

## RESEARCH ARTICLE

# Optimization of Predictors of Ewing Sarcoma Cause-specific Survival: A Population Study

Min Rex Cheung

### Abstract

**Background:** This study used receiver operating characteristic curve to analyze Surveillance, Epidemiology and End Results (SEER) Ewing sarcoma (ES) outcome data. The aim of this study was to identify and optimize ES-specific survival prediction models and sources of survival disparities. **Materials and Methods:** This study analyzed socio-economic, staging and treatment factors available in the SEER database for ES. 1844 patients diagnosed between 1973-2009 were used for this study. For the risk modeling, each factor was fitted by a Generalized Linear Model to predict the outcome (bone and joint specific death, yes/no). The area under the receiver operating characteristic curve (ROC) was computed. Similar strata were combined to construct the most parsimonious models. **Results:** The mean follow up time (S.D.) was 74.48 (89.66) months. 36% of the patients were female. The mean (S.D.) age was 18.7 (12) years. The SEER staging has the highest ROC (S.D.) area of 0.616 (0.032) among the factors tested. We simplified the 4-layered risk levels (local, regional, distant, un-staged) to a simpler non-metastatic (I and II) versus metastatic (III) versus un-staged model. The ROC area (S.D.) of the 3-tiered model was 0.612 (0.008). Several other biologic factors were also predictive of ES-specific survival, but not the socio-economic factors tested here. **Conclusions:** ROC analysis measured and optimized the performance of ES survival prediction models. Optimized models will provide a more efficient way to stratify patients for clinical trials.

**Keywords:** SEER - ROC - Ewing - disparity - predictor models

*Asian Pac J Cancer Prev*, 15 (10), 4143-4145

### Introduction

Ewing sarcoma (ES) is the second most common bone sarcoma of children and young adults (Herzog, 2005; Schrager et al., 2011). The ES-specific survival is about 75% overall [(Schrager et al., 2011) and this study]. EW is under active research to identify biological, pathological, and socio-economic barriers to improvement of the clinical outcome (Friedman et al., 2010; Sultan et al., 2010; Worch et al., 2010; 2011; Jawad et al., 2011; Mukherjee et al., 2011; Schrager et al., 2011; Applebaum et al., 2012). Along this line, this study used the receiver operating characteristic (ROC) curve to analyze Surveillance, Epidemiology and End Results (SEER) EW outcome data. The aim of this study was to identify and optimize predictive EW models to aid treatment and patient selection. By analyzing why some predictive models did not perform as expected, this study also discovered potential sources of outcome disparity.

Surveillance Epidemiology and End Results (SEER) (<http://seer.cancer.gov/>) is a public use cancer registry of United States of America (US). SEER is funded by National Cancer Institute and Center for Disease Control to cover 28% of all oncology cases in US. SEER started collecting data in 1973 for 7 states and cosmopolitan registries. Its main purpose is through collecting and

distributing data on cancer, it strives to decrease the burden of cancer. SEER data are used widely as a bench-mark data source for studying ES cancer outcomes in US and in other countries. The extensive ground coverage by the SEER data is ideal for identifying the disparity in oncology outcome and treatment in different geographical and cultural areas. In addition to the biological staging factors and the treatment factors, this database also contains a large number of county level socio-economic factors data. This study aimed to identify barriers to good treatment outcome that may be discernable from a national database.

SEER registry has massive amount of data available for analysis, however, manipulating this data pipeline could be challenging. SEER Clinical Outcome Prediction Expert (SCOPE) is designed and implemented to mine SEER data and construct accurate and efficient prediction models.

### Materials and Methods

The data were obtained from SEER 18 database. SEER is a public use database that can be used for analysis with no internal review board approval needed. SEER\*Stat (<http://seer.cancer.gov/seerstat/>) was used for listing the cases. The filter used was: 'Site and Morphology.AYA site recode'='Site and Morphology.ICCC site recode ICD-O-3'='VIII(c) Ewing tumor and related sarcomas of

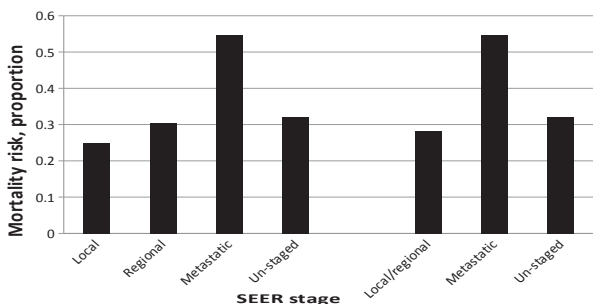
bone'. This study explored a long list of socio-economic, staging and treatment factors that were available in the SEER database. We have designed and implemented SEER Clinical Outcome Prediction Expert (SCOPE) for this purpose (Cheung, 2012). The codes of SCOPE have been posted on Matlab Central (www.mathworks.com). SCOPE has a number of utility programs that are adapted to handle the large SEER data pipeline. All statistics and programming were performed in Matlab. Each risk factor was fitted by a Generalized Linear Model to predict the outcome (SEER cause of death: Bones and Joints). The areas under the receiver operating characteristic curve (ROC) were computed. Similar strata were fused to make more efficient models if the ROC performance did not degrade (Cheung et al., 2001a; 2001b). In addition, it also implemented binary fusion and optimization to streamline the risk stratification by combining risk strata when possible. SCOPE uses Monte Carlo sampling and replacement to estimate the modeling errors and allows t-testing of the areas under the ROC. SCOPE provides SEER-adapted programs for user friendly exploratory studies, univariate recoding and parsing.

**Results**

There were 1844 patients included in this study (Table 1). The mean follow up time (S.D.) was 74.48(89.66) months. 36% of the patients were female. The mean (S.D.) age was 18.7 (12) years. The SEER staging has the highest ROC (S.D.) area of 0.616 (0.032) among the factors tested. Based on ROC areas, the other ES specific survival models included age, sex, radiation treatment, surgery (Table 1). County level percent college graduate, family income and rural-metropolitan continuum had ROC areas close to the 0.5 expected for random variables (Table 1).

To correlate with the ROC analysis, the percent ES-specific mortality was computed (Table 2). There were 44 black patients, the ROC area was close to 0.5. However, when the risk of ES death was calculated (Table 2), African American patients had worse outcome. RT was related to worse outcome marginally. Surgery improved the treatment outcome.

SEER Clinical Outcome Prediction Expert (SCOPE) simplified the 4-layered risk levels (local, regional, distant, un-staged) to a simpler non-metastatic (I and II) versus metastatic (III) versus un-staged model. The ROC area (S.D.) of the 3-tiered model was 0.612 (0.008). Several other biologic factors were also predictive of ES-specific survival, but not the socio-economic factors tested here.



**Figure 1. The Mortality Rate of Ewing Sarcoma Patients by Stage of Original and Optimized Models**

**Table 1. The Risk Models Include the Socio-demographic, Tumor and Treatment Factors for Ewing Sarcoma**

%	Model	ROC Area	S.D.	ROC area of each run				
32.10		0.548	0.015	57	54	55	53	55
67.90								
36.42276423		0.534	0.009	53	54	52	54	54
63.52303523								
24.17344173	I,II,III,IV	0.616	0.032	57	65	60	62	64
37.45257453	optimized							
28.88888889	(I,II),III, IV	0.612	0.008	62	62	61	60	61
9.430894309								
58.21138211		0.506	0.005	50	51	51	50	51
22.38482385								
1.89701897								
1.463414634								
7.696476965								
3.468834688								
1.029810298								
3.089430894								
0.54200542								
0.162601626								
48.10		0.512	0.011	50	51	53	51	51
51.90								
54.23		0.516	0.011	50	51	52	53	52
45.77								
97.45		0.502	0.004	50	50	51	50	50
2.55								
51.43631436		0.54	0.016	56	52	54	53	55
44.82384824								
0.704607046								
1.951219512								
0.596205962								
0.162601626								
0.216802168								
0.054200542								
17.23577236		0.57	0.029	57	59	59	52	58
51.8699187								
3.197831978								
1.08401084								
25.90785908								
0.325203252								
0.325203252								
52.5203252								
36.2601626								
11.17								

**Table 2. Risk of Cause Specific Mortality Associated with Different Models**

Initial univariate risk models	No. at risk	% mortality
Study population	1844	0.36
Age of diagnosis (years)		
<20	1252	0.33
≥20	592	0.43
Sex		
Female	672	0.33
Male	1172	0.38
Rural-Urban Continuum Code 2003		
Metropolitan	1629	0.37
Rural	215	0.34
County Family Income		
≥50000	887	0.37
<50000	957	0.36
County % college graduate		
≥25%	1000	0.38
<25%	844	0.34
Race		
Others	1797	0.36
Black	47	0.43
Radiation treatment given		
Beam radiation	949	0.40
None	895	0.32
Reason no cancer-directed surgery		
Surgery performed	957	0.28
Surgery not performed	887	0.45

SCOPE was used to perform ROC curve and area under the curve calculations. In this example, the ROC area of the 3-tiered SEER staging model as computed for 5 random samples (Table 1). The results are shown in the upper panels. In the lower panels, SCOPE simplified the 4-layered risk levels (local, regional, distant, and un-staged) to a simpler non-metastatic (I and II) versus metastatic (III) versus un-staged model. The ROC area (S.D.) of the 4-tiered model was 0.616 based on 5 random samples with replacement from the SEER data. The optimized model had a more efficient 3-layered structure without losing ROC area (Table 1, Figure 1).

## Discussion

This study is interested in constructing models that will aid patient and treatment selection for ES cancer patients. To that end, this study examined the ROC models (Hanley and McNeil, 1982) of a long list of potential explanatory factors (Table 1). ROC analysis takes into account both sensitivity and specificity of the prediction. Ideal model would have a ROC area of 1 and a random model is expected to have an area of 0.5 (Hanley and McNeil, 1982). The SEER staging is most predictive of patient outcome (Table 1). After binary fusion, it reduces to non-metastatic versus metastatic classification versus un-staged model (Table 1). When there are competing prediction or prognostic models, the most efficient (i.e. the simplest) model is thought to prevail (D'Amico et al., 1998). This has an information theoretic (D'Amico et al., 1998) under-pinning. However, for practical purposes, simpler models require fewer patients for a randomized trials because fewer risk strata need to be balanced. In the clinic, simpler models are easier to use. SCOPE streamlined ROC models by binary fusion (Table 1). Two adjacent strata were tested iteratively to see if they could be combined without sacrificing the higher predictive power usually belong to the more complex models. This study has shown that SCOPE can built efficient and accurate prediction models.

For surgery and radiotherapy, the ROC areas were modest 0.57 and a 0.54 respectively. Low ROC areas imply the information content (i.e. the staging accuracy) of the models may be limited. Curiously, radiotherapy was associated with higher risk of ES mortality (Table 2). Although, the overall percentage of patients completed their staging was just over 90% (Table 1). Figure 1 shows the percentage of ES-specific mortality by stage. Un-staged patients had a rate in between the early and metastatic stages of ES. Since the staging is accurate and aids treatment selection, eliminating the 9.5% un-staged rate may better pair patients with their treatments. This in turn may improve the treatment outcome.

In conclusion, this study has identified the staging models that are the most prognostic of treatment outcomes of Ewing sarcoma patients. The high under-staging rates may have prevented patients from selecting definitive local therapy. The poor rates of radiotherapy after surgery use may have contributed to the poor outcome in these patients with this aggressive disease.

## References

- Applebaum MA, Goldsby R, Neuhaus J, Dubois SG (2012). Clinical features and outcomes in patients with Ewing sarcoma and regional lymph node involvement. *Pediatr Blood Cancer*, **59**, 617-20.
- Cheung R (2012). Poor treatment outcome of neuroblastoma and other peripheral nerve cell tumors may be related to under usage of radiotherapy and socio-economic disparity: a US SEER data analysis. *Asian Pac J Cancer Prev*, **13**, 4587-91.
- Cheung R, Altschuler MD, D'Amico AV, et al (2001). ROC-optimization may improve risk stratification of prostate cancer patients. *Urology*, **57**, 286-90.
- Cheung R, Altschuler MD, D'Amico AV, et al (2001). Using the receiver operator characteristic curve to select pretreatment and pathologic predictors for early and late post-prostatectomy PSA failure. *Urology*, **58**, 400-5.
- D'Amico AV, Desjardin A, Chung A, et al (1998). Assessment of outcome prediction models for patients with localized prostate carcinoma managed with radical prostatectomy or external beam radiation therapy. *Cancer*, **82**, 1887-96.
- Friedman DL, Whitton J, Leisenring W, et al (2010). Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. *J Natl Cancer Inst*, **102**, 1083-95.
- Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiol*, **143**, 29-36.
- Herzog CE (2005). Overview of sarcomas in the adolescent and young adult population. *J Pediatr Hematol Oncol*, **27**, 215-8.
- Jawad MU, Haleem AA, Scully SP (2011). Malignant sarcoma of the pelvic bones: treatment outcomes and prognostic factors vary by histopathology. *Cancer*, **117**, 1529-41.
- Mukherjee D, Chaichana KL, Adogwa O, et al (2011). Association of extent of local tumor invasion and survival in patients with malignant primary osseous spinal neoplasms from the surveillance, epidemiology, and end results (SEER) database. *World Neurosurg*, **76**, 580-5.
- Mukherjee D, Chaichana KL, Gokaslan ZL, et al (2011). Survival of patients with malignant primary osseous spinal neoplasms: results from the Surveillance, Epidemiology, and End Results (SEER) database from 1973 to 2003. *J Neurosurg Spine*, **14**, 143-50.
- Schrager J, Patzer RE, Mink PJ, Ward KC, Goodman M (2011). Survival outcomes of pediatric osteosarcoma and Ewing's sarcoma: a comparison of surgery type within the SEER database, 1988-2007. *J Registry Manag*, **38**, 153-61.
- Sultan I, Rihani R, Hazin R, Rodriguez-Galindo C (2010). Second malignancies in patients with Ewing Sarcoma Family of Tumors: A population-based study. *Acta Oncol*, **49**, 237-44.
- Worch J, Cyrus J, Goldsby R, et al (2011). Racial differences in the incidence of mesenchymal tumors associated with EWSR1 translocation. *Cancer Epidemiol Biomarkers Prev*, **20**, 449-53.
- Worch J, Matthay KK, Neuhaus J, Goldsby R, DuBois SG (2010). Ethnic and racial differences in patients with Ewing sarcoma. *Cancer*, **116**, 983-8.