Workplace Project Portfolio

A comparison of the 7th edition American Joint Committee on Cancer (AJCC) and N1S3 nodal staging systems for metastatic cutaneous squamous cell carcinoma

Jonathan Clark

Statistical content Supervisor:

Pavlina Rumcheva, PhD

Lecturer in Biostatistics

School of Public Health University of Sydney

Table of Contents

Preface

Student's role

My role included formulation of the study question, treatment of patients contained within the database, collection of patient data, submission to the institutional ethics committee, requesting data from the corresponding data managers, merging and cleaning data sets, literature review to determine appropriate variables for multivariable models and to determine the optimal methodology to answer the study question. I performed the data analysis and modelling, model assessment, model diagnostics and clinical interpretation of the analysis. Following this I wrote the project report in the form of a scientific manuscript suitable for a peer-reviewed journal submission.

Reflection on learning

Communication skills and work planning:

The primary communication challenges in this project were crossing the barriers between clinician and statistician. I needed to describe the clinical research question and background to my statistical supervisor and write the manuscript in a manner suitable for a clinical journal appropriate for a non-statistician to read. Communicating the research question to my supervisor was relatively straight-forward, and my supervisor was very helpful in developing the approach to analysis and presentation of results. However, preparing this project for a clinical journal so that the results not only could be understood by a non-statistician but also be of interest was challenging. The central nature of the statistical analysis to the results and discussion differed from prior research projects I have undertaken.

As a full-time surgeon, planning meetings with my statistical supervisor was challenging. The flexibility of my supervisor enabled us to meet on multiple occasions to discuss the general research aims and methodology. We used email as the primary method of assessing progress and discussion of specific statistical techniques, their application and presentation in this project. Whilst in principle clinical research is encouraged, there is no practical support from hospitals to facilitate this type of research undertaken by surgeons. Therefore this project was undertaken primarily on nights and weekends amidst a busy work schedule trying to balance family commitments. Delays in receiving institutional ethics approval made commencing this project difficult even though the data was contained within established databases. These constraints mean that efficient time management, planning and flexibility were essential to meet project deadlines.

Statistical principles and methods:

This project has expanded my knowledge in a number of areas related to survival analysis. In particular, multiple texts and articles were consulted to begin to understand the appropriate application of frailty models even though this had been covered to some degree in the survival analysis unit of study. Determining whether clustering was appropriate outside of repeated events analysis was a major issue of concern. The use of failure rates was a familiar concept but not a technique that I had previously encountered during the survival analysis unit. This provided a useful graphical adjunct to the usual Kaplan Meier survival curves and was easier to apply confidence intervals where the Kaplan Meier curves became too cluttered when confidence intervals were included. My statistical supervisor suggested using lift

curves to compare the staging systems ability to capture deaths due to disease. I was not familiar with the use of lift curves for other statistical models and do not believe that this has been applied to survival analysis previously. The alternative of a time-adjusted receiver operator curve was also considered. The first decision to be made was whether raw stage should be used to rank patients or whether this should be based on an adjusted hazard coefficient. In the end I decided to include both. The second problem was that using death due to disease alone did not take into consideration that a death after a shorter period of time implies a worse outcome than death after a long period of time. To try to account for this, which is critical in most time to event analyses, I decided to weight death inversely by the time to death from treatment. As there was no literature on the use of lift curves for survival analysis, I could not be sure whether this was a valid approach.

Apart from these specific statistical techniques, a major challenge for this project was deciding what criteria should be used to compare the staging systems. A literature review was performed which was helpful to determine the broad criteria. The Cox proportional hazards model does not allow assessment of all of the predetermined criteria, however the models were helpful in assessing whether staging systems are monotonic and linear in terms of increasing risk with increasing stage. Deciding whether the staging systems should be assessed as raw or adjusted variables was also challenging since there are differing arguments for either approach. To try and overcome this, the final decision was made to assess the staging systems as both raw and after adjusting for appropriate covariates.

The statistical methods applied in this project were variable exploration and distribution, univariable comparisons using Kaplan-Meier survival estimates and the Log-rank test, applying the Cox proportional hazards model for multivariable comparisons. Within the Cox models the important principles included appropriate selection of covariates, assessment of the proportional hazards assumptions, assessment of linearity assumptions for continuous variables and appropriate transformations, assessment of potential significant interactions, assessment of influential outliers and consideration of frailty models.

Statistical computing:

The data was extracted from Microsoft excel and SPSS databases and imported to Stata version 11 for analysis. This provided an opportunity to improve my knowledge of Stata data cleaning, manipulation, analysis and generation of appropriate graphs.

Teamwork

Communication with other team members

As I was responsible for the majority of the project there was very little teamwork required for this project other than discussion with my supervisor as discussed above. The data was sourced from two different data managers and after identification of potential influential outliers, I consulted the respective clinician to confirm the veracity of the data.

Working within timelines:

A set timeline and schedule was not created for this project. Since there were no other clinicians or statisticians directly involved with this project, I was not dependent on

others performing analyses or review of the analysis before proceeding. Analyses were performed on a weekly basis followed by presentation of interim results to the statistical supervisor and email discussion.

Ethical considerations

The data was de-identified and databases are maintained in a secure environment. Patients give consent for their data to be collected and added to the database. As I was not privy to any identifying variables, the only patient confidentiality issue arose when potential outliers were identified. This was managed by using database codes rather than requesting names or medical record numbers from the data managers. An alternative issue arose when one of the interim models suggested that patient outcomes were better at one institution compared to the other. Reporting this sort of information is associated with a myriad of concerns ranging from clinician ire to patient distress over where they were treated. Furthermore, this analysis demonstrated how easily introduction of an additional variable (in this case an interaction term) can change the interpretation of the results. Comparing outcomes between treating institutions or clinicians requires complex multilevel modelling and close scrutiny of both the results and their interpretation is essential before presentation.

Front Sheet

Title

A comparison of the 7th edition American Joint Committee on Cancer (AJCC) and N1S3 nodal staging systems for metastatic cutaneous squamous cell carcinoma

Location

This project combined data from the Sydney Head and Neck Cancer Institute (SHNCI) and Westmead Head and Neck Cancer Centre (WHNCC) databases.

Dates

March – June 2011

Context

This project utilises data from two Australian cancer centres, one of which is the SHNCI, Royal Prince Alfred Hospital where I currently practice as a Head and Neck oncologic and reconstructive surgeon. Metastatic cutaneous cancer is a common condition encountered within Australia but is relatively uncommon in many other parts of the world. Until recently, nodal metastases were not given any prognostic stratification within the AJCC TNM staging system other than being present (N1) or absent (N0). The latest (7th) edition of the AJCC staging manual introduced complex nodal staging criteria in line with mucosal cancer of the Head and Neck. This was in response to a number of alternative staging systems recently published, some of which originated from the SHNCI. Australian cancer centres are in a unique position to examine whether the current AJCC staging system is an advance in the optimal staging of this malignancy.

Contribution of student

- Contributed to management of patients and collection of data pertaining to subjects within the SHNCI database

- Formulation of study aims and hypotheses, literature review and determination of appropriate confounding variables
- Obtained data from data managers of respective cancer centres
- Merged and cleaned data sets
- Created staging variables for data set, Data exploration and description
- Univariable comparisons and survival curves to determine variables for multivariable models. Comparison of failure rates.
- Creation of multivariable models and predicted survival curves
- Model diagnostics and examination of outliers
- Under direction of statistic supervisor – creation of lift curves to compare performance of staging systems
- Discussion with supervisor regarding statistical methods

Statistical issues

- Determining methodology to compare staging systems
- Univariable survival analysis
- Selection of appropriate variables for multivariable models
- Multivariable model diagnostics and testing model assumptions
- Exploration of potential frailty models
- Understanding use of lift curves for assessing staging systems efficiency and application of lift curves to survival analysis

Student declaration

I declare this project is evidence of my own work, with direction and assistance provided by my project supervisor. This work has not been previously submitted for academic credit.

…….. Jonathan Clark

Supervisor's Statement

Jonathan Clark has completed this research project independently with little supervision in selecting and applying the appropriate statistical methods and procedures. The delivered portfolio serves as evidence of Jonathan's ability to carry out a research study from start to finish, successfully managing any challenges arising in the communication of ideas between researchers with diverse backgrounds, aggregation and preparation of data for the statistical analyses, and implementation of alternative approaches to properly address and answer the primary research question. His work on examining and comparing previously established and newly proposed staging systems for metastatic cutaneous squamous cell carcinoma delivered important results and conclusions, and provided valuable basis for further research.

……..

Pavlina Rumcheva

A comparison of the 7th edition American Joint Committee on Cancer (AJCC) and N1S3 nodal staging systems for metastatic cutaneous squamous cell carcinoma

Abstract

Background: The American Joint Committee on Cancer (AJCC) substantially changed the staging of cutaneous squamous cell carcinoma (cSCC) in the 7th edition of its staging manual. We aim to compare the 7th edition AJCC staging of nodal metastases from cSCC with the 'N1S3' staging system.

Methods: Analysis of 603 patients from two prospective cancer centre databases. Multivariable analysis was performed using a Cox proportional hazards model adjusting for the effect of immunosuppression, treating institution, adjuvant radiotherapy, nodal margins and extracapsular spread. Criteria used for comparing staging systems were distribution of patients, stratification of patients according to risk of death from cSCC and model performance.

Results: The N1S3 staging system functioned well in terms of distribution and stratification of patients. The distribution of patients within the AJCC staging system was problematic with three groups (N2a, N2c and N3) containing less than 10% of patients without any prognostic relevance. Stratification of patients within the AJCC staging system was poor in terms of monotonicity (N2c) and distinctiveness (N2a). The performance of the AJCC and N1S3 staging systems was similar despite the AJCC staging being more complex.

Conclusion: The N1S3 staging system is preferred on the grounds of distribution, stratification and parsimony.

Introduction

Non-melanoma skin cancer (NMSC) is the most common malignancy in Australia and the majority of NMSCs occur on sun exposed regions, such as the head and neck. Cutaneous squamous cell carcinoma (cSCC) constitute 25% of NMSC and the incidence directly relates to proximity to the equator, ranging from 16/100,000/year in central Europe to 300/100,000/year in Australia[1]. Within Australia, the highest incidence is seen in Northern Queensland, where the annual rate exceeds 1300/100,000 males[2, 3]. Despite the frequency of cSCC, nodal metastases occur in less than 5% of patients [3, 4] and there are relatively few studies on metastatic cSCC of the head and neck with large enough samples to power reliable conclusions about the behaviour of this disease.

Head and Neck mucosal SCC (mSCC) is less common than cSCC in most countries, particularly Australia, but mSCC exhibits a much greater propensity for developing nodal metastases. In 1977 the American Joint Committee on Cancer (AJCC) introduced a staging system for mucosal cancer in the first edition of its staging manual that stratifies patients according to the estimated to the risk of death from nodal metastases (N stage) [5] and this has been modified over time. Currently, nodal metastases from mSCC are divided into three main groups (N1, N2 and N3) on the basis of the size and number of lymph nodes involved. N2 is further sub-divided into three groups (N2a, N2b and N2c) on the basis of lymph node number and their laterality (side of the neck) giving a total of six groups including N0, which denotes the absence of nodal metastases. In contrast, until recently nodal metastases from cSCC have been divided into only two groups by the AJCC: N0 in patients with no nodal metastases and N1 in patients with nodal metastases.

The discrepancy between mucosal and cutaneous SCC staging systems prompted several alternative staging systems for cSCC to be evaluated [6-10]. Most of these alternative staging systems have stratified nodal metastases by the size and number of lymph node metastases in a similar fashion to mSCC. The most recent alternative staging system, called "N1S3" proposed by Forest et al [7] was published prior to the7th edition of the AJCC staging manual. It is relatively simple, allocating all single nodal metastases less than (or equal to) 3cm as nodal stage I, multiple nodes less than (or equal to) 3cm or single nodes greater than 3cm as nodal stage II and multiple nodes with at least one node greater than 3cm as stage III (Table 1).

Table 1.

N1S3 Staging system of nodal metastases for cutaneous squamous cell carcinoma

I	Single node measuring ≤ 3 cm in diameter
II	Single node measuring > 3 cm or multiple nodes ≤ 3 cm
III	Multiple nodes measuring > 3 cm in diameter

In response to increasing literature suggesting the need for a more complex staging system for cSCC, the AJCC revised the TNM staging of cSCC in the 7th edition of its staging manual[11] by adopting the same nodal (N) staging as that used for mSCC (Table 2). The advantages of using an established staging system are obvious, in terms of simplicity and ready acceptance due to familiarity[12]. However, the new staging system has not been evaluated previously. Its validity is questionable since

some components such as the laterality of lymph nodes (N2c) are not supported by any existing studies.

Table 2.

Nodal Staging of Mucosal and Cutaneous Squamous Cell Carcinoma from the 7th Edition of AJCC TNM Staging Manual

N1	Single ipsilateral node ≤ 3cm in greatest dimension
N2a	Single ipsilateral node > 3cm, ≤ 6 cm in greatest dimension.
N2b	Multiple ipsilateral nodes ≤ 6 cm in greatest dimension
N2c	Bilateral or contralateral nodes, ≤ 6 cm in greatest dimension
N3	Any node > 6cm in greatest dimension

There is minimal literature regarding what constitutes a good staging system for cancer and what criteria should be used to determine if one staging system is superior to another. Staging systems such as the AJCC TNM system have evolved over time to become a complex mix of anatomic disease extent, tumour grade and other prognostic factors. In many cancers, disease extent accounts for only a modest proportion of the variation in survival observed. In some tumours, disease extent has very limited prognostic value. For example, prognosis in papillary thyroid cancer is highly dependent on patient age and in patients under the age of 45 years, nodal metastases have minimal impact on survival and the presence of distant metastases is only considered to be stage II disease[13]. Despite this, within each T or N group an increasing number should correlate with worse prognosis rather than just depicting the anatomic extent of disease. The AJCC staging manuals state that the goal of cancer

staging is to group cancer characteristics for which patient survival differs between groups (distinctiveness), consistently decreases with increasing stage group (monotonicity), and is similar within a group (homogeneity) [14-16]. This is intended to be applied to the overall stage (I – IV) but the concepts of distinctiveness and monotonicity should apply within the T and N stages also. Distinctive groups should be clinically useful by avoiding stages that are rarely or too frequently applied and the monotonic decrease in survival should be sufficiently different as to be clinically relevant. Furthermore staging systems should not be unnecessarily complex and therefore the concept of parsimony should also apply.

The primary aim of this study is to compare the 7th edition AJCC TNM staging of nodal metastases in patients with cSCC with that of the N1S3 staging system. The criteria used to compare these staging systems is their ability to stratify patients according to risk of death due to cSCC taking into consideration distinctiveness and monotonicity, appropriate distribution of patients and staging model performance.

Methods

After institutional ethics approval was obtained, data on patients with nodal metastases from cSCC of the head and neck was obtained from two Australian cancer centre prospective databases, the Sydney Head and Neck Cancer Institute (SHNCI) database and Westmead Head and Neck Cancer Centre (WHNCC) database. This is a retrospective analysis of the two combined datasets which includes 331 patients from WHNCC and 272 from the SHNCI, giving a total of 603 patients with sufficiently complete data treated with curative intent between 1980 and 2010. During this time, management of patients has not differed substantially with surgery remaining the

primary treatment modality. The routine use of adjuvant radiotherapy, however, has become more routine over this time-frame and the differing application of radiotherapy is potentially an important confounding factor. The role of chemotherapy remains unproven and has only been used in a negligible proportion of patients. All patients underwent surgery (parotidectomy and / or neck dissection) and median duration of follow-up of survivors was 2.5 years (range 0.1 – 17 years). Only variables considered important for analysis based on existing literature[10, 17-20] were obtained including maximal nodal diameter, number of involved lymph nodes, presence of extracapsular spread (ECS) of tumour in lymph nodes, location of involved lymph nodes, margin status of lymph nodes, administration of radiotherapy, dose of radiotherapy, gender, age and presence of major immunosuppression. The pathological number, size and location of lymph nodes were used to calculate the AJCC nodal (N) stage and N1S3 stage. A simplified N stage was also devised where all N2 patients were combined. For simplicity, Nx (eg. N2b) will refer to the AJCC N stage and N1S3-I, II, III will refer to the N1S3 stage. These clinicopathological variables are summarised in Table 3 according to Hospital.

Statistical Analysis:

Data was collated and filtered using Excel (Microsoft, USA) and SPSS version 17.0 (IBM, USA). The data was then imported and merged using Stata version 11.1. The end point for analysis was disease-specific survival and was calculated from the date of surgery to date of death from cSCC or last follow-up. Patients who died from causes other than cSCC were censored at the time of death. Less than 10% of any variable contained missing data and most variables were complete. Missing values were imputed using other available data without statistical modeling where possible

(for example maximal lymph node diameter was based on pathology reports but clinical measures were used where pathological data was absent). In the case of missing categorical data the variable was assumed to be absent (for example extracapsular spread and immunosuppression). Differences in survival were determined using the general log-rank test, log-rank test for trend and univariable Cox proportional hazards model analysis. Failure rates were estimated for each staging system and compared. Preselected covariates (radiotherapy, ECS, nodal margin, age, hospital, radiotherapy dose and immunosuppression) were included in a multivariable Cox proportional hazards model where N1S3 stage and AJCC stage were included as categorical variables to avoid assumptions regarding linearity and monotonicity allowing individual estimates of effects of the levels of the two staging systems. Age and radiotherapy dose were removed from the model as they did not significantly contribute to the model, despite transformation. A significant interaction between hospital and radiotherapy was identified and an interaction term was included. Due to dependence amongst subjects within hospitals, the model was adjusted by clustering by institutions. To assess stage system performance and monotonicity, the N1S3 and AJCC stages were included as continuous variables. Proportion of explained variation (PVE) (R^2 and R_D^2) were calculated using Stata ado-files developed by Royston[21]. Survival curves were generated using the Kaplan-Meier method and also based on hazard estimates from the multivariable Cox regression models. Lift curves were generated by ranking patients by their adjusted estimated risk of failure ($\exp(x_j\beta_x)$) and plotting this against the proportion of patients who died from cSCC. Weighted lift curves were generated by ranking patients and plotting this against the proportion of deaths weighted by follow-up time ($\sum_1^{92} 1/\text{time to death from cSCC}$).

Table 3.

Clinicopathological data of patients with nodal metastases from cutaneous squamous cell carcinoma of the head and neck according to institution

Variable	**SHNCI**	**WHNCC**	**Total**
N	272	331	603
Age mean (SD)	72.4 (10.83) years	68.0 (12.41) years	70.0 (11.92) years
Male:Female ratio	240 : 32	275 : 56	525 : 88
(% male)	(88.2%)	(83.1%)	(85.4%)
Duration of follow up	1.7 years	3.8 years	2.5 years
median (range)	(0.1 – 13.7)	(0.1 – 17.6)	(0.1 – 17.6)
Involved lymph nodes	1 node	1 node	1 node
median (range)	(1 – 67)	(1 - 29)	(1 – 67)
Largest metastatic node	25 mm	23 mm	25 mm
median (range)	(5 – 100)	(3 – 92)	(3 – 100)
Involved margin n (%)	71 (26.1%)	191 (57.7%)	262 (43.5%)
ECS n (%)	138 (50.7%)	265 (80.1%)	403 (66.8%)
RT n (%)	159 (58.7%)	296 (89.4%)	455 (75.6%)
RT dose median (range)	54 (0 – 66) Gy	60 (0 – 74) Gy	60 (0 -74) Gy
Immunosuppression n (%)	4 (1.5%)	22 (6.7%)	26 (4.3%)
AJCC N stage n (%)			
▪ **N1**	121 (44.5%)	133 (40.2%)	254 (42.1%)
▪ **N2a**	25 (9.2%)	33 (10.0%)	58 (9.6%)
▪ **N2b**	97 (35.7%)	145 (43.8%)	242 (40.1%)
▪ **N2c**	2 (0.7%)	10 (3.0%)	12 (2.0%)
▪ **N3**	27 (9.9%)	10 (3.0%)	37 (6.1%)

Variable	SHNCI	WHNCC	Total
N1S3 stage n (%)			
▪ **I**	121 (44.5%)	135 (40.8%)	256 (42.5%)
▪ **II**	123 (45.2%)	137 (41.4%)	260 (43.12%)
▪ **III**	28 (10.3%)	59 (17.8%)	87 (14.4%)
Death from cSCC n	40	52	92
Total Deaths n	63	117	180

SHNCI – Sydney Head and Neck Cancer Institute

WHNCC – Westmead Head and Neck Cancer Centre

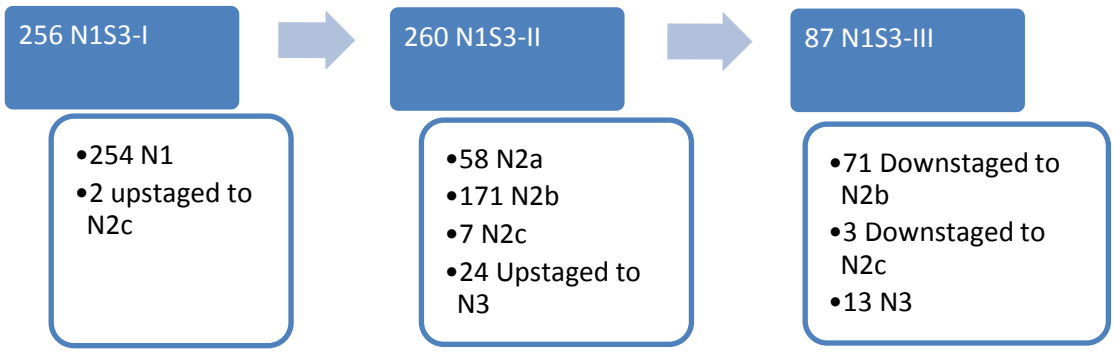
RT - Radiotherapy

ECS – Extracapsular spread

cSCC – cutaneous squamous cell carcinoma

Figure 1

Stage Shift from N1S3 to AJCC N Stage



Legend

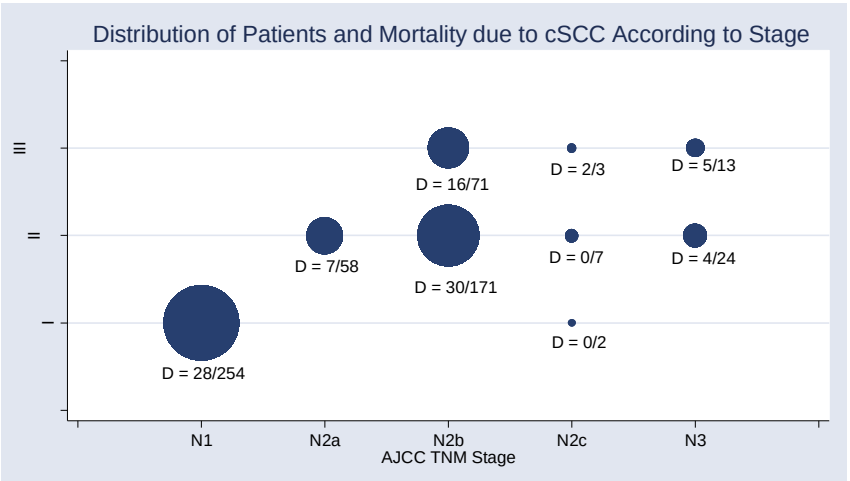
Shift in Stage when applying the AJCC N stage to patients previously staged according to the N1S3 staging system.

Results

Distribution by stage

The distribution of patients by stage is shown in Table 3 and the shift in stage from N1S3 to AJCC is shown in Figure 1. Three groups from the AJCC staging system (N2a, N2c and N3) contained less than 10% of all patients implying that that these groups would be used infrequently. In particular the N2c group, which denotes contralateral or bilateral nodal metastases, contained only 12 patients (2%). In contrast the smallest N1S3 group, N1S3-III, contained 87 patients (14%). N1 and N1S3-I have very similar criteria, with the only difference being that the AJCC system requires the solitary involved lymph node to be ipsilateral to the primary tumour. Two patients were both N2c and N1S3-I. Figure 2 shows the number of deaths due to cSCC as a proportion of patients in each AJCC-N1S3 subgroup. Within each AJCC group, an increasing proportion of patients died with increasing N1S3 sub-group. In particular, within N2c, only N1S3-III patients died. The converse was not observed within each N1S3 group. However, the proportion of N1S3-II/N2a patients who died was similar to N1S3-I patients. Despite any differences, the two staging systems were highly correlated (Spearman's rho = 0.88, p< 0.0001; Kendall's tau-b = 0.81 p < 0.0001).

Figure 2



Legend.

Scatterplot of patients with metastatic cutaneous squamous cell carcinoma (cSCC) according to N1S3 stage versus AJCC N stage. The size of the balloons corresponds to the number of patients in each subgroup. D is the number of patients who died due to cSCC over the total number of patients in the subgroup.

Table 4.

Estimated two- and five-year disease specific survival and 95% confidence intervals according to stage

Stage	2 Year DSS	95% CI	5 Year DSS	95% CI
N1S3				
I	91%	(85.7 – 94.1)	83%	(75.1 – 88.0)
II	85%	(79.3 – 89.5)	78%	(70.7 – 83.5)
III	75%	(62.7 – 83.3)	63%	(48.6 – 74.3)
AJCC N Stage				
N1	91%	(85.6 – 94.1)	83%	(75.0 – 88.0)
N2a	86%	(70.1 – 94.1)	79%	(60.8 – 89.6)
N2b	82%	(75.9 – 86.9)	74%	(66.2 – 80.2)
N2c	81%	(42.4 – 94.9)	81%	(42.4 – 94.9)
N3	80%	(59.9 – 90.4)	65%	(42.1 – 80.3)

Stratification of risk by stage

Kaplan-Meier disease specific survival curves generated according to staging system are shown in figures 3a – 3c with Kaplan-Meier estimates of two- and five-year disease specific survival are shown in Table 4. The curves demonstrate good stratification of survival according to N1S3 stage and there is strong evidence for a

difference in survivor functions across the groups (log-rank test $\chi^2(2)$ =12.64, p = 0.0018, log-rank test for trend $\chi^2(1)$ = 11.89, p = 0.0006, deviation from linear trend $D\chi^2(1)$ =0.75, p = 0.61). The AJCC N2a-c groups did not stratify patients well, with the N2a and N2c curves overlapping with N1 and N2b. The difference across AJCC groups was not as strong (log-rank test $\chi^2(4)$ =9.95, p = 0.041, log-rank test for trend $\chi^2(1)$ = 9.44, p = 0.002) however there was no statistical deviation from linearity (deviation from linear trend $D\chi^2(2)$ =0.51, p = 0.23). When N2 patients were combined, the condensed N stage appeared to stratify patients in a similar fashion to the N1S3 staging system though due to the small number of events in the N3 group, the evidence for a difference across the groups was also not as strong (log-rank test $\chi^2(2)$ =8.88, p = 0.012). After adjusting for the effect of immunosuppression, extracapsular spread, nodal margins, treating institution and radiotherapy using a Cox regression (no clustering) the N1S3 stage overall was significant as a categorical variable (W(2) =6.34, p = 0.04), however AJCC stage was not (W(4) =5.74, p = 0.22).

Figure 3a

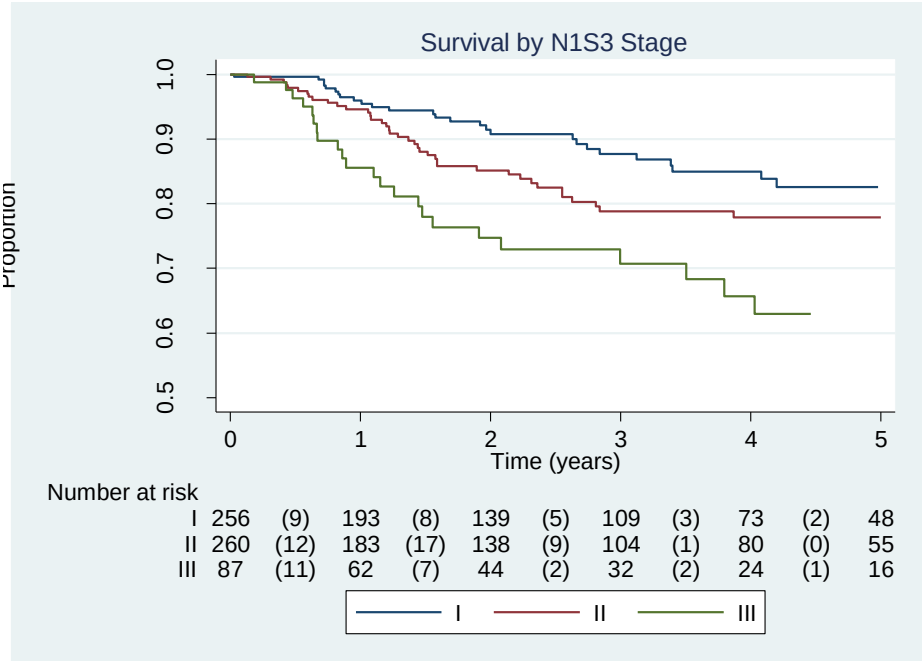


Figure 3b

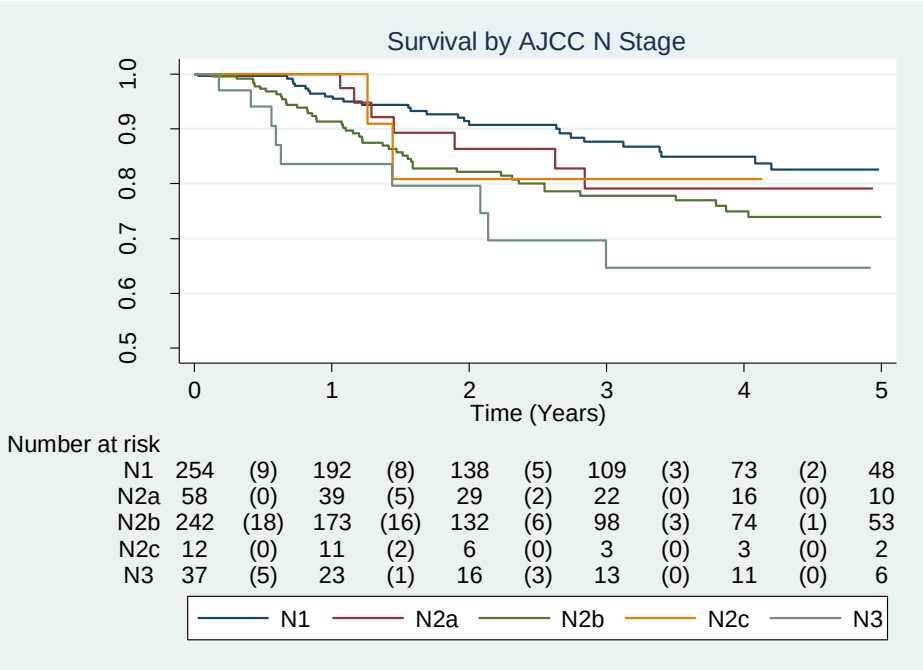
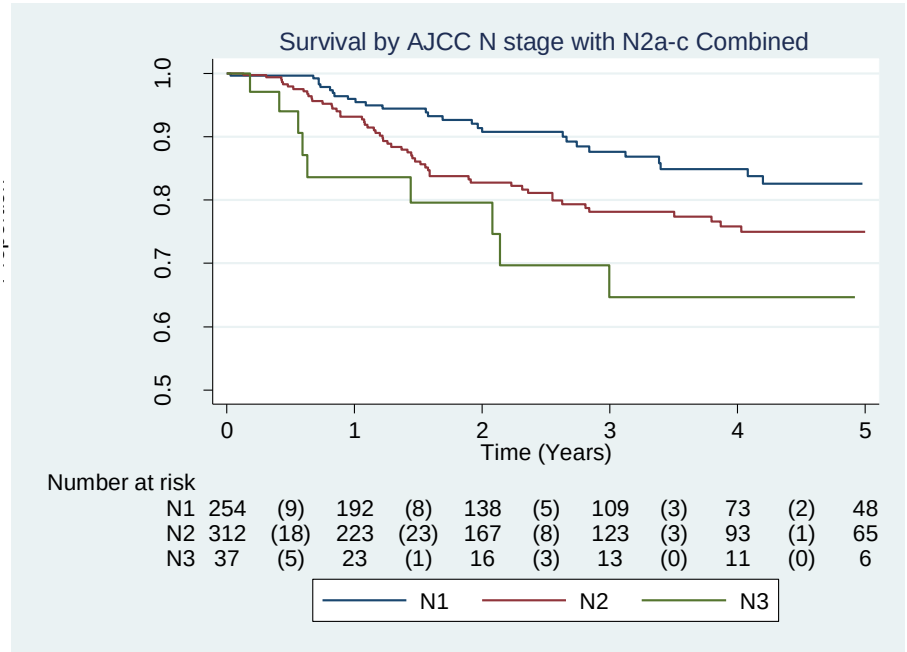


Figure 3c

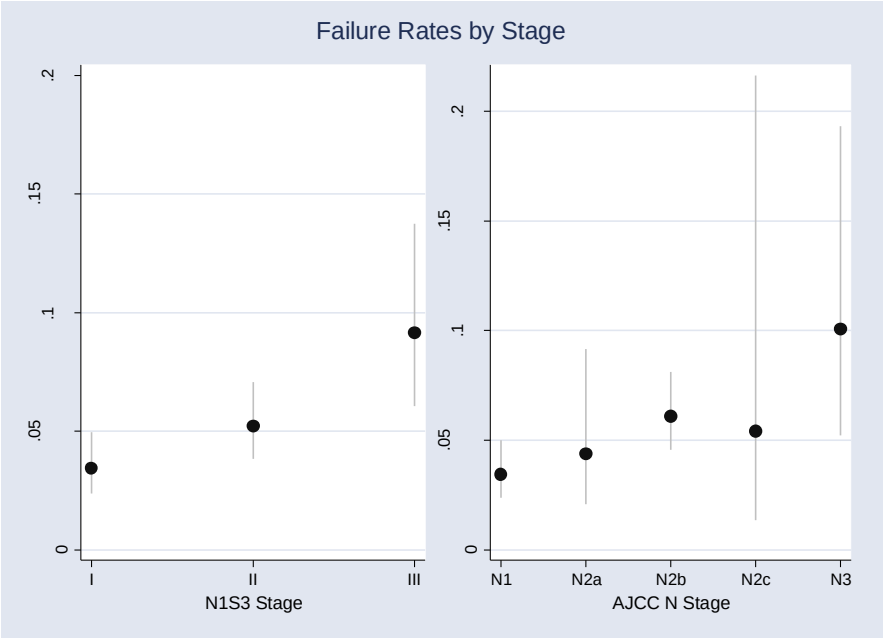


Legend

Kaplan-Meier disease specific survival curves according to a) N1S3 stage, b) AJCC stage and c) AJCC stage with N2a-c combined. Numbers at risk at each time period are provided with the number of events in parentheses. Note the scale of the curves has been changed to aid identification of the different groups (lower limit 0.5).

Estimated failure rates for each staging system are shown in Figure 4. The point estimates of failure for N1S3 stage rise at each level without confidence intervals overlapping with the point estimate. In contrast, the N2c point estimate is less than N2b and the 95% confidence intervals for N2a, N2c and N3 are very broad due to the small number of events (death due to disease) in these groups. The estimated failure rates for N3 and N1S3-III are similar; however the confidence intervals for N3 are much broader. Comparisons between groups within staging systems were limited to two tests per staging system. There was weak evidence for a difference in survival between N1S3-I and N1S3-II ($\chi^2(1) = 3.30$, $p = 0.069$) but was stronger between N1S3-II and N1S3-III ($\chi^2(1) = 5.86$, $p = 0.015$). There was a difference in survival between N1 and N2 (combined N2a-c) ($\chi^2(1) = 8.77$, $p = 0.012$) but not between N2 (combined N2a-c) and N3 ($\chi^2(1) = 1.04$, $p = 0.31$). There was no difference between N2a, N2b or N2c.

Figure 4



Legend

Comparison of failure rates (death from cSCC per Person-Year) for each staging system. Point estimates and 95% confidence intervals are provided for each stage.

Unadjusted hazard ratios (HR) and adjusted HRs in the final multivariable model are shown in Table 5 where the staging systems were fitted as categorical variables, therefore making no assumption about order or monotonicity. The effect of adjusting for other covariates was to reduce the hazard for both staging systems since there is considerable correlation between node size (an element in both staging systems) and both ECS and positive node margins. The estimated HR for N1S3-II and N1S3-III was 1.4 and 2.1, respectively, indicating a clinically useful, monotonic and linear increase in risk. The estimated HR for N2a, N2b, N2c and N3 was 1.1, 1.6, 1.4 and 2.2 indicating that the increase in risk was neither clinically useful nor monotonic, in particular for N2a (similar to N1) and N2c (less than N2b) as illustrated in Figure 4. The increase in risk for N3 compared to N1S3-III was similar but included less patients in the highest risk group. The adjusted survival curves (Figure 5) generated from the regression models demonstrate poor discrimination between N1, N2a, N2b and N2c patients and also that N2c having improved survival compared to N2b.

Staging System Performance

The PVE (R_D^2) for the model incorporating N1S3 stage was 31.0%, which was similar to the model incorporating TNM stage at 30.6%. In both cases, the staging system alone explained only a small proportion of the total variation and although the point estimate for N1S3 (8.3%, 95% CI 1.58-18.50) was slightly higher than AJCC (6.7%, 95% CI 0.87 – 16.27) there was no statistically significant difference. The overall model fit for using Cox-Snell residuals and predictive power of the models as assed by Harrell's C (0.72 v 0.71) and Somer's D (0.42 v 0.41) was similar for both models as shown in Figure 6.

Table 5.

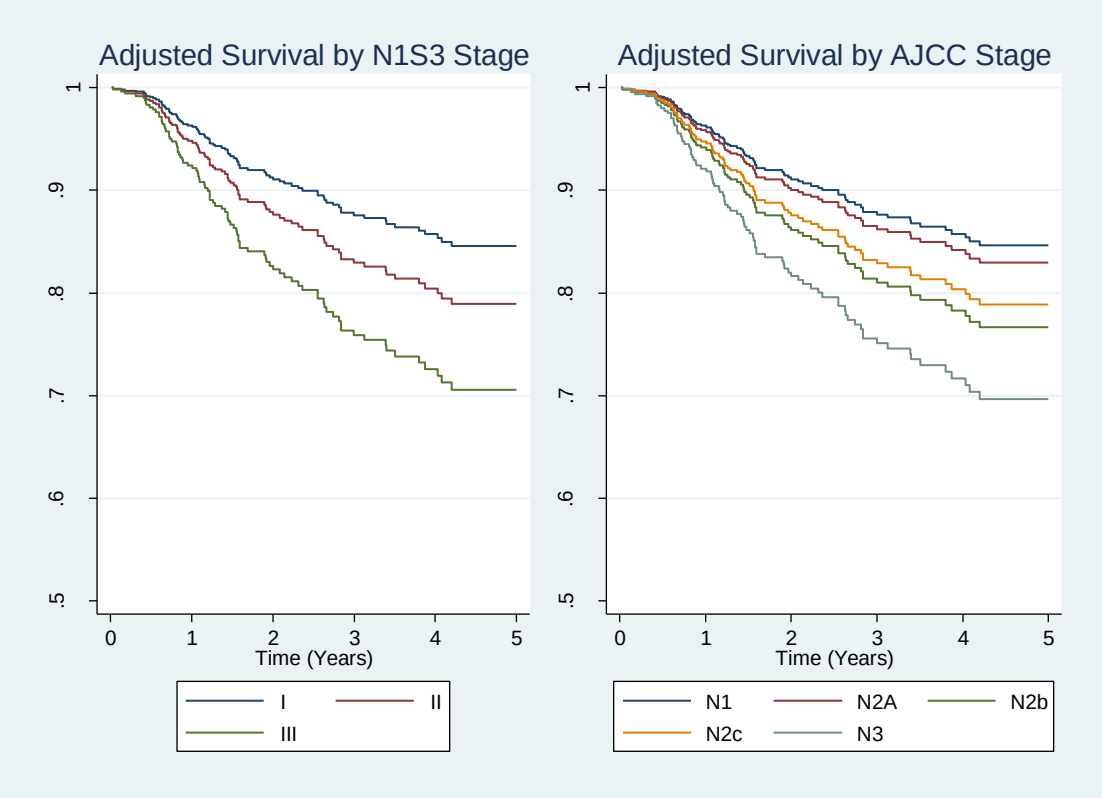
Raw and Adjusted Effect of N1S3 and AJCC Staging Systems

	Unadjusted			Adjusted†		
N1S3 Model	**HR**	**95% CI**	**p**	**HR**	**95% CI**	**p**
N1S3-II v I	1.5	0.95 - 2.49	0.078	**1.4**	1.21 - 1.65	<0.001
N1S3-III v I	2.6	1.53 - 4.60	0.001	**2.1**	1.97 - 2.19	<0.001
ECS				2.8	2.02 - 3.80	<0.001
Immunosuppression				3.3	3.30 - 3.38	<0.001
Involved Node Margin				2.0	1.73 - 2.34	<0.001
Radiotherapy				0.06	0.06 - 0.07	<0.001
Hospital				0.6	0.58 - 0.70	<0.001
Interaction RT*Hospital				5.4	5.21 - 5.62	<0.001
AJCC Model	**HR**	**95% CI**	**p**	**HR**	**95% CI**	**p**
N2a v N1	1.2	0.54 – 2.85	0.61	**1.1**	0.62 - 2.02	0.71
N2b v N1	1.8	1.13 – 2.90	0.013	**1.6**	0.98 - 2.57	<0.001
N2c v N1	1.5	0.36 – 6.40	0.56	**1.4**	0.44 - 4.58	0.56
N3 v N1	2.7	1.27 – 5.68	0.010	**2.2**	1.73 - 2.72	<0.001
ECS				2.9	2.21 - 3.88	<0.001
Immunosuppression				3.3	3.08 - 3.58	<0.001
Involved Node Margin				2.0	1.77 - 2.22	<0.001
Radiotherapy				0.06	0.05 - 0.07	<0.001
Hospital				0.6	0.51 - 0.69	<0.001
Interaction RT*Hospital				5.6	5.07 - 6.24	<0.001

† Cox proportional hazards model Standard Errors adjusted for 2 clusters in Hospital

HR – Hazard ratio. CI – confidence interval. RT - radiotherapy

Figure 5.



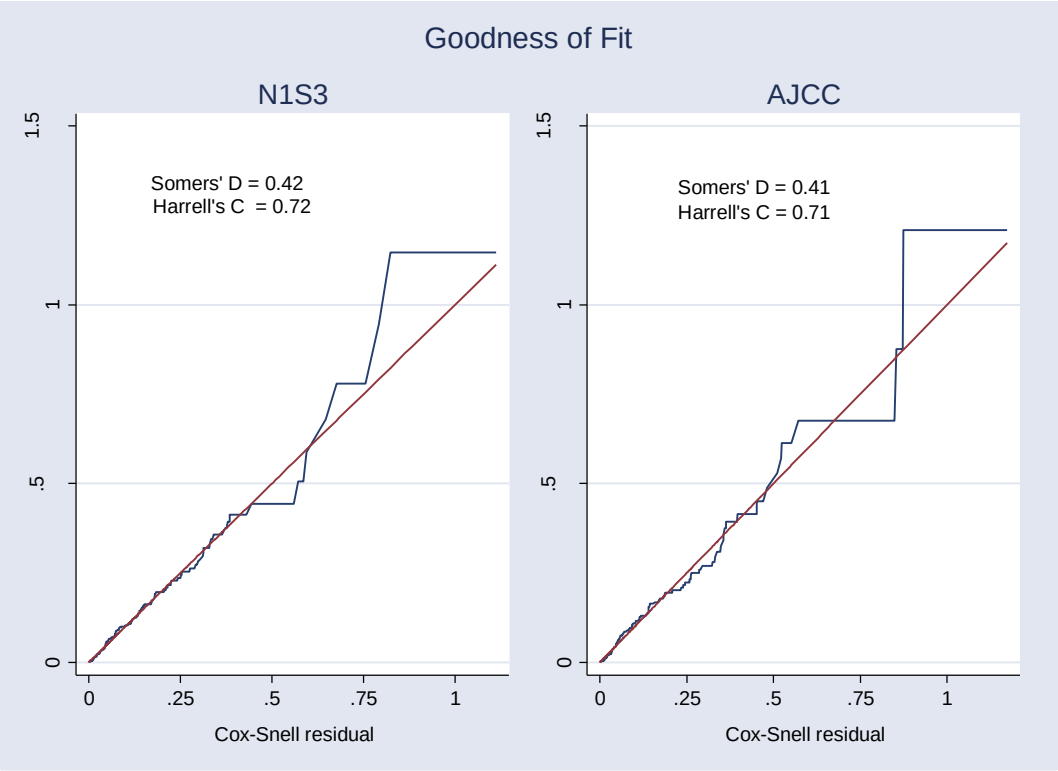
Legend

Predicted survival with other covariates (immunosuppression, involved node margin, radiotherapy, ECS and treating institution) fixed at their means using the N1S3 staging system on the left and the AJCC staging system on the right.

To assess the ability of the models to capture deaths due to cSCC, lift curves were generated by calculating the relative hazard ($\exp(x\beta_x)$) for each patient in the dataset based on the multivariable models, except AJCC and N1S3 stages were entered as continuous variables to force an assumption of increasing hazard with increasing stage. The relative hazard was ranked and plotted against the proportion of deaths (Figure 7a) and then weighted inversely by time to death (Figure 7b and 7c). All curves were significantly better than a random (uniform) distribution as indicated by the diagonal line where x% of deaths equals x% of patients (Kolmogorov-Smirnov

test $p < 0.001$). The lift curves demonstrate similar capacity for both staging systems to capture deaths due to disease according to estimated hazard, suggesting that the AJCC staging system does not offer any increase in performance over the N1S3 system despite being more complicated (Kolmogorov-Smirnov test $p = 0.51$ for unweighted data and $p = 0.65$ for weighted data). When ranked according to raw stage (without adjusting for other covariates) and inversely weighted by time from treatment to death, the N1S3 stage appeared to perform marginally better than the AJCC stage as shown in Figure 7c (Kolmogorov-Smirnov test $p = 0.06$).

Figure 6.



Legend:

Cox –Snell residuals (dashed lined) plotted against Nelson-Aelan cumulative hazard (solid line) for both staging systems. Note closer approximation of the two lines indicates better model fit. Harrell's C and Somers' D are measures of the ordinal predictive power of a model.

Figure 7a

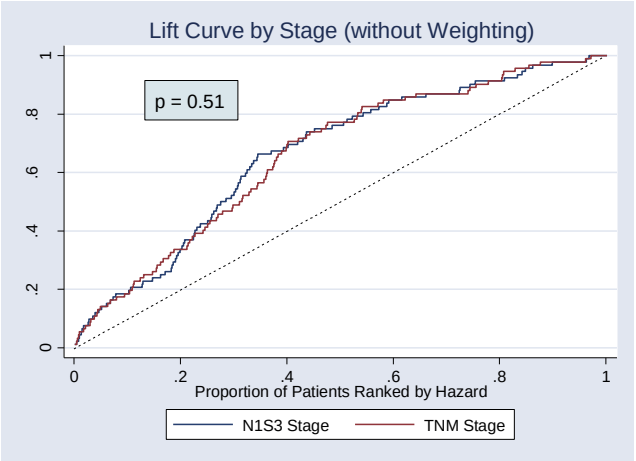


Figure 7b

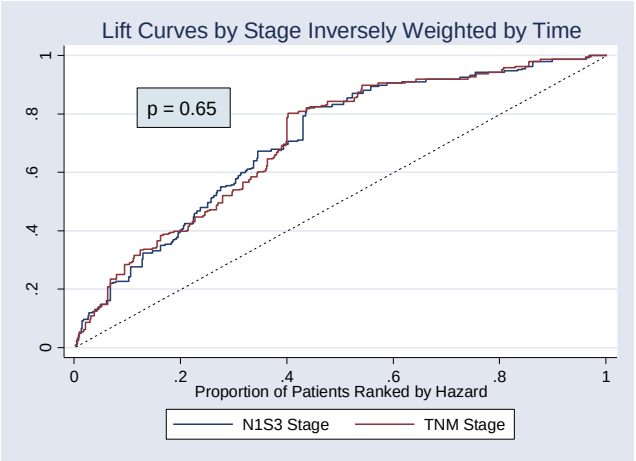
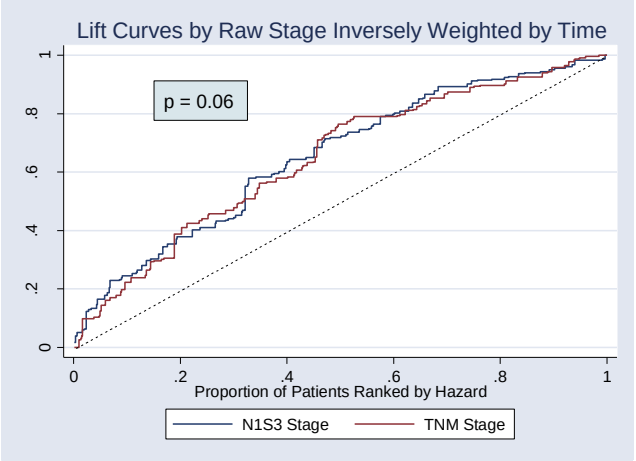


Figure 7c



Legend

Lift curves comparing N1S3 and AJCC TNM stage. Improved performance is indicated by greater area under the individual lift curves. No significant difference between curves was observed between staging systems.

Discussion

The present study incorporates data from two Australian cancer centres and represents the largest study of metastatic cSCC to date. Given that only 15% of patients died from metastatic cSCC in this study, large cohorts are required to generate more complex models that can adjust for other clinicopathological variables in an attempt to determine the independent effect of stage alone. Whilst overall survival could be used to provide more events, in a sample with a mean age of 70 years, many patients will die from unrelated causes introducing more variability which cannot be explained without incorporating comorbidity data and other predictors of non-cancer mortality. Disease specific survival is favoured as an outcome measure in less aggressive malignancies, such as cSCC for similar reasons[15].

The AJCC staging manual for cancer intends for primary tumour (T), nodal (N) and distant metastatic (M) data to be incorporated together to generate an overall stage I – IV and therefore assessing N stage in isolation is somewhat dubious. However, important differences between cSCC and other malignancies, such as mSCC, are that nodal metastases do not frequently present concurrently with the primary tumour, and in most Australian patients there are multiple potential primary tumours over long time periods thus many patients would need to be excluded if primary tumour factors were to be included in the analysis. Not only may it be impossible to determine the responsible primary tumour but it is unknown whether the primary factors are of any importance in patients with nodal metastases from cSCC. The metachronous nature of cSCC and our own (unpublished) data suggest that factors related to the nodal metastases are of principle importance.

To determine which staging system is more appropriate we have examined several criteria considered to be important for allocating stage to a malignancy. These include the distribution of patients by stage, stratification of patients by stage and performance of statistical models incorporating both the staging system alone and when combined with other potential confounding variables that may adjust the effect of stage. There is no uniform approach to how this should be undertaken, for example Brierley et al used a combination of the sum of observed deviations, mortality ratios and PVE to compare staging systems for thyroid cancer[22] where Wang et al used linear trend, likelihood ratio, and Akaike information criterion to comparing staging systems for gastric cancer[23]. Each approach has its own inherent limitations and it is difficult to know the whether any specific approach is superior to just looking at a Kaplan-Meier curve with confidence intervals. Whether raw or adjusted staging data should be used is contentious, therefore we have considered both. The adjusted effect is more useful in clinical practice where stage is rarely looked at in isolation, in particular adjusting for the effect of adjuvant treatment (not given to all patients) which alters the natural course of a disease is important.

Distribution

Although there is no stipulation that patients should be evenly distributed into staging groups, it is not beneficial to create groups that are rarely used unless they convey a unique clinical significance. Within the AJCC N stage there are three groups that apply to less than 10% of patients. This occurs mainly due to the separation of N2 into N2a, N2b and N2c paralleling the mSCC system, however in the case of cSCC the separation appears to create an irrelevant complexity. The strongest argument can be made against N2c which applies to only 2% of patients with metastatic cSCC in this

two-centre cohort and is estimated to have a better prognosis than N2b, though the confidence intervals are so large that the true effect could reasonably overlap with any group (42% – 95% two and five year disease specific survival). There does not appear to be any particular clinical significance attached to contralateral nodal metastases to warrant a distinct group. For example, a cSCC on the right midface with spread to the left submandibular lymph nodes would not be expected to carry a worse a prognosis than spread to the ipsilateral parotid nodes. The same could be argued for non-lateralised mucosal cancer, of course, however the unpredictable nature of cutaneous sentinel nodes provides pathophysiological evidence as to why contralateral nodal metastases are unlikely to carry the same clinical significance as in mucosal cancer [24]. The N3 group was also relatively small (6%) and whether, it is more appropriate to have a higher proportion of patients in the most adverse prognostic group is arguable and depends on how the staging is applied. If the most adverse group will be given more aggressive therapy, then one may argue that more patients will be exposed to either the benefit or toxicity. It is important to mention, however, that staging systems such as the AJCC and N1S3 were not primarily designed to determine what therapy should be administered but rather to provide a common language for disease processes and to predict prognosis in terms of recurrence and survival[25]. Given that prognosis in N3 and N1S3-III was similar and N1S3-III still only represents 14% of patients, it is reasonable to favour using the larger group.

Stratification

The N1S3 staging system stratifies well in terms of discriminating risk of death and creating a monotonic and linear increase in risk with the adjusted hazard ratio for N1S3-I, II and III being 1.0, 1.4 and 2.1, respectively. This is easier to achieve in a

three-level than a five-level staging system and many of the problems associated with the AJCC staging system may be related to the small number of subjects / events within each group making the estimates unreliable. Despite this, the estimated hazards do not support the use of the current five-level staging system, particularly after adjusting for the effect of confounding variables. In particular the hazard ratio for N2a (1.1 adjusted, 1.2 raw) is too similar to that of N1 to be clinically useful. The problems with N2c have already been discussed. The increase in risk of death for the AJCC stage is not monotonic and thus does not conform to this basic staging principle. This is partially overcome in the 7th edition of the AJCC staging system by grouping all N2 and N3 patients together as stage IV. However, this grouping seems inappropriate since N2 and N3 patients have an estimated 5 year disease specific survival 75% and 65%, respectively whereas no patient with distant metastases (M1) survived five years (0% 5 year DSS).

Performance

The evidence for a difference in survivor function for both the raw and adjusted N1S3 stage was considerably stronger than for the corresponding AJCC stage. However, both the N1S3 and AJCC regression models performed similarly in terms of capturing deaths due to cSCC as shown by the lift curves and other predictive measures. Schemper argues that if the prognostic importance of factors is to be compared, then PVE is the most appropriate measure[26]. It is useful to note that the estimated variation in survival time explained by N1S3 alone (R_D^2 8.3%) is similar to AJCC stage (R_D^2 6.7%) as the confidence intervals broadly overlap. Given that the AJCC stage is considerably more complex (with two extra sub-categories) than the N1S3

stage, without any gain in performance, the empiric evidence would support a more parsimonious staging system.

Limitations

The retrospective nature of this study increases the potential for error and bias, however as the data was collected prospectively for the relevant databases (i.e. not specifically for this study) with less than 10% of data being imputed, this is minimized. The benefit of a prospective study to assess staging of patients is considerably less than that for an interventional study because confounding variables are invariably not evenly distributed either between or within staging systems and therefore one still needs to adjust for treatment and pathological variables regardless of the way in which data is collected. Even though radiotherapy is included as a variable in the model, there is limited ability to adjust for its effect and correlated confounders. This is clearly demonstrated by the interaction term between hospital and radiotherapy, where radiotherapy given in one institution has the opposite effect in the other. This can be explained by policy differences between institutions, correlation with other adverse factors and also because the effect is not based on 'intention to treat' and therefore patients too sick to receive radiotherapy have an inflated adverse effect where the intention would have been to treat. This may be overcome by a prospective study. The long time-frame of data collection over 30 years also represents a problem due to changes in treatment philosophy and techniques. Whilst it is reasonable to assume that surgical techniques have not changed substantially, their application has become more standardised in terms of the extent of neck dissection. Radiotherapy techniques, on the other hand, have

undergone major changes in terms of dose and conformation. Furthermore, data supporting the use of adjuvant radiotherapy has led its routine application[10].

These staging systems only account for a small amount of the variability of patient outcomes (PVE < 10%) and even when other established variables are included, only one third of variability can be explained by the models. Therefore the importance of any particular prognostic variable could be argued to be minimal compared to the un-explained variation. As the N1S3 staging system was developed from SHNCI patients, there are potential problems with over-fitting of data in this study. However, only node number and size were directly modelled (rather than N1S3 stage directly) and only one-third of the current sample was used for this modelling. The N1S3 stage was applied separately and validated on patients where the initial sample were excluded in the study by Forest et al [7], and this was repeated using a subset of patients as part of the current analysis. However, due to the low rate of disease-related death, the entire dataset was required to provide sufficient power for comparison of the staging systems.

Recommendations

There is no measure by which the AJCC staging system functions better than N1S3 and in several categories it is worse despite being more complicated. Generally one would expect a more complex model to perform better than a simple model as it can use more variables to explain the variation observed. Whilst it is sensible and convenient to use an already established staging system, there is no evidence that cSCC should adopt the same staging as mSCC of the Head and Neck. In particular the N2a and N2c groups increase the complexity without any additional functionality. A

number of alternatives would seem reasonable. Firstly the N1S3 staging system could be adopted, secondly the N2a, N2b and N2c groups could be combined to one N2 group and lastly, based on the present data it would be practical to combine N2a with N1 and eliminate laterality of nodes (N2c). If the latter was done, this would result in the staging system summarised in Table 6 and would also require external validation. Kaplan-Meier disease specific survival curves according to this alternative TNM staging system is shown in Figure 8.

Conclusion

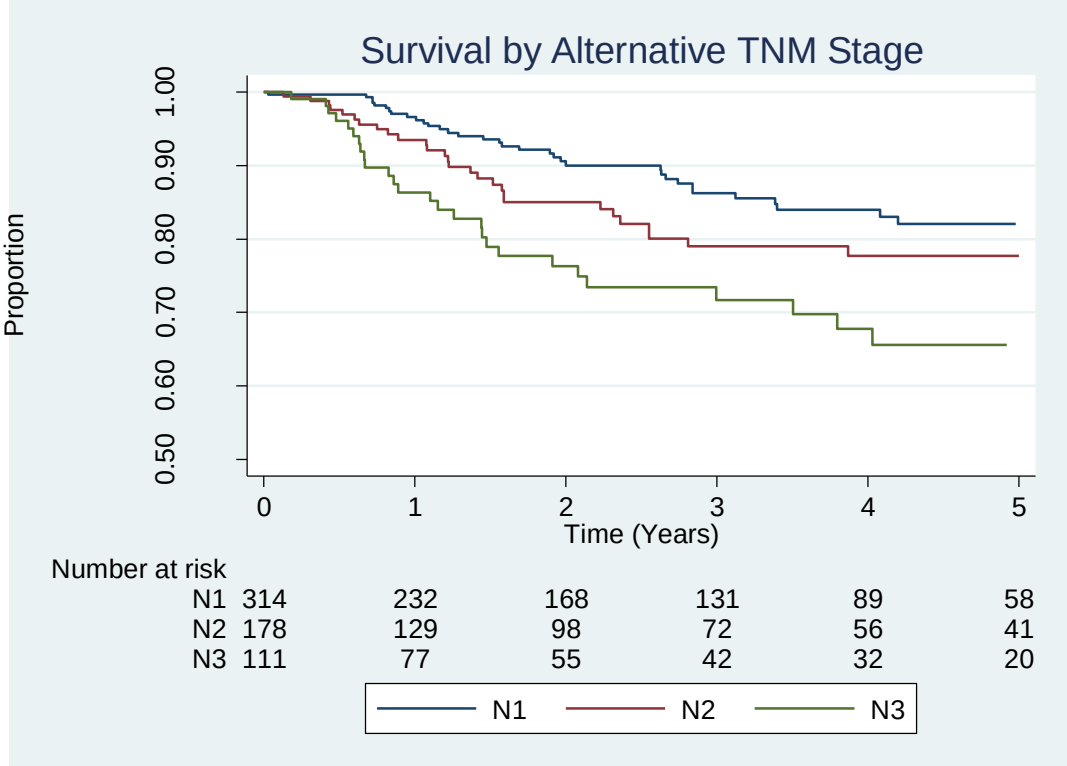
The 7th edition of the AJCC staging manual for cSCC is a major advance over the 6th edition, however the AJCC staging system does not stage patients as well as the N1S3 staging system despite being more complicated.

Table 6.

Alternative TNM staging based on current data

N1	Single node ≤ 6cm in greatest dimension
N2	Multiple nodes ≤3 cm in greatest dimension
N3	Multiple nodes > 3cm in greatest dimension or Any node > 6cm in greatest dimension

Figure 8



Legend

Kaplan-Meier disease specific survival curves according to the alternative TNM stage described in Table 6. Note the scale of the curves has been changed to aid identification of the different groups (proportion surviving lower limit 0.5).

References

1. Staples M, M.R., Giles G., *Trends in the incidence of non-melanocytic skin cancer treated in Australia 1985-1995: are primary prevention programs starting to have an effect?* Int J Cancer, 1998. **78**(2): p. 144-148.
2. Buettner PG, R.B., *Incidence rates of skin cancer in Townsville.* Australia. Int J Cancer, 1998. **78**: p. 587-593.
3. Veness MJ, P.S., Palme CE, et al. , *Cutaneous head and neck squamous cell carcinoma metastatic to parotid and cervical lymph nodes.* Head Neck, 2007. **29**: p. 621–631.
4. Brantsch KD, M.C., Schönfisch B, et al. , *Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study.* Lancet Oncol 2008. **9**: p. 713-720.
5. *Manual for Staging of Cancer*. 1977, American Joint Committee for Cancer Staging and End Results Reporting.
6. Andruchow JL, V.M., Morgan GJ, et al. , *Implications for clinical staging of metastatic cutaneous squamous carcinoma of the head and neck based on a multicenter study of treatment outcomes.* . Cancer, 2006. **106**(5): p. 1078-1083.
7. Forest VI, C.J., Veness MJ et al. , *N1S3: a revised staging system for head and neck squamous cell carcinoma with lymph node metastases.* . Cancer, 2010. **116**(5): p. 1298-1304.
8. O'Brien CJ, *The parotid gland as a metastatic basin for cutaneous cancer.* Arch Otolaryngol Head Neck Surg. , 2005. **131**: p. 551-555.

9. O'Brien CJ, M.E., McMahon JD, et al. , *Significance of clinical stage, extent of surgery, and pathological findings in metastatic cutaneous squamous cell carcinoma of the parotid gland.* . Head Neck., 2002. **24**: p. 417-422.

10. Oddone N, M.G., Palme CE, et al. , *Metastatic cutaneous squamous cell carcinoma of the head and neck: the Immunosuppression, Treatment, Extranodal spread, and Margin status (ITEM) prognostic score to predict outcome and the need to improve survival.* Cancer, 2009. **115**(9): p. 1883-1891.

11. *American Joint Committee on Cancer. AJCC cancer staging handbook. 7th edition*. 2010, Springer: New York.

12. Lardaro T, S.S.e.a., *Improvements in the Staging of Cutaneous Squamous Cell Carcinoma in the 7th Edition of the AJCC Cancer Staging Manual.* Annals of Surgical Oncology, 2010. **17**: p. 1979-1980.

13. Tuttle RM, B.D., Byrd D et al, *Thyroid Carcinoma* Natl Compr Canc Netw, 2010(8): p. 1228-1274.

14. Ishwaran, H., et al., *A novel approach to cancer staging: application to esophageal cancer.* Biostatistics, 2009. **10**(4): p. 603-20.

15. *American Joint Committee on Cancer. AJCC cancer staging handbook. 6th edition*. 2002, Springer: New York

16. Greene, F.L. and L.H. Sobin, *The staging of cancer: a retrospective and prospective appraisal.* CA Cancer J Clin, 2008. **58**(3): p. 180-90.

17. Clayman GL, L.J., Holsinger FC, et al. , *Mortality risk from squamous cell skin cancer.* . J Clin Oncol 2005, 2005. **23**(759-765).

18. Palme, C.E., et al., *The need for a better prognostic staging system in patients with metastatic cutaneous squamous cell carcinoma of the head and neck.* Curr Opin Otolaryngol Head Neck Surg, 2007. **15**(2): p. 103-6.

19. Veness, M.J., *High-risk cutaneous squamous cell carcinoma of the head and neck.* J Biomed Biotechnol, 2007. **2007**(3): p. 80572.

20. Veness, M.J., *Defining patients with high-risk cutaneous squamous cell carcinoma.* Australas J Dermatol, 2006. **47**(1): p. 28-33.

21. Royston, *Explained variation for survival models.* The Stata Journal, 2006. **6**(1): p. 83-96.

22. Brierley, J.D., et al., *A comparison of different staging systems predictability of patient outcome. Thyroid carcinoma as an example.* Cancer, 1997. **79**(12): p. 2414-23.

23. Wang, W., et al., *Comparison of the 6th and 7th editions of the UICC TNM staging system for gastric cancer: results of a Chinese single-institution study of 1,503 patients.* Ann Surg Oncol, 2011. **18**(4): p. 1060-7.

24. de Wilt, J.H., et al., *Correlation between preoperative lymphoscintigraphy and metastatic nodal disease sites in 362 patients with cutaneous melanomas of the head and neck.* Ann Surg, 2004. **239**(4): p. 544-52.

25. Kassis, E.S., et al., *Application of the revised lung cancer staging system (IASLC Staging Project) to a cancer center population.* J Thorac Cardiovasc Surg, 2009. **138**(2): p. 412-418 e1-2.

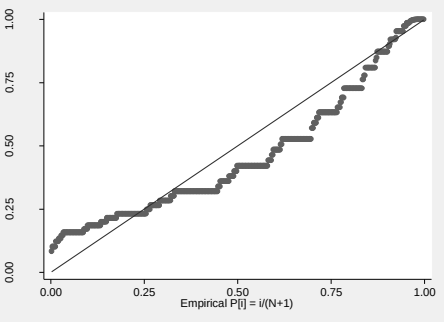
26. Schemper, M., *The relative importance of prognostic factors in studies of survival.* Stat Med, 1993. **12**(24): p. 2377-82.

27. Hosmer D.W., L.S., May S., *Applied Survival Analysis*. Second ed. 2008, New Jersey: John Wiley & Sons.

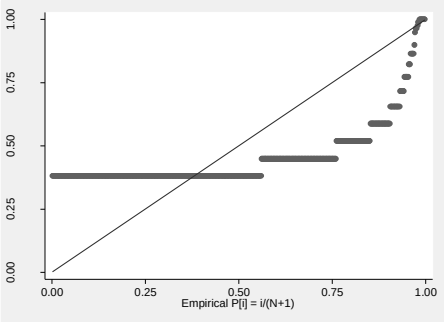
Statistical Appendix

Normal Probability Plots of Continuous Variables

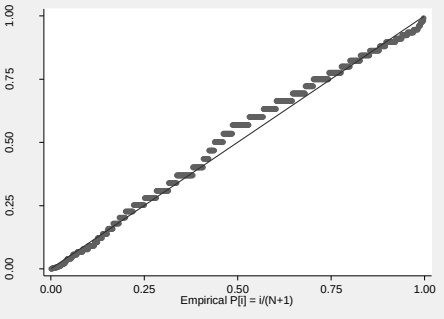
Largest metastasis



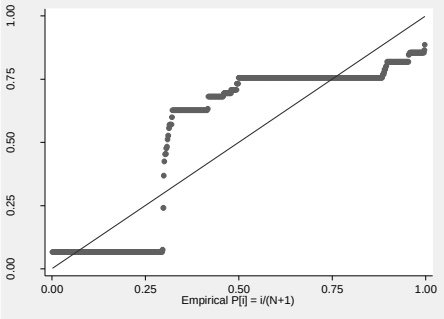
Number of nodes



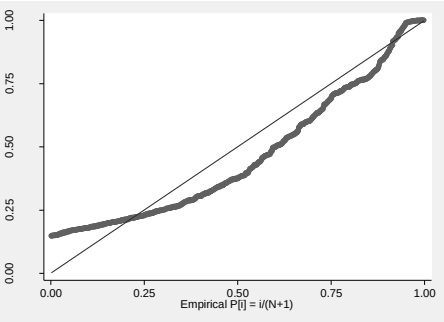
Age



Radiotherapy dose

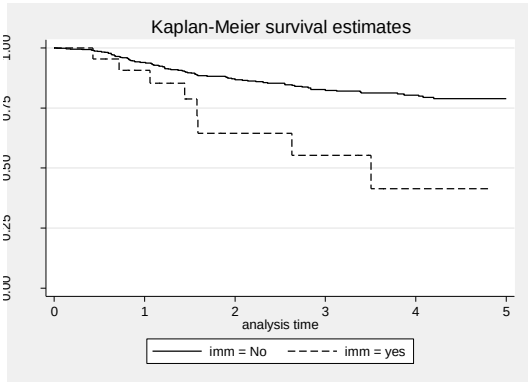


Duration FU



Univariable survival comparisons using Kaplan-Meier Curves and Log-rank test

Immunosuppression

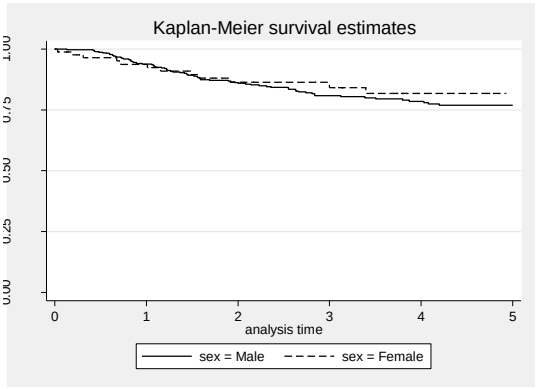


Log-rank test for equality of survivor functions

imm	Events observed	Events expected
No	84	89.15
yes	8	2.85
Total	92	92.00

chi2(1) = 9.64
Pr>chi2 = 0.0019

Gender

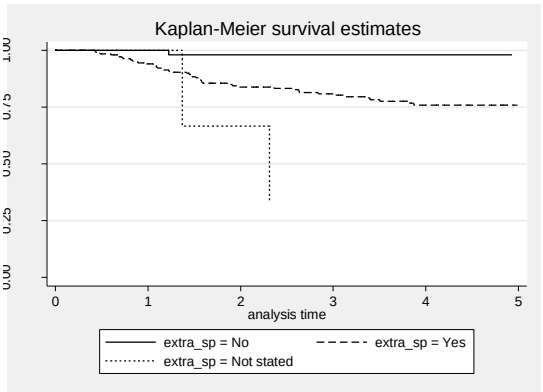


Log-rank test for equality of survivor functions

sex	Events observed	Events expected
Male	80	78.04
Female	12	13.96
Total	92	92.00

chi2(1) = 0.33
Pr>chi2 = 0.5679

Extracapsular spread



Log-rank test for equality of survivor functions

extra_sp	Events observed	Events expected
No	1	10.94
Yes	49	40.55
Not stated	2	0.51
Total	52	52.00

chi2(2) = 15.17
Pr>chi2 = 0.0005

Radiotherapy Dose

Test for trend of survivor functions

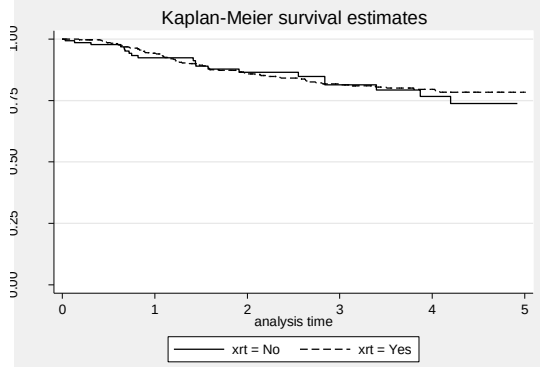
chi2(1) = 1.00
Pr>chi2 = 0.3174

Age

Test for trend of survivor functions

chi2(1) = 1.67
Pr>chi2 = 0.1969

Radiotherapy

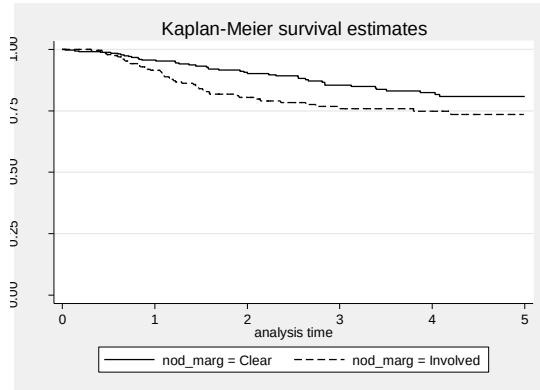


Log-rank test for equality of survivor functions

xrt	Events observed	Events expected
No	20	18.86
Yes	72	73.14
Total	92	92.00

chi2(1) = 0.09
Pr>chi2 = 0.7676

Node Margin

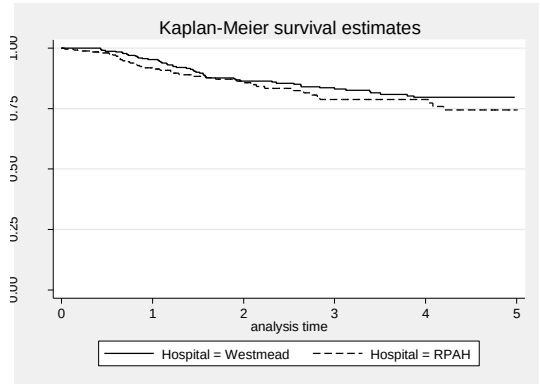


Log-rank test for equality of survivor functions

nod_marg	Events observed	Events expected
Clear	43	55.17
Involved	49	36.83
Total	92	92.00

chi2(1) = 6.73
Pr>chi2 = 0.0095

Hospital



Log-rank test for equality of survivor functions

Hospital	Events observed	Events expected
Westmead	52	57.45
RPAH	40	34.55
Total	92	92.00

chi2(1) = 1.39
Pr>chi2 = 0.2389

Selection of covariates and model construction

The model was constructed as per that described by Hosmer, Lemeshow and May[27]. As there were 92 deaths due to cSCC it was reasonable to include up to 9 covariates. From previous literature it was felt that in addition to stage, radiotherapy and extracapsular spread (ECS) needed to be included in the model. Based on univariable comparisons immunosuppression and involved nodal margin were included. Age was considered, although it was felt that this was less important for disease specific rather than overall survival. It was also felt that there were likely to be institutional differences that needed to be considered and hence treating hospital would need to be included in the model. Only N1S3 models will be demonstrated in this section of the appendix.

The initial model constructed (shown below) surprisingly showed that the treating institution was significant but radiotherapy was not. Age was not significant, however before removing age from the model the linearity assumption was assessed using martingale residuals and age was found to be approximately linear. Age was also dichotomised into greater and less than 70 years but was not significant. It was postulated that the reason why radiotherapy was not significant may be due to differing radiotherapy doses delivered.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.235763	1.266913	3.00	0.003	1.50211	6.970303
xrt	1.00762	.279741	0.03	0.978	.5847648	1.736249
Hospital	2.213671	.5714403	3.08	0.002	1.334697	3.6715
_IN1S3_2	1.415428	.353189	1.39	0.164	.8679388	2.308269
_IN1S3_3	2.117652	.6183699	2.57	0.010	1.194807	3.753287
ECS	2.57004	.7989712	3.04	0.002	1.397395	4.726726
nod_marg	1.738441	.3929111	2.45	0.014	1.116289	2.707341
age	1.011313	.0093548	1.22	0.224	.9931429	1.029815

The next model removed age and included radiotherapy dose. Radiotherapy dose was not linear and multiple transformations were performed with the best approximation

being with a radiotherapy dose squared transformation shown below. The smoothed martingale residual plots are shown in the model diagnostics section.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.057938	1.190459	2.87	0.004	1.425787	6.55847
xrt	1.17696	.4704902	0.41	0.684	.5376421	2.576498
Hospital	2.162151	.6119024	2.72	0.006	1.241625	3.765142
_IN1S3_2	1.371341	.3404308	1.27	0.203	.8430197	2.230763
_IN1S3_3	2.029183	.5897605	2.43	0.015	1.147965	3.586856
ECS	2.616381	.8121525	3.10	0.002	1.423898	4.807541
nod_marg	1.821799	.4092554	2.67	0.008	1.172961	2.829551
maxRTsq	.9999303	.0001118	-0.62	0.533	.9997113	1.000149

The introduction of radiotherapy dose to the model did not contribute significantly to the model or alter the other variable coefficients so it was removed. The next step was introduction of any clinically reasonable interaction terms. Since several prior publications from Westmead hospital had demonstrated a clinically and statistically significant effect of radiotherapy on survival, the most plausible interaction would be between the treating institution and radiotherapy. The model shown below demonstrates that both radiotherapy and the interaction term were significant, however the treating institution was no longer significant. This clearly showed how incorrect conclusions can be reached without consideration of potential interactions. The appropriate interpretation is not that outcomes at one institution are superior to the other (as the model above would suggest), however differences associated with radiotherapy administration between hospitals altered survival. Whilst it is not possible to determine what these differences are with the existing data I think it is likely this represents differences in treatment philosophies where in one institution radiotherapy is almost universally administered and in the other a more risk-adjusted approach is used. This means that in one hospital only patients who refuse or are too sick to receive radiotherapy are spared (hence the beneficial effect is exaggerated) and in the other institution only patients with more adverse features are given radiotherapy

(hence the beneficial effect is under-estimated). The adjusted effect of radiotherapy now estimates a 66% reduction in risk of death from cSCC with the administration of radiotherapy. Radiotherapy dose was reintroduced to the model but again was not significant.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.292301	1.29014	3.04	0.002	1.527366	7.096696
_Ixrt_1	.3368084	.1266004	-2.90	0.004	.1612245	.7036144
_IHospital_2	.6374435	.2885656	-0.99	0.320	.2624881	1.54801
IxrtXHos~2	5.408739	2.822376	3.23	0.001	1.94502	15.0407
_IN1S3_2	1.411957	.3528379	1.38	0.167	.8651913	2.304255
_IN1S3_3	2.079099	.6048128	2.52	0.012	1.1756	3.676977
ECS	2.771659	.8625334	3.28	0.001	1.50608	5.100722
nod_marg	2.009846	.4558626	3.08	0.002	1.28854	3.13493

The unexpected interaction between hospital and radiotherapy raise the likelihood that there may be a random effect or dependence among patients (clustering) within a hospital. As a result shared-frailty, correlated-frailty and stratified models were considered as shown below. Whilst clustering models are usually applied to multivariate survival data with repeated events for individuals it can be applied whenever failures times are correlated.

Standard model

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.292301	1.29014	3.04	0.002	1.527366	7.096696
_Ixrt_1	.3368084	.1266004	-2.90	0.004	.1612245	.7036144
_IHospital_2	.6374435	.2885656	-0.99	0.320	.2624881	1.54801
IxrtXHos~2	5.408739	2.822376	3.23	0.001	1.94502	15.0407
_IN1S3_2	1.411957	.3528379	1.38	0.167	.8651913	2.304255
_IN1S3_3	2.079099	.6048128	2.52	0.012	1.1756	3.676977
ECS	2.771659	.8625334	3.28	0.001	1.50608	5.100722
nod_marg	2.009846	.4558626	3.08	0.002	1.28854	3.13493

Shared frailty model

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.292301	1.29014	3.04	0.002	1.527366	7.096696
_Ixrt_1	.3368084	.1266004	-2.90	0.004	.1612245	.7036144
_IHospital_2	.6374435	.2885656	-0.99	0.320	.2624881	1.54801
IxrtXHos~2	5.408739	2.822376	3.23	0.001	1.94502	15.0407
_IN1S3_2	1.411957	.3528379	1.38	0.167	.8651913	2.304255
_IN1S3_3	2.079099	.6048128	2.52	0.012	1.1756	3.676977
ECS	2.771659	.8625334	3.28	0.001	1.50608	5.100722
nod_marg	2.009846	.4558626	3.08	0.002	1.28854	3.13493
theta	2.11e-16	7.85e-13				

Likelihood-ratio test of theta=0: chibar2(01) = 1.5e-09 Prob>=chibar2 = 0.500

Stratified model

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.318109	1.307843	3.04	0.002	1.532452	7.184464
_Ixrt_1	.3407206	.128081	-2.86	0.004	.1630878	.7118283
_IHospital_2	(omitted)					
IxrtXHos~2	5.287743	2.76028	3.19	0.001	1.900774	14.70991
_IN1S3_2	1.408486	.3519329	1.37	0.170	.8631095	2.29847
_IN1S3_3	2.068144	.6021529	2.50	0.013	1.168822	3.659429
ECS	2.723583	.8471455	3.22	0.001	1.480411	5.010708
nod_marg	2.012249	.4569151	3.08	0.002	1.289443	3.14023

Stratified by Hospital

Correlated –frailty model

(Std. Err. adjusted for 2 clusters in Hospital)

_t	Haz. Ratio	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
imm	3.292301	.0438303	89.51	0.000	3.207507	3.379338
_Ixrt_1	.3368084	.0120704	-30.37	0.000	.3139625	.3613167
_IHospital_2	.6374435	.0298987	-9.60	0.000	.581456	.6988219
IxrtXHos~2	5.408739	.103622	88.11	0.000	5.20941	5.615696
_IN1S3_2	1.411957	.1134184	4.29	0.000	1.206276	1.652707
_IN1S3_3	2.079099	.0559756	27.19	0.000	1.972234	2.191756
ECS	2.771659	.4475095	6.31	0.000	2.019788	3.803418
nod_marg	2.009846	.1558786	9.00	0.000	1.726417	2.339806

Within all of the preliminary and final models the coefficients remain stable, particularly with respect to the N1S3 and AJCC staging systems indicating that the estimates are reliable with the data available. The clustered model had smaller standard errors than the unclustered models due to negative correlation of residuals. Whilst all of the models seem reasonable and do not alter the final results, the correlated frailty model was used to account for a likely dependence among patients within each institution. This is because patients are nested within hospitals and cannot be assumed to be independent.

Model Diagnostics

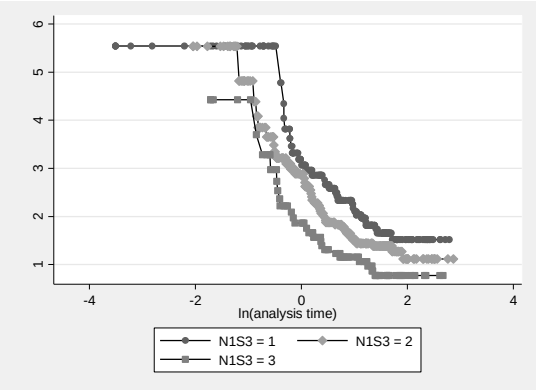
Test of proportional-hazards assumption

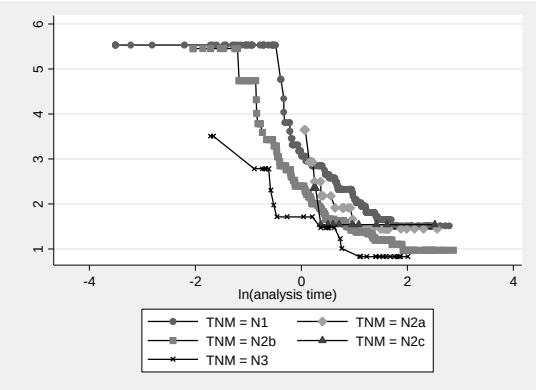
N1S3 Model

	rho	chi2	df	Prob>chi2
imm	0.09172	0.78	1	0.3783
_Ixrt_1	-0.09893	0.91	1	0.3396
_IHospital_2	-0.16098	2.31	1	0.1287
IxrtXHos~2	0.16177	2.49	1	0.1143
_IN1S3_2	-0.07582	0.55	1	0.4575
_IN1S3_3	-0.10308	0.95	1	0.3290
ECS	-0.02249	0.05	1	0.8285
nod_marg	-0.02848	0.07	1	0.7884
global test		5.06	8	0.7509

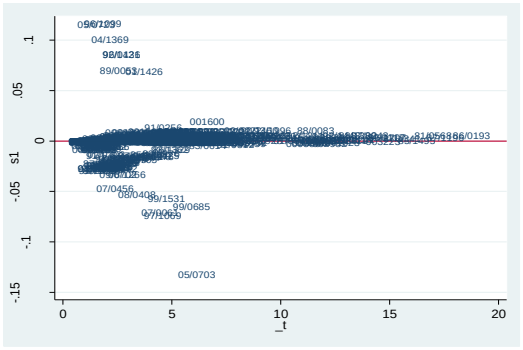
AJCC Model

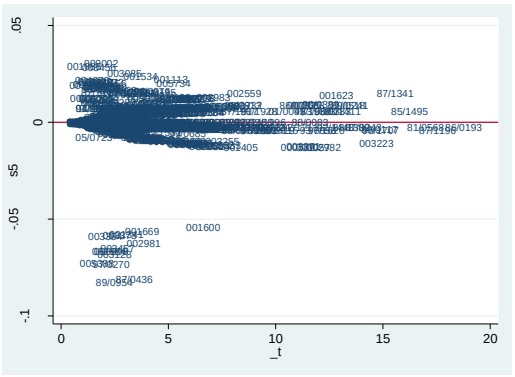
	rho	chi2	df	Prob>chi2
imm	0.09478	0.85	1	0.3573
_Ixrt_1	-0.09291	0.80	1	0.3723
_IHospital_2	-0.13985	1.81	1	0.1788
IxrtXHos~2	0.15097	2.24	1	0.1347
_ITNM_2	-0.02698	0.07	1	0.7933
_ITNM_3	-0.08842	0.75	1	0.3860
_ITNM_4	-0.03920	0.13	1	0.7162
_ITNM_5	-0.09188	0.77	1	0.3809
ECS	-0.02371	0.05	1	0.8183
nod_marg	-0.02992	0.08	1	0.7812
global test		5.08	10	0.8855

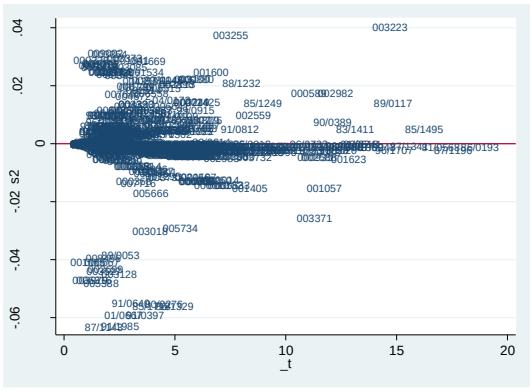


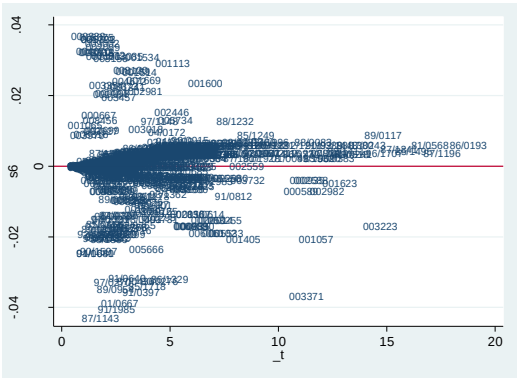


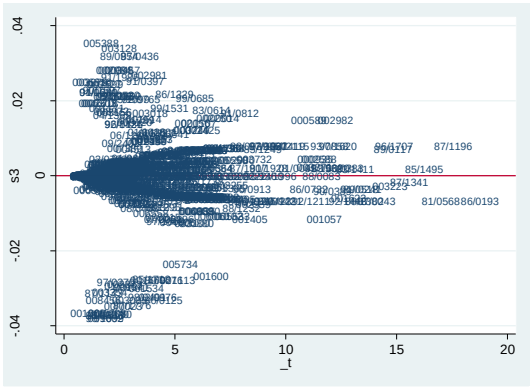
Identification of influential outliers using Schoenfeld residuals

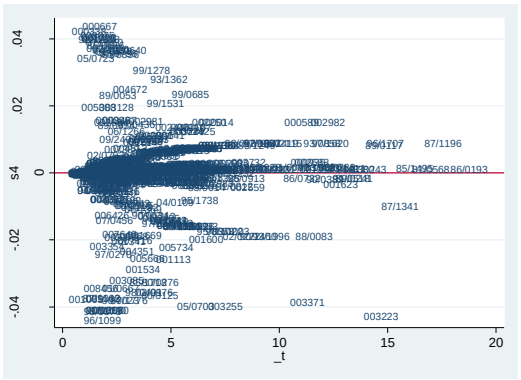








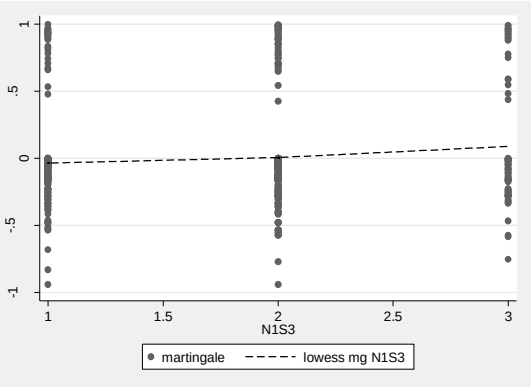




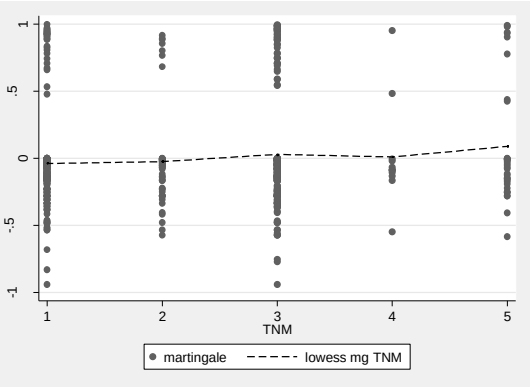
Outliers 05/0703 / 04/1369 / 05/0723 / 96/1099 / 89/0954 / 87/0436 / 97/0270 / 005388 / 003128 / 002984 / 000098 / 003457 / 002981 / 003354 / 003373 / 001341 / 001669 / 001600 were identified and their files extracted to confirm survival status and clinicopathological variables. All data was confirmed to be correct and no patients were excluded.

Assessment of linearity assumptions for staging variables using Martingale residuals

N1S3

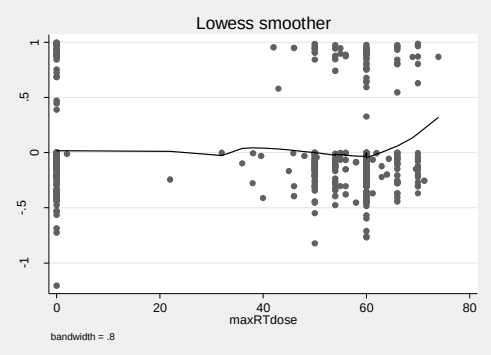


AJCC

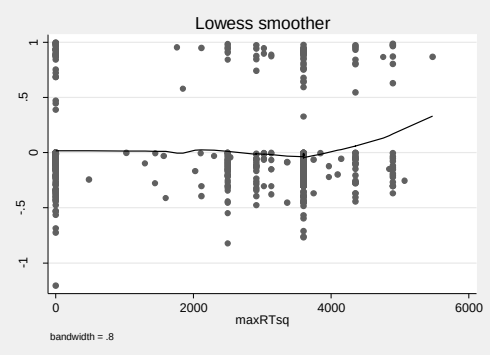


Radiotherapy dose (RTdose) and transformations

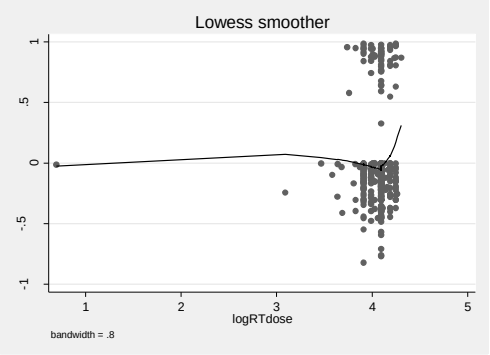
RT Dose (no transformation)



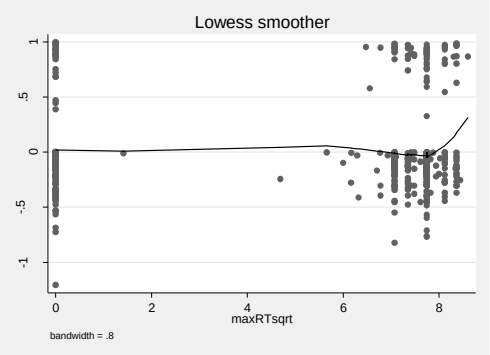
RT dose squared



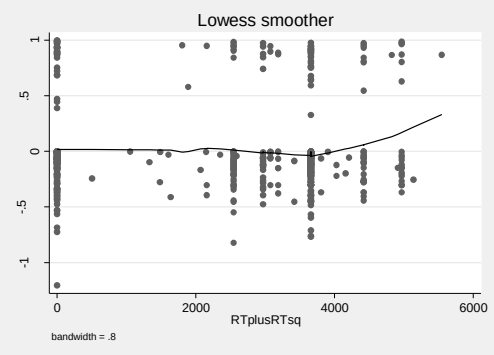
Log RT dose



Square root RT dose



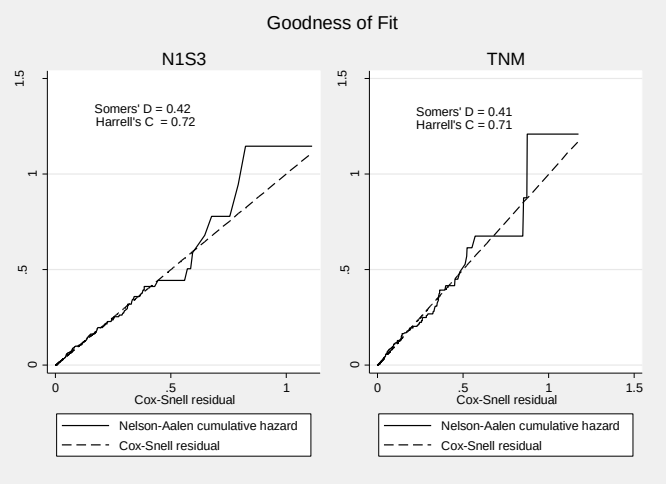
RT dose + RT dose squared



Assessment of Goodness of fit using Cox-Snell Residuals

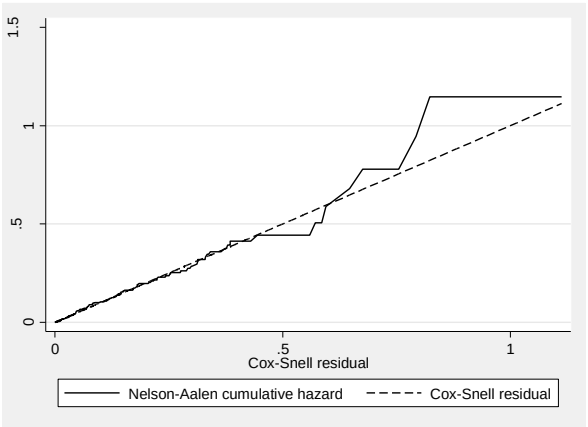
The unclustered models were first assessed

```
N1S3
Cox regression -- Efron method for ties
No. of subjects =          602                     Number of obs   =       602
No. of failures =           92
Time at risk    =  1854.983286
                                                    LR chi2(8)      =     50.95
Log likelihood  =    -514.0584                      Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         imm |   3.292261   1.290122     3.04   0.002     1.527348    7.096602
     _Ixrt_1 |   .3367946   .1265951    -2.90   0.004      .161218    .7035851
_IHospital_2 |   .6374646   .2885748    -0.99   0.320      .262497     1.54806
_IxrtXHos_~2 |   5.409894   2.823018     3.24   0.001     1.945407    15.04413
    _IN1S3_2 |   1.411954   .3528384     1.38   0.167     .8651887    2.304255
    _IN1S3_3 |   2.079538   .6049378     2.52   0.012     1.175851    3.677744
         ECS |   2.771764   .8625684     3.28   0.001     1.506134    5.100923
    nod_marg |   2.010123   .4559328     3.08   0.002     1.288708    3.135384
------------------------------------------------------------------------------
Harrell's C concordance statistic

  Number of subjects (N)              =      602
  Number of comparison pairs (P)      =    34773
  Number of orderings as expected (E) =    23964
  Number of tied predictions (T)      =     1360

          Harrell's C = (E + T/2) / P =    .7087
                        Somers' D =    .4174


AJCC
Cox regression -- Efron method for ties
No. of subjects =          602                     Number of obs   =       602
No. of failures =           92
Time at risk    =  1854.983286
                                                    LR chi2(10)     =     50.48
Log likelihood  =   -514.29409                      Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         imm |   3.320736    1.31152     3.04   0.002     1.531276     7.20137
     _Ixrt_1 |   .3384321   .1281478    -2.86   0.004      .161125    .7108537
_IHospital_2 |   .5898643   .2729426    -1.14   0.254     .2381679    1.460901
_IxrtXHos_~2 |   5.628694   2.962251     3.28   0.001     2.006504    15.78975
     _ITNM_2 |   1.120705   .4796412     0.27   0.790     .4843876    2.592925
     _ITNM_3 |   1.593284   .3915728     1.90   0.058     .9842335    2.579221
     _ITNM_4 |   1.420321   1.055898     0.47   0.637     .3308142    6.098018
     _ITNM_5 |   2.166823   .8614827     1.94   0.052     .9940395    4.723273
         ECS |   2.923738   .9016245     3.48   0.001     1.597513    5.350972
    nod_marg |   1.983819   .4537287     3.00   0.003     1.267126    3.105878
------------------------------------------------------------------------------
  Harrell's C concordance statistic

  Number of subjects (N)              =      602
  Number of comparison pairs (P)      =    34773
  Number of orderings as expected (E) =    23833
  Number of tied predictions (T)      =     1428

          Harrell's C = (E + T/2) / P =    .7059
                        Somers' D =    .4118
```



This was compared with the clustered models and the same results observed (only N1S3 model shown)

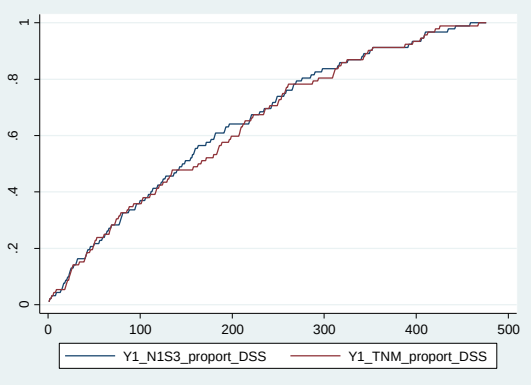
```
N1S3
Cox regression -- Efron method for ties
No. of subjects      =          602                 Number of obs    =       602
No. of failures      =           92
Time at risk         =  1854.983286
                                                    Wald chi2(1)     =     39.89
Log pseudolikelihood =    -514.0584                 Prob > chi2      =    0.0000

                              (Std. Err. adjusted for 2 clusters in Hospital)
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         imm |   3.292261   .0437341    89.70   0.000      3.20765    3.379104
     _Ixrt_1 |   .0622553   .0034459   -50.16   0.000     .0558548    .0693892
    Hospital |   .6374646   .0298748    -9.61   0.000     .5815198    .6987916
_IxrtXHosp~1 |   5.409894   .1051462    86.86   0.000     5.207687    5.619952
     _IN1S3_2 |   1.411954   .1134065     4.30   0.000     1.206294    1.652678
     _IN1S3_3 |   2.079538    .056946    26.74   0.000     1.970869      2.1942
          ECS |   2.771764   .4474134     6.32   0.000     2.020025    3.803257
    nod_marg |   2.010123   .1564317     8.97   0.000      1.72576    2.341342
------------------------------------------------------------------------------

  Harrell's C concordance statistic

  Number of subjects (N)                =       602
  Number of comparison pairs (P)        =     34773
  Number of orderings as expected (E)   =     23964
  Number of tied predictions (T)        =      1360

          Harrell's C = (E + T/2) / P =     .7087
                           Somers' D =     .4174
```



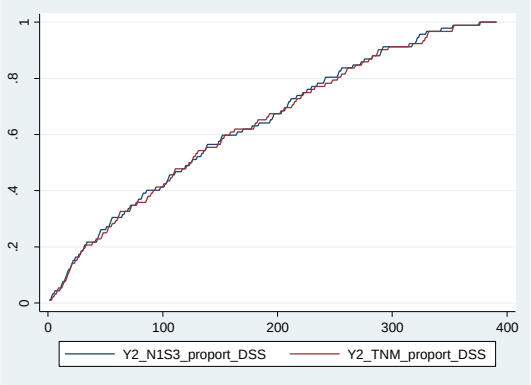
.

Additional Lift curves

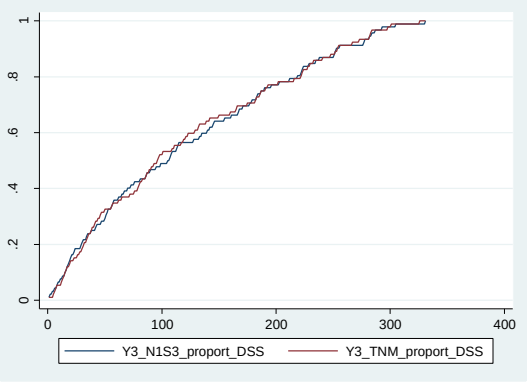
COMBINED LIFT CURVES FOR PATIENTS DEAD OR SURVIVING GREATER THAN ONE YEAR



COMBINED LIFT CURVES FOR PATIENTS DEAD OR SURVIVING GREATER THAN TWO YEARS



COMBINED LIFT CURVES FOR PATIENTS DEAD OR SURVIVING GREATER THAN THREE YEARS



COMBINED LIFT CURVES FOR PATIENTS DEAD OR SURVIVING GREATER THAN FIVE YEARS