}$. In node-positive cases the hazard ratios were 1.003–1.088 for the number of involved nodes, 0.966–1.005 for the number of excised nodes, 1.015–1.017 for the percentage, and 1.344–1.381 for the log odds. $R^2_{N}$ improved from 0.067 (no nodal covariate) to 0.102 (models based on counts only) and to 0.108 (models based on ratios).

Discussion  Ratios are simple optimal predictors, in that they provide at least the same prognostic value as the more traditional staging based on counting of involved nodes, without replacing them with a needlessly complicated alternative. They can be viewed as a per patient standardization in which the number of involved nodes is standardized to the number of nodes excised. In an extension to the study, ratios were validated in a comparison with categorized staging measures using blinded data from the San Jose–Monterey cancer registry. A ratio based prognostic index was also derived. It improved the Nottingham Prognostic Index without compromising on simplicity.

Keywords: axillary lymph node ratio, breast neoplasm, functional form, loco-regional, log odds, nodal ratio, Nottingham Prognostic Index, predictive utility, prognostic factors, proportional hazards, proportion based, ratio-based prognostic index, SEER program, staging, survival

Introduction  Breast cancer is the most common neoplasm in women. Nodal status as determined by pathologic examination of lymph nodes has repeatedly been shown to be the single most important predictor of survival in breast cancer [1]. The absolute number of pathologically involved nodes has also been shown to be an important prognostic factor in breast cancer survival [2-6]. The extent of lymph node

L = empirical logistic transform (estimated log odds); nn = number of axillary lymph nodes free from tumor involvement; np = number of pathologically involved axillary lymph nodes; nx = number of axillary lymph nodes examined (excised); $R^2_{N}$ = Nagelkerke R² index.
involvement is incorporated into prognostic indices such as the Nottingham Prognostic Index [7-9] (see Additional files 1,2,3,4,5,6). Old lymph node stage measures categorized cases according to whether they had none, one to three, or four or more involved nodes, and recently according to more detailed subdivisions [10,11] (see Additional files 2,3,4,5 and 7).

However, several authors have noted the inherent confounding by the number of excised nodes [12,13]. To address the variability of nodal examination, an intuitive approach is to use the proportion or the percentage of involved nodes, as was suggested by Rostgaard and coworkers [14]. The proportion can immediately be derived from pathology reports that clearly state the total number of lymph nodes examined and the total number of involved nodes [1,15]. The proportion has received increasing attention in the literature, providing a reference base on which its clinical relevance may be discussed [13,14,16-27].

In this report the modeling utility of the proportion of involved nodes is compared with the absolute numbers of involved nodes and of examined nodes. There is a one-to-one correspondence between proportion and ratio between involved and uninvolved nodes. A previous study hinted at an apparent linear relationship with survival between involved and uninvolved nodes (Fig. 1) [28], and therefore this report also examines the utility of expressing ratios as odds instead of proportions.

The absolute numbers considered in the study were the number of nodes examined (excised; nx), the number of involved nodes (np), and the number of uninvolved nodes (nn).

**Methods**

The SEER (Surveillance, Epidemiology, and End Results) program of the USA [29] provides extensive cancer incidence data from 11 population-based registries. The data used in the present study were extracted from nine of those registries: San Francisco-Oakland, Connecticut, Metropolitan Detroit, Hawaii, Iowa, New Mexico, Seattle (Puget Sound), Utah, and Metropolitan Atlanta.

Selected patients were women without a previous history of cancer who presented with a noninflammatory invasive breast carcinoma, which was histologically confirmed and diagnosed between 1988 and 1997, with specified tumor size no larger than 50 mm (T1 and T2), strictly confined to breasts without distant metastasis, and in which curative surgery and axillary lymph node dissection were performed with removal of at least one node. Cases with involvement of skin, hypodermis or pectoral muscles, or with deep fixation were excluded. Patients who had undergone subcutaneous mastectomy, radical mastectomy, or preoperative or intraoperative radiotherapy were excluded. Data on systemic treatment were not available and therefore could not be taken into account. Certain records were rejected because of data quality concerns: uncertain sequence of treatment, nonhospital based data records, month of diagnosis unknown, or race unknown. Examination of outliers (scarce and extreme values) resulted in further exclusion of cases with more than 50 nodes examined, '0 months' of follow up, and age at diagnosis under 25 years or older than 95 years.

The follow-up cut-off date was 31 December 1999. The survival end event was defined as death from breast cancer.

The proportions of involved nodes were expressed as percentages ([np/nx] × 100%). The log odds of nodal involvement were computed using the empirical logistic transform: L = Loge[(np + 0.5) / (nn + 0.5)] [30]. The transform, also called the sample logit, avoids singularities caused by null observations, and is the least biased estimator of the true log odds [31]. (Note that, with hindsight, Fig. 1 shows a logarithmic relationship.) Unadjusted mortality (the number of patients who died divided by the number of patients at
risk) as a function of the ratios was used for descriptive purposes.

The utilities of the percentage and log odds were evaluated in different multivariate Cox proportional hazards models [32]. The numbers np and nx, the percentage (np/nx) × 100%, and the L transform were entered as quantitative continuous variables in different combinations. The corresponding hazard ratios were each time computed within a Cox model that included tumor size, age at diagnosis, and year of diagnosis modeled as quantitative continuous variables; and the registry area, race, marital status, tumor topography, histologic type and grade, estrogen and progesterone receptor status, type of primary surgery, and administration of postoperative radiotherapy modeled as qualitative variables. The qualitative variables were converted or expanded as needed into dummy variables to allow binary coding ('married' versus 'not married', 'high grade' versus 'not high grade', for example, and so on). A first order interaction between type of surgery and postoperative radiotherapy was included for consistency with a previous analysis [33]. The models were computed in all cases irrespective of nodal status, and then as a function of positive or negative nodal status. The functional forms were

| Percentage of involved nodes (np/nx [%]) | Number of patients at risk | All-cause mortality (%) | Breast cancer mortality (%) | Non-breast-cancer mortality (%) |
|----------------------------------------|---------------------------|-------------------------|----------------------------|---------------------------------|
| 0                                      | 58070                     | 15.9                    | 5.2                        | 10.7                            |
| 1–10                                   | 8695                      | 21.6                    | 11.7                       | 9.9                             |
| 11–20                                  | 6350                      | 26.3                    | 15.7                       | 10.6                            |
| 21–30                                  | 3003                      | 30.8                    | 19.9                       | 10.9                            |
| 32–40                                  | 2055                      | 33.1                    | 22.7                       | 10.4                            |
| 41–50                                  | 1487                      | 41.0                    | 29.6                       | 11.4                            |
| 51–60                                  | 889                       | 37.8                    | 28.2                       | 9.6                             |
| 61–70                                  | 719                       | 47.0                    | 33.7                       | 13.4                            |
| 71–80                                  | 717                       | 50.3                    | 40.3                       | 10.0                            |
| 81–90                                  | 658                       | 52.6                    | 40.4                       | 12.2                            |
| 91–100                                 | 1043                      | 58.6                    | 45.2                       | 13.4                            |

np, number of pathologically involved axillary lymph nodes; nx, number of axillary lymph nodes examined (excised).

| L = empirical logistic transform (estimated log odds). | All patients | Node-negative patients | Node-positive patients |
|------------------------------------------------------|--------------|------------------------|------------------------|
|                                                      | Number of patients at risk | Breast cancer mortality | Non-breast-cancer mortality | Number of patients at risk | Breast cancer mortality | Non-breast-cancer mortality | Number of patients at risk | Breast cancer mortality | Non-breast-cancer mortality |
| L ≤ 4                                                | 3053         | 5.4                    | 9.2                     | 3053                     | 5.4                    | 9.2                     | 0                       | 13.5                     | 7.8                      |
| -4 < L ≤ -3                                          | 44695        | 5.3                    | 10.2                    | 44503                    | 5.2                    | 10.2                    | 192                     | 13.5                     | 7.8                      |
| -3 < L ≤ -2                                          | 17619        | 8.0                    | 11.5                    | 10187                    | 5.4                    | 12.8                    | 7432                    | 11.7                     | 9.7                      |
| -2 < L ≤ -1                                          | 9173         | 15.4                   | 10.9                    | 327                      | 4.0                    | 19.3                    | 8846                    | 15.8                     | 10.6                     |
| -1 < L ≤ 0                                           | 5120         | 24.0                   | 11.0                    | 0                        | 5120                   | 24.0                    | 11.0                    | 11.0                     | 11.0                     |
| 0 < L ≤ 1                                            | 1943         | 31.9                   | 11.3                    | 0                        | 1943                   | 31.9                    | 11.3                    | 11.3                     | 11.3                     |
| 1 < L ≤ 2                                            | 1113         | 40.3                   | 11.1                    | 0                        | 1113                   | 40.3                    | 11.1                    | 11.1                     | 11.1                     |
| 2 < L ≤ 3                                            | 536          | 44.4                   | 13.8                    | 0                        | 536                    | 44.4                    | 13.8                    | 13.8                     | 13.8                     |
| 3 < L ≤ 4                                            | 376          | 47.6                   | 13.0                    | 0                        | 376                    | 47.6                    | 13.0                    | 13.0                     | 13.0                     |
| 4 < L                                                | 58           | 56.9                   | 10.3                    | 0                        | 58                     | 56.9                    | 10.3                    | 10.3                     | 10.3                     |
assessed using the generalized additive model procedure [34].

The Nagelkerke $R^2$ index ($R^2_N$) was used to score the different Cox models [35]. $R^2$ represents the proportion of variation explained by covariates in regression models [35-37]. $R^2_N$ divides $R^2$ by its maximum attainable value to scale it to within the range 0–1. $R^2_N$ is close to 1 for a perfectly predictive model, and close to 0 for a model that does not discriminate between short and long survival times.

Statistical analyses were performed using Splus (Insightful Corporation, Seattle, WA, USA) statistical software.

Results
In the 2002 SEER release [29], 188,410 women were diagnosed with breast tumors from 1988 to 1997, of whom 132,457 had a hospital based histopathologic diagnosis of unilateral invasive carcinoma. A total of 83,686 cases matched the selection criteria; 58,070 were node-negative and 25,616 node-positive. The median follow-up time was 73 months (range 1–143 months) for patients still alive at the follow-up cut-off date (31 December 1999). Characteristics of the patients were presented elsewhere [33]. Except for some additional cases due to updated registration minus the exclusion of outliers resulting in 90 fewer cases, there were no noticeable differences in the distribution of the characteristics.

The median number of nodes examined was 15 (range 1–50, mean ± standard deviation 15.4 ± 6.5). Among the node-positive patients, the median number of involved nodes was 2 (1–46, 4.1 ± 4.8).

Table 1 shows the distribution of the percentages of involved nodes. Figure 2 is a plot of the corresponding breast cancer mortality, which appears to increase linearly with the np/nx percentage.

Table 2 shows the distribution of the log odds for nodal involvement. Figure 3 plots the corresponding breast cancer mortality. There is an initial, almost flat segment for values of L ≤ -3, which is followed by a steeply sloping upward segment. The initial flat segment corresponds mostly to node-negative cases. The sloping upward segment corresponds to node-positive cases, with more positive L values indicating more involved nodes and/or fewer uninvolved nodes. There is an overlap between node-negative and node-positive cases for L values between -3.5 and -1.

In multivariate analyses, np and nx exhibited marked nonlinearity and widely diverging confidence intervals (Fig. 4a,4b). The linearity improved for the percentage (np/nx) × 100% and the L transform, which also showed more homogeneously distributed confidence intervals (Fig. 4c,4d).

Table 3 shows a comparison between proportional hazards models that included different combinations of np, nx, np/nx, and L for all patients, irrespective of nodal status. Based on $R^2_N$, the best predictive covariate was L (model 6), with a small improvement contributed by nx (model 10). In the model with nx alone (model 3), nx was statistically significant but its contribution to global model fit appeared negligible because the $R^2_N$ did not change from the baseline 0.069 (model 1). The contribution of np alone was substantial, with a change of $R^2_N$ from baseline 0.069 to 0.093 (model 2). However, adding np and/or nx onto L or onto np/nx provided no improvement, except in the already mentioned model 10.

The middle section of Table 3 shows multivariate analysis performed for node-positive cases only. Models based on separately expressed numbers provided the lowest $R^2_N$ (models 2–4). The largest $R^2_N$ values were all observed in models incorporating L or np/nx (models 5–11). The simplest model appeared to be based on np/nx alone (model 5; $R^2_N = 0.108$). A small improvement was contributed by np (model 7; $R^2_N = 0.109$).
The lower section of Table 3 shows the analysis performed for node-negative cases only. Because np, by definition, equals 0, there are only four models. They show that nx (model 3) and L (model 6) are statistically significant, but these variables either alone or in combination did not improve the index $R^2_N$.

The multivariate computations were also performed by considering death from any cause as the end-point. There were no notable discrepancies.

Discussion
Although many data were evaluated in the present study, there are weaknesses. The data are heterogeneous. Histopathologic characteristics such as grade could not be verified. Neoadjuvant systemic treatment might have modified the yield of nodes [38]. Important information such as how patients were selected for any particular treatment is missing. Undocumented comorbidity might have affected the extent of nodal dissection. An imbalance in the delivery of chemotherapy or hormone therapy could have affected the distribution of deaths. For all of these reasons, the present results should be considered explorative and must be validated independently.

Since about 1999 a growing number of studies have investigated nodal ratios. In the studies that compared the numbers of involved nodes with ratios in multivariate models, the majority found that ratios were better than numbers as prognostic indicators [13,16,18,19,24,26]. Ratios (expressed as percentages or log odds) have a better prognostic impact than do isolated numbers and, unlike numbers, they are not associated with inconsistent findings. Part of the explanation might be that a ratio can be interpreted as a form of standardization in which the number of involved nodes found in a patient is standardized to the number of nodes examined in that same patient [20]. It is noteworthy that the hazard ratios for np/nx were almost unaffected by the model (column np/nx [%] in Table 3),
whereas the hazard ratios for \( np \) and/or \( nx \) exhibited more variability (columns \( np \) and \( nx \) in Table 3).

As a prognostic factor, \( np/nx \) appears the most convenient. Figure 2 shows that, for node-positive cases presenting with 0–10% involved nodes, the crude breast cancer mortality risk for an average follow up of 6 years is about 5%, and with 90–100% involved nodes the mortality is about 45%. For any intermediate value for the percentage of node involvement, the mortality risk is easily interpolated.

### Table 3

| Patients studied | \( R^2_N \) | Hazard ratios (95% confidence interval) |
|------------------|-----------|----------------------------------------|
|                  |           | \( np \) | \( nx \) | \( np/nx (\%) \) | \( L \) |
| All cases        |           |         |         |                 |       |
| 1: no nodal variables | 0.069    |         |         |                 |       |
| 2: \( np \)       | 0.093    | 1.093 (1.090–1.097) |         |                 |       |
| 3: \( nx \)       | 0.069    | 0.998 (0.995–1.001) |         |                 |       |
| 4: \( np, nx \)   | 0.096    | 1.112 (1.108–1.116) | 0.970 (0.966–0.974) | 1.023 (1.022–1.024) | 1.459 (1.442–1.477) |
| 5: \( np/nx (\%) \) | 0.104    |         |         |                 |       |
| 6: L             | 0.108    |         |         |                 |       |
| 7: \( np, np/nx (\%) \) | 0.104    | 1.010 (1.004–1.017) | 1.021 (1.020–1.023) |       |
| 8: \( nx, np/nx (\%) \) | 0.104    | 1.002 (0.999–1.005) | 1.023 (1.022–1.024) |       |
| 9: \( np, L \)    | 0.108    | 0.998 (0.992–1.005) |         |                 | 1.466 (1.435–1.497) |
| 10: \( nx, L \)   | 0.109    | 1.009 (1.006–1.013) |         |                 | 1.462 (1.445–1.479) |
| 11: \( np, nx, np/nx (\%) \) | 0.104    | 1.013 (1.004–1.022) | 0.998 (0.994–1.002) | 1.021 (1.019–1.022) |
| 12: \( np, nx, L \) | 0.109    | 0.957 (0.957–0.976) | 1.021 (1.017–1.026) |         | 1.599 (1.554–1.646) |
| Node-positive patients |         |         |         |                 |       |
| 1: no nodal variables | 0.067    |         |         |                 |       |
| 2: \( np \)       | 0.095    | 1.066 (1.062–1.071) |         |                 |       |
| 3: \( nx \)       | 0.067    | 0.996 (0.992–1.000) |         |                 |       |
| 4: \( np, nx \)   | 0.102    | 1.088 (1.083–1.093) | 0.966 (0.961–0.971) |       |
| 5: \( np/nx (\%) \) | 0.108    |         |         |                 |       |
| 6: L             | 0.108    |         |         |                 |       |
| 7: \( np, np/nx (\%) \) | 0.109    | 1.013 (1.006–1.020) | 1.015 (1.014–1.017) |       |
| 8: \( nx, np/nx (\%) \) | 0.108    | 1.005 (1.001–1.009) | 1.017 (1.016–1.018) |       |
| 9: \( np, L \)    | 0.108    | 1.008 (1.001–1.016) |         |                 | 1.344 (1.306–1.384) |
| 10: \( nx, L \)   | 0.108    | 1.005 (1.001–1.009) |         |                 | 1.381 (1.357–1.405) |
| 11: \( np, nx, np/nx (\%) \) | 0.109    | 1.016 (1.005–1.028) | 0.998 (0.991–1.004) | 1.015 (1.013–1.017) |
| 12: \( np, nx, L \) | 0.108    | 1.003 (0.990–1.017) | 1.003 (0.996–1.010) |         | 1.386 (1.304–1.431) |
| Node-negative patients |         |         |         |                 |       |
| 1: no nodal variables | 0.045    | NA      |         |                 |       |
| 3: \( nx \)       | 0.045    | NA      | 0.991 (0.986–0.997) |       |
| 6: L             | 0.045    | NA      |         |                 |       |
| 10: \( nx, L \)   | 0.045    | NA      | 1.001 (0.984–1.019) | 1.169 (0.902–1.514) |       |

Shown are all nodal status combined, and node-positive patients and node-negative patients separately. A hazard ratio >1 indicates increased risk for death from breast cancer. All models are multivariate, adjusting for the effect of covariates listed in the Methods section: tumor size, age at diagnosis, year of diagnosis, registry area, race, marital status, tumor topography, histologic type and grade, estrogen and progesterone receptor status, type of primary surgery, and administration of postoperative radiotherapy. NA, not applicable; \( np \), number of involved nodes; \( nx \), number of nodes examined; \( np/nx (\%) \), percentage of nodes involved; \( L \), log-odds of node involvement; \( R^2_N \), Nagelkerke \( R^2 \) index of global model fit (0 = lack of fit, 1 = perfect fit).
The estimated log odds L provided results very similar to those with np/nx. Overall, L improves on np/nx when all cases are considered together (column L in Table 3). The log odds appears useful for integrating node-negative and node-positive cases while avoiding more complex modeling, which we performed previously [39]. However, there is a range of L values in which node-negative and node-positive cases overlap (Table 2, Figure 3). In an analysis of all cases pooled, the overlap might blur the prognostic difference between node-negative status based on a very small number of excised nodes, and node-positive status based on a large number of excised nodes but with few involved nodes. The literature on the log odds of node involvement is scarce, and the utility of the L transform needs independent confirmation.

The present findings indicate that the favorable survival attributed to higher numbers of nodes removed, as suggested by Krag and Single [40], might be due to different model specifications. The number of patients is huge and statistical significance can easily be demonstrated but without necessarily implying any major clinical impact. Undoubtedly, the uncertainty about node negativity increases when nx (the number of excised nodes) is small. However, the predictive utility attributable to nx is exceedingly small (Table 3, lower section). This dissociation between statistical significance and predictive utility appears counterintuitive. Nevertheless, it is in keeping with findings from Fisher and coworkers [41], who noted that prognosis was unaffected by the number of excised nodes when nodal status was reported to be negative. This is also supported by a recent report based on 3800 patients [42] in which the number of excised nodes was predictive of the risk for recurrence in node-positive but not in node-negative patients.

Sentinel node biopsy has gained wide acceptance since 1997 and it is used to determine the need for axillary dissection [43]. Because our selection of patients was from 1988 until 1997, it is unlikely that sentinel nodes could have represented any substantial part of the present study. The prognostic impact of one involved node in patients who had one node removed in this study cannot be extrapolated to the patient found with one involved node in a sentinel node procedure. However, in the prediction of nonsentinel node involvement when one or more sentinel nodes are found to be involved, Cserni and coworkers [44] reported that the number of sentinel nodes and the percentage of positive sentinel nodes were jointly significant predictors. A closely related finding that also highlights the predictive role of ratios was reported in a recent Australian study [45], in which the prediction model was determined by patient age, by the number of sentinel nodes, and by the proportion of involved sentinel nodes.

**Conclusion**
We found the percentage of involved nodes to be the most directly useful indicator of nodal involvement, but this is limited to node-positive cases. The log odds of nodal involvement performed equally well in node-positive and node-negative patients. The log odds might provide a unified approach to the modeling of nodal involvement. The present results and the growing literature argue that ratios should be considered in the staging of axillary dissection.

**Competing interests**
The author(s) declare that they have no competing interests.

### Additional files

The following Additional files are available online:

**Additional File 1**
Text on deriving a ratio based prognostic index.
See http://www.biomedcentral.com/content/supplementary/bcr934-S1.doc

**Additional File 2**
Table evaluating nodal staging measures using breast cancer data from the San Jose–Monterey registry: all cases irrespective of nodal status.
See http://www.biomedcentral.com/content/supplementary/bcr934-S2.doc

**Additional File 3**
Table evaluating nodal staging measures using breast cancer data from the San Jose–Monterey registry: node-positive patients.
See http://www.biomedcentral.com/content/supplementary/bcr934-S3.doc

**Additional File 4**
Table providing a simulation of small datasets of 300 breast cancer patients, irrespective of nodal status.
See http://www.biomedcentral.com/content/supplementary/bcr934-S4.doc

**Additional File 5**
Table providing a simulation of small datasets of 300 node-positive breast cancer patients.
See http://www.biomedcentral.com/content/supplementary/bcr934-S5.doc

**Additional File 6**
Figure showing Kaplan–Meier survival estimates for T1–T2 breast cancer abstracted from the San Jose–Monterey registry.
See http://www.biomedcentral.com/content/supplementary/bcr934-S6.pdf

R686
Acknowledgements

We are grateful to the population that made the SEER data available. We are grateful to the reviewers who placed signposts to new directions, gave precious information, raised the research stakes, and finally sent us to explore unexpected realms. Use of the R® owes to queries and answers from Lawrence Husnick and Terry Therneau posted on the mail-list s-news@lists.biostat.wustl.edu. Part of this work was originally discussed in the onco@yahoogroups.com, an experimental mail-list for data studies in oncology. Gabor Cserni was supported by a Janos Bolyai Research Fellowship from the Hungarian Academy of Sciences. This work was initially presented at the 26th Annual San Antonio Breast Cancer Symposium [46].

References

1. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, Ruby SG, O'Malley F, Simpson JP, Connolly JL, et al.: Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000, 124:9678.

2. Atkinson EN, Brown BW, Montague ED: Work was initially presented at the 26th Annual San Antonio Breast Cancer Research Fellowship from the Hungarian Academy of Sciences. This work was originally discussed in the onco@yahoogroups.com, an experimental mail-list for data studies in oncology. Gabor Cserni was supported by a Janos Bolyai Research Fellowship from the Hungarian Academy of Sciences. This work was initially presented at the 26th Annual San Antonio Breast Cancer Symposium [46].

3. Carter CL, Allen C, Henson DE: A modified Nottingham prognostic index for breast cancer patients diagnosed in Denmark 1978–1994. Acta Oncol 2001, 40:838-843.

4. Sloane JP: Quality assurance guidelines for pathology in mammography screening. Open biopsy and resection specimens. In European Guidelines for Quality Assurance in Mammography Screening Edited by: Ferry N, Broders M, de Wolf C, Törnberg S, Luxemburg: European Communities; 2001:173-212.

5. Sant M, Allemani C, Capocaccia R, Hakulinen T, Aareleid T, Coe- yellow P, Julious S, Laux M: Prognostic importance of the axillary lymph node ratio in autologous stem-cell transplantation. Clin Cancer Res 1999, 5:3425-3431.

6. Rostgaard K, Mouridsen HT, Vaeth M, Holst H, Olesen KP, Lynge E: A modified Nottingham prognostic index for breast cancer patients [abstract 611]. Proc Am Soc Clin Oncol 2001, 20:329b [http://www.asco.org/lej/1/1003_12_002490-00_18-002001-00_19-003069,00.asp].

7. Martinez-Trufero J, Artal-Cortes A, Zorrilla M, Puertolas T, Vicente Alonso V, Herrero A, Corral M, Maurel J, Anton A: A new distinctive feature on the study of prognostic factors in patients with locally advanced breast cancer (LABC) treated with neoadjuvant chemotherapy (NAC) [abstract 3069]. Proc Am Soc Clin Oncol 2001, 20:329b [http://www.asco.org/lej/1/1003_12_002490-00_18-002001-00_19-003069,00.asp].

8. Schmoor C, Sauerbrei W, Baster G, Bojar H, Schumacher M: Long-term prognosis of breast cancer patients with more positive lymph nodes treated with CMF. Eur J Cancer 2001, 37:1123-1131.

9. Schneeweiss A, Goerner R, Hensel MA, Lauschner I, Sinn P, Kaul S, Egerer G, Beldermann F, Geberth M, Solomayer E, et al.: A predictive model for relapse after high-dose chemotherapy with autologous stem-cell transplantation for stage IV oligometastatic breast cancer. J Clin Oncol 2001, 19:707-718.

10. van der Wal BCh, Butzelaar RMJ, van der Meij S, Boermeester MA: Axillary lymph node ratio and total number of removed lymph nodes: predictors of survival in stage I and II breast cancer. Eur J Surg Oncol 2002, 28:461-468.

11. Grills IS, Kestin LL, Goldstein N, Mitchell G, Martinez A, Ingold J, Vicini FA: Risk factors for regional nodal failure after breast-conserving therapy: regional nodal irradiation reduces rate of axillary failure in patients with four or more positive lymph nodes. Int J Radiat Oncol Biol Phys 2003, 56:658-670.

12. Sanchez Costa L, Soares HP, Gaspar HA, Trujillo LG, Santi P, Pereira RS, de Santana TL, Pinto FN, del Giglio A: Ratio between positive lymph nodes and total dissected axillaries lymph nodes as an independent prognostic factor for disease-free survival in patients with breast cancer. Am J Clin Oncol 2004, 27:304-306.

13. Langoj M, Lingnau A, Dagnall A, Tu TT, Larochelle M: The value of the number of nodes removed (or the ratio of involved nodes) as a prognostic factor in breast cancer [abstract 611]. Proc Am Soc Clin Oncol 2004, 22:29.

14. Vinch-Hung V, Cserni G, Burzykowski T, Van de Steene J, Voordecker M, Storme G: Effect of the number of uninvolved nodes on survival in early breast cancer. Oncol Rep 2003, 10:363-368 [http://147.35.2.10/1003/volume10/number2/363.pdf].

15. National Cancer Institute: Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973–1999), National Cancer Institute, DCCPS, Surveillance Research Pro-
gram, Cancer Statistics Branch, released April based on the November 2001 submission Bethesda, MD: National Cancer Institute; 2002.

30. Vinh-Hung V, Storme G: From data mining to biological hypotheses: modelling the effect of nodes and tumour size in breast cancer. In Bulletin of the International Statistical Institute, Contributed Papers 2004:586-587 [http://www.isi-2003.de/request?De284d7f6b00fc17/Mival=WhosWhoView&PRSSID=7344]. Berlin: International Statistical Institute.

31. Agresti A: Logistic regression. In Categorical Data Analysis 2nd edition. Hoboken, NJ: Wiley; 2002:165-210.

32. Therneau TM, Grambsch PM: Testing proportional hazards. In Modeling Survival Data: Extending the Cox Model New York, NY: Springer-Verlag; 2000:127-152.

33. Vinh-Hung V, Burzykowski T, Van de Steene J, Storme G, Soete G: Post-surgery radiation in early breast cancer: survival analysis on registry data. Radiother Oncol 2002, 64:281-290.

34. Therneau TM, Grambsch PM: Functional form. In Modeling Survival Data: Extending the Cox Model New York, NY: Springer-Verlag; 2000:87-126.

35. Harrell FE Jr. Cox proportional hazards. In Regression Modeling Strategies with Applications to Survival Analysis and Logistic Regression Charlottesville, VA: University of Virginia; 2000:453-499.

36. Schenper M: Predictive accuracy and explained variation. Stat Med 2003, 22:2299-2308.

37. Royston P, Sauerbrei W: A new measure of prognostic separation in survival data. Stat Med 2004, 23:723-748.

38. Baslaim MM, Al Malik OA, Al Sobhi SS, Ibrahim E, Ezzat A, Ajarim D, Tulbah A, Chauchardy MA, Sorbris RA: Decreased axillary lymph node retrieval in patients after neoadjuvant chemotherapy. Am J Surg 2002, 184:299-301.

39. Vinh-Hung V, Burzykowski T, Cserni G, Voordeckers M, Van de Steene J, Storme G: Functional form of the effect of the numbers of axillary nodes on survival in early breast cancer. Int J Oncol 2003, 22:697-704 [http://147.52.72.117/IJO/2003/volume22/number3/697.pdf].

40. Krag DN, Single RM: Breast cancer survival according to number of nodes removed. Ann Surg Oncol 2003, 10:1152-1159.

41. Fisher B, Wolmark N, Bauer M, Redmond C, Gehradt M: The accuracy of clinical nodal staging and of limited axillary dissection as a determinant of histologic nodal status in carcinoma of the breast. Surg Gynecol Obstet 1981, 152:765-772.

42. Thieleke W, Janni W, Rack B, Strobl B, Riosk D, Sommer H, Gerber B, Fries K: The number of resected axillary lymph nodes (ALN) influences the risk for axillary recurrences in node-positive, but not in node-negative patients [abstract 396]. Eur J Cancer Suppl 2004, 2:171.

43. Edge SB, Niland JC, Bookman MA, Theriault RL, Ottesen R, Liao E, Weeks JC: Emergence of sentinel node biopsy in breast cancer as standard-of-care in academic comprehensive cancer centers. J Natl Cancer Inst 2003, 95:1514-1521.

44. Cserni G, Burzykowski T, Vinh-Hung V, Kocsis L, Borose G, Sinko M, Tarjan M, Bori N, Rajtar M, Tekle E, et al.: Axillary sentinel node and tumour-related factors associated with non-sentinel node involvement in breast cancer. Jpn J Clin Oncol 2004 in press.

45. Farshid G, Pradhan M, Kollias J, Gill P: Independent predictors of non-sentinel node involvement in women with breast cancer and at least one positive sentinel node. Volume of metastatic disease in sentinel node, patient age and number of sentinel nodes are predictive of non-sentinel node involvement in a multivariate analysis. The Breast 2004 in press.

46. Vinh-Hung V, Verschraegen C, Voordeckers M, Van de Steene J, Storme G: Proportion, percentage, or ratio of axillary nodes: which is the best prognostic factor? [abstract 524]. Breast Cancer Res Treat 2003, 829(suppl):S126.