RESEARCH ARTICLE

Analysis of SEER Adenosquamous Carcinoma Data to Identify Cause Specific Survival Predictors and Socioeconomic Disparities

Rex Cheung

Abstract

Background: This study used receiver operating characteristic curve to analyze Surveillance, Epidemiology and End Results (SEER) adenosquamous carcinoma data to identify predictive models and potential disparities in outcome. Materials and Methods: This study analyzed socio-economic, staging and treatment factors available in the SEER database for adenosquamous carcinoma. For the risk modeling, each factor was fitted by a generalized linear model to predict the cause specific survival. An area under the receiver operating characteristic curve (ROC) was computed. Similar strata were combined to construct the most parsimonious models. Results: A total of 20,712 patients diagnosed from 1973 to 2009 were included in this study. The mean follow up time (S.D.) was 54.2 (78.4) months. Some 2/3 of the patients were female. The mean (S.D.) age was 63 (13.8) years. SEER stage was the most predictive factor of outcome (ROC area of 0.71). 13.9% of the patients were un-staged and had risk of cause specific death of 61.3% that was higher than the 45.3% risk for the regional disease and lower than the 70.3% for metastatic disease. Sex, site, radiotherapy, and surgery had ROC areas of about 0.55-0.65. Rural residence and race contributed to socioeconomic disparity for treatment outcome. Radiotherapy was underused even with localized and regional stages when the intent was curative. This under use was most pronounced in older patients. Conclusions: Anatomic stage was predictive and useful in treatment selection. Under-staging may have contributed to poor outcome.

Keywords: Adenosquamous carcinoma - radiotherapy - SEER registry - under usage - cause specific survival

Asian Pac J Cancer Prev, **17 (1), 347-352**

Introduction

SEER registry has massive amount of data available for analysis, however, manipulating this data pipeline could be challenging. SEER Clinical Outcome Prediction Expert (SCOPE) (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b) was used mine SEER data and construct accurate and efficient prediction models (Cheung et al., 2001a; Cheung et al., 2001b). 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.ICD-O-3 Hist/behav, malignant = '8560/3: Adenosquamous carcinoma'. This study explored a long list of socio-economic, staging and treatment factors that were available in the SEER database (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)).

The codes of SCOPE are 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 (www.mathworks.com) (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). The areas under the receiver operating characteristic curve (ROC) were computed (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). Similar strata were fused to make more efficient models if the ROC performance did not degrade (Cheung et al., 2001a; Cheung et al., 2001b). In addition, it also implemented binary fusion and optimization to streamline the risk stratification by combining risk strata when possible (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). SCOPE provides SEER-adapted programs for user friendly exploratory studies, univariate recoding and parsing (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)).

Materials and Methods

SEER registry has massive amount of data available for analysis, however, manipulating this data pipeline could be challenging. SEER Clinical Outcome Prediction Expert (SCOPE) (Cheung, 2014a; 2014b; 2014c; 2014b) was used mine SEER data and construct accurate and

*Bryn Mawr Ave, Bryn Mawr, USA *For correspondence: cheung.r100@gmail.com*

Table 1. The Risk Models Include the Socio-demographic, Tumor and Treatment Factors for Adenosquamous Carcinoma

	Number	%
	20712	
Mean	63	
S.D.	14	
< 20 years	7	0.03
≥20 years old	20705	99.97
Mean	54	
S.D.	78	
Female	13109	63.29
Male	7603	36.71
Localized, I	6801	32.84
Regional, II	6196	29.92
Distant, III	4836	23.35
Unstaged/others, IV	2879	13.90
Lung and bronchus	9472	45.73
Others	11240	54.27
Well differentiated; Grade I	790	3.81
Moderately differentiated; Grade II	4017	19.39
Poorly differentiated; Grade III	8793	42.45
Undifferentiated; anaplastic; Grade IV	780	3.77
Unknown	6332	30.57
Counties in metropolitan areas ge 1 million pop/ Counties in metropolitan areas of 250,000 to 1 million pop/ Urban pop of ge 20,000 adjacent to a metropolitan area	18446	89.06
versus		
Others	2266	10.94
≥$50000	11952	57.71
< $50000	8760	42.29
≥25%	10557	50.97
< 25%	10155	49.03
White/others	18634	89.97
Black	2078	10.03
None	11291	54.51
Beam radiation	6725	32.47
Combination of beam with implants or isotopes	1491	7.20
Refused	172	0.83
Other radiation (1973-1987 cases only)	264	1.27
Recommended, unknown if administered	293	1.41
Radioisotopes	17	0.08
Radioactive implants	192	0.93
Radiation, NOS method or source not specified	149	0.72
Unknown	118	0.57
Surgery performed	13639	65.85
Recommended but not performed, unknown reason	1992	9.62
Not recommended	4227	20.41
Recommended but not performed, patient refused	100	0.48
Not recommended, contraindicated due to other conditions	389	1.88
Recommended, unknown if performed	66	0.32
Not performed, patient died prior to recommended surgery	7	0.03
Unknown; death certificate or autopsy only case	292	1.41
Dead	9264	44.73
N/A not first tumor	3285	15.86
Alive or dead of other cause	8163	39.41

efficient prediction models (Cheung et al., 2001a; 2001b). 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.ICD-O-3 Hist/behav, malignant = '8560/3: Adenosquamous carcinoma'. This study explored a long list of socio-economic, staging and treatment factors that were available in the SEER database (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)).

The codes of SCOPE are 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 (www.mathworks.com) (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). The areas under the receiver operating characteristic curve (ROC) were computed (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). Similar strata were fused to make more efficient models if the ROC performance did not degrade (Cheung et al., 2001a; Cheung et al., 2001b). In addition, it also implemented binary fusion and optimization to streamline the risk stratification by combining risk strata when possible (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). SCOPE provides SEER-adapted programs for user friendly exploratory studies, univariate recoding and parsing (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)).

Results

There were 20712 patients included in this study (Table 1). The follow up (S.D.) was 54.2 (78.4) months. 64% of the patients were female. The mean (S.D.) age was 63 (13.8) years. There were 60% adenosquamous carcinoma patients listed from SEER database were adults. There were 7 patients younger than 20 years old in the SEER data, and it was a poor prognostic factor (Table 1 and Table 2). There is a significant female to male difference in risk of cause specific death (Table 2) favoring the female sex. 46% of the patients had lung cancers. Uterus and uterine cervix were also the common anatomic sites (Table 3). 30.6% of the tumors were not graded. Unknown grade has the highest risk of cause specific death at 51.8%. SEER stage model (localized, regional, distant, un-staged/others) was the most predictive model (ROC area or 0.71). A 4-tiered staging model was optimized to a 3-tiered model (with a ROC area of 0.67) by SCOPE (Table 1). ROC areas were used to optimize the risk models. For example, the SEER staging could be slimmed down to 3-tiered structure while not abandoning the poor (Table 1, 2 and 3). Among the socioeconomic factors studies, African American patients had 53.8% risk of death compared with 43.7% of others. However, this level of difference increased the ROC area mildly to 0.52 (Table 1). A rural residence and living a cosmopolitan area have respectively 48.7% and 44.2% risk of cause specific death (Table1, 2 and 3).

There is about 44.73% overall risk of adenosquamous carcinoma death for patients listed in SEER. The risks were 19.1% and 45.3% for localized and regional adenosquamous carcinoma respectively (Table 2). Age older than 20 years old did correlate with higher percentage mortality during this study period from 1973 to 2009 (Table 1 and Table 2). RT with external beam was associated with 54.5% risk of death, and 32.5% risk of death without RT (Table 2). Patients had surgery had 34% risk of death, 66% risk of death among patients who did not have surgery.

Table 1. The Risk Models Include the Socio-demographic, Tumor and Treatment Factors for Adenosquamous Carcinoma (Continue)

		Modelaverage ROC area	S.D.	ROC area				
Mean								
S.D.								
< 20 years		0.5	0.000	0.5	0.5	0.5	0.5	0.5
≥20 years old								
Mean								
S.D.								
Female		0.582	0.004	0.58	0.58	0.58	0.59	0.58
Male								
Localized, I	I, II, III, IV	0.714	0.005	0.71	0.72	0.71	0.71	0.72
Regional, II	optimized							
Distant, III	I, (II,III), IV	0.676	0.005	0.68	0.670	0.68	0.68	0.67
Unstaged/others, IV								
Lung and bronchus		0.574	0.005	0.57	0.57	0.57	0.58	0.58
Others								
Well differentiated; Grade I		0.59	0.007	0.58	0.59	0.6	0.59	0.59
Moderately differentiated; Grade II								
Poorly differentiated; Grade III								
Undifferentiated; anaplastic; Grade IV								
Unknown								
Counties in metropolitan areas ge 1 million pop/ Counties in metropolitan areas of 250,000 to 1 million pop/ Urban pop of ge 20,000 adjacent to a metropolitan area		0.51	0.000	0.51	0.51	0.51	0.51	0.51
versus								
Others								
≥ $50000		0.506	0.005	0.51	0.5	0.51	0.5	0.51
< $50000								
≥25%		0.502	0.004	0.5	0.5	0.5	0.51	0.5
< 25%								
White/others		0.52	0.000	0.52	0.52	0.52	0.52	0.52
Black								
None	Beam vs. others	0.572	0.004	0.57	0.57	0.57	0.57	0.58
Beam radiation								
Combination of beam with implants or isotopes								
Refused								
Other radiation (1973-1987 cases only)								
Recommended, unknown if administered								
Radioisotopes								
Radioactive implants								
Radiation, NOS method or source not specified								
Unknown								
Surgery performed	Surgery vs. others	0.65	0.000	0.65	0.65	0.65	0.65	0.65
Recommended but not performed, unknown reason								
Not recommended								
Recommended but not performed, patient refused								
Not recommended, contraindicated due to other conditions								
Recommended, unknown if performed								
Not performed, patient died prior to recommended surgery								
Unknown; death certificate or autopsy only case								
Dead								
N/A not first tumor								
Alive or dead of other cause								

Discussion

This study is interested in constructing models that will aid patient and treatment selection for adenosquamous carcinoma 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 models take 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; Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b; Cheung, 2014e; Cheung, 2015b; Cheung, 2015a; Cheung, 2015 (In press)). For example, a clinical ROC model can be used to predict if a patient receiving the recommended treatment will die from the disease. SEER stage in order to be consistent over decades, it abstracts the staging into simple but important stages for cancer progression: localized, regional and distant. Stage was the most predictive of patient outcome (Table 1). Stage has ROC of 0.71 was higher than the

0.65 of surgery. Thus complete staging is important in this disease and it may improve patient selection and council.

After binary fusion by SCOPE (Table 1), the 4 tiered stage was reduced to a 3 tiered grade based on ROC area calculations (Table 1). Un-staged grade was associated with high risk of cause specific death (Table 2). However, there is no a priori reason to put it between localized and distant. Thus it was left as a high risk factor. The solution to the uncertainty of placement of these cases is to complete the staging (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b). The binary fusion was performed to demonstrate how a complex predictive model could be numerically optimized to a much simpler model that may also be useful (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b).

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 under-pinning (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b). For practical purposes, simpler models require fewer patients for a randomized trials because fewer risk strata need to be balanced using epidemiology data (Cheung, 2014a; 2014b; 2014c; 2014b; 2014e; Cheung, 2015a; 2015b; Cheung, 2015 (In press)). 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 (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b). This study has shown that SCOPE can build efficient and accurate prediction models (Cheung, 2014c; Cheung, 2014a; Cheung, 2014d; Cheung, 2014b).

For optimized stage model (Table 1), the ROC area of 0.67 was modestly more than that of surgery. For a point of reference, using we computed the prostate risk model was 0.75 in its accuracy of predicting biochemical failure(Cheung et al., 2001a; Cheung et al., 2001b). Low ROC areas imply the information content (i.e. the staging accuracy) of the models may be limited. It is consistent with the fact that most patients did not have complete grading or staging (Table 2). This is an area of improvement. It may be a consequence of having a better guidance model in treatment and patient selection.

Adenosquamous carcinoma is an aggressive disease, there was a 19% risk of adenosquamous carcinoma death (Table 2) despite treatments even for early stage cancer.

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

References

Adegoke O, Kulasingam S, Virnig B (2012). Cervical cancer trends in the United States: a 35-year population-based analysis. *J Womens Health*.

Bhatia S (2011). Disparities in cancer outcomes: lessons learned

Table 2. Risk of Cause Specific Mortality (%) Associated with Different Models

Variables	Risk models	Number at risk	expected risk of death
Age of diagnosis	< 20 years	7	0.57
	≥20 years old	20705	0.45
Sex	Female	13109	0.38
	Male	7603	0.56
Site rec with Kaposi and mesothelioma	Lung and bronchus	9472	0.53
	Others	11240	0.38
Grade	Well differentiated; Grade I	790	0.13
	Moderately differentiated; Grade II	4017	0.32
	Poorly differentiated; Grade III	8793	0.48
	Undifferentiated; anaplastic; Grade IV	780	0.48
	Unknown	6332	0.52
SEER Staging	Localized	6801	0.19
	Regional	6196	0.45
	Distant	4836	0.70
	Unstaged/ohters	2879	0.61
Rural-Urban Continuum Code 2003	Counties in metropolitan areas ge 1 million pop/ Counties in metropolitan areas of 250,000 to 1 million pop/Urban pop of ge 20,000 adjacent to a metropolitan area	18446	0.44
	versus		
	Others	2266	0.49
County Family Income	≥ $50000	11952	0.44
	< $50000	8760	0.46
County % college graduate	≥ 25 college graduate	10557	0.45
	< 25 % college graduate	10155	0.45
Race	White/others	18634	0.44
	Black	2078	0.54
Radiation treatment given	Beam radiation	6725	0.56
	Others	13987	0.39
Reason no cancer-directed surgery	Surgery performed	13639	0.34
	Others	7073	0.66

Table 3. The Site Distribution of Adenosquamous Cancers

Site	Number	Percentage
Lung and Bronchus	9472	45.72973495
Miscellaneous	379	1.829768744
Corpus Uteri	4221	20.37850625
Breast	213	1.02833969
Cervix Uteri	2813	13.58084295
Pancreas	656	3.16709313
Other Non-Epithelial Skin	137	0.661420364
Rectum	113	0.545551103
Stomach	361	1.742866799
Gallbladder	436	2.104958239
Esophagus	512	2.471877565
Other Biliary	75	0.36209144
Nose, Nasal Cavity and Middle Ear	33	0.159320234
Kidney and Renal Pelvis	12	0.05793463
Salivary Gland	78	0.376575098
Other Female Genital Organs	34	0.16414812
Cecum	87	0.420026071
Larynx	61	0.294501038
Urinary Bladder	58	0.28001738
Small Intestine	17	0.08207406
Transverse Colon	28	0.135180804
Ovary	243	1.173176266
Anus, Anal Canal and Anorectum	82	0.395886641
Ascending Colon	51	0.246222179
Splenic Flexure	10	0.048278859
Uterus, NOS	42	0.202771206
Descending Colon	10	0.048278859
Sigmoid Colon	65	0.313812581
Vulva	28	0.135180804
Prostate	35	0.168976005
Vagina	70	0.337952011
Tonsil	22	0.106213489
Rectosigmoid Junction	35	0.168976005
Gum and Other Mouth	34	0.16414812
Hypopharynx	23	0.111041375
Floor of Mouth	14	0.067590402
Tongue	43	0.207599092
Trachea, Mediastinum and Other Respiratory Organs	10	0.048278859
Penis	3	0.014483658
Thyroid	5	0.024139429
Oropharynx	12	0.05793463
Other Urinary Organs	11	0.053106745
Other Digestive Organs	7	0.033795201
Lip	2	0.009655772
Eye and Orbit	2	0.009655772
Peritoneum, Omentum and Mesentery	4	0.019311543
Other Oral Cavity and Pharynx	5	0.024139429
Liver	7	0.033795201
Hepatic Flexure	14	0.067590402
Nasopharynx	8	0.038623087
Ureter	4	0.019311543
Large Intestine, NOS	7	0.033795201
Intrahepatic Bile Duct	4	0.019311543
Other Endocrine including Thymus	1	0.004827886
Appendix	2	0.009655772
Soft Tissue including Heart	1	0.004827886

from children with cancer. *Pediatr Blood Cancer*, **56**, 994-1002.

Cheung MC, Zhuge Y, Yang R, et al (2010). Incidence and outcomes of extremity soft-tissue sarcomas in children. *J Surg Res*, **163**, 282-9.

Cheung MR (2014a). Optimization of predictors of Ewing sarcoma cause-specific survival: a population study. *Asian Pac J Cancer Prev*, **15**, 4143-5.

Cheung MR (2014b). Receiver operating characteristic curve analysis of SEER medulloblastoma and primitive neuroectodermal tumor (PNET) outcome data: identification and optimization of predictive models. *Asian Pac J Cancer Prev*, **15**, 6781-5.

Cheung MR (2014c). Surveying and optimizing the predictors for ependymoma specific survival using SEER data. *Asian Pac J Cancer Prev*, **15**, 867-70.

Cheung MR (2014d). Under-use of radiotherapy in stage III bronchioaveolar lung cancer and socio-economic disparities in cause specific survival: a population study. *Asian Pac J Cancer Prev*, **15**, 4091-4.

Cheung R (2014e). Epidemiology and radiotherapy of hepatocellular carcinoma. *Int J Cancer Clin Res*, **1**, 1.

Cheung R (2015a). Smoking, air pollution and cancer: global epidemiology, public health and genomics. *Ann Transl Med Epidemiol*, **2**, 1-7.

Cheung R 2015b. Topics on radiotherapy, global cancer epidemiology and public health. *Lambert Academic Publishing*.

Cheung R 2015 (In press). Determining best contours in radiotherapy treatment in a modern era, and asian American medical epidemiology: public health point of view.

Cheung R 2015, in press. Contemporary notes on public health and radiotherapy topics, Austin Publishing Group, Austin eBooks.

Cheung R, Altschuler MD, D'Amico AV, et al (2001a). ROC-optimization may improve risk stratification of prostate cancer patients. *Urology*, **57**, 286-90.

Cheung R, Altschuler MD, D'Amico AV, et al (2001b). 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.

Galic V, Herzog TJ, Lewin SN, et al (2012). Prognostic significance of adenocarcinoma histology in women with cervical cancer. *Gynecol Oncol*, **125**, 287-91.

Hanley JA, McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.

Marcus DM, Goodman M, Jani AB, et al (2012). A comprehensive review of incidence and survival in patients with rare histological variants of prostate cancer in the United States from 1973 to 2008. *Prostate Cancer Prostatic Dis*, 15, 283-8.

Masoomi H, Ziogas A, Lin BS, et al (2012). Population-based evaluation of adenosquamous carcinoma of the colon and rectum. *Dis Colon Rectum*, **55**, 509-14.

McDowell HP, Foot AB, Ellershaw C, et al (2010). Outcomes in paediatric metastatic rhabdomyosarcoma: results of The International Society of Paediatric Oncology (SIOP) study MMT-98. *Eur J Cancer*, **46**, 1588-95.

Ognjanovic S, Linabery AM, Charbonneau B, et al (2009). Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. *Cancer*, **115**, 4218-26.

Pappo AS, Krailo M, Chen Z, et al (2010). Infrequent tumor initiative of the Children's Oncology Group: initial lessons learned and their impact on future plans. *J Clin Oncol*, **28**, 5011-6.

Perez EA, Kassira N, Cheung MC, et al (2011). Rhabdomyosarcoma in children: a SEER population based

study. J Surg Res, 170, e243-51.
Sultan I, Qaddoumi I, Yaser S, et al (2009). Comparing adult and pediatric rhabdomyosarcoma in the surveillance, epidemiology and end results program, 1973 to 2005: an analysis of 2,600 patients. *J Clin Oncol*, **27**, 3391-7.
Travis WD, Lubin J, Ries L, et al (1996). United States lung carcinoma incidence trends: declining for most histologic types among males, increasing among females. *Cancer*, **77**, 2464-70.
Wang J, Wang FW, Lagrange CA, et al (2010). Clinical features and outcomes of 25 patients with primary adenosquamous cell carcinoma of the prostate. *Rare Tumors*, **2**, 47.