RESEARCH ARTICLE

Analysis of SEER Glassy Cell Carcinoma Data: Underuse of Radiotherapy and Predicators of Cause Specific Survival

Rex Cheung

Abstract

Background: This study used receiver operating characteristic curve to analyze Surveillance, Epidemiology and End Results (SEER) for glassy cell carcinoma data to identify predictive models and potential disparities in outcome. Materials and Methods: This study analyzed socio-economic, staging and treatment factors. For risk modeling, each factor was fitted by a generalized linear model to predict the cause specific survival. Area under the receiver operating characteristic curves (ROCs) were computed. Similar strata were combined to construct the most parsimonious models. A random sampling algorithm was used to estimate modeling errors. Risk of glassy cell carcinoma death was computed for the predictors for comparison. Results: There were 79 patients included in this study. The mean follow up time (S.D.) was 37 (32.8) months. Female patients outnumbered males 4:1. The mean (S.D.) age was 54.4 (19.8) years. SEER stage was the most predictive factor of outcome (ROC area of 0.69). The risks of cause specific death were, respectively, 9.4% for localized, 16.7% for regional, 35% for the un-staged/others category, and 60% for distant disease. After optimization, separation between the regional and unstaged/others category was removed with a higher ROC area of 0.72. Several socio-economic factors had small but measurable effects on outcome. Radiotherapy had not been used in 90% of patients with regional disease. Conclusions: Optimized SEER stage was predictive and useful in treatment selection. Underuse of radiotherapy may have contributed to poor outcome.

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

Introduction

Glassy carcinoma is a relatively new and rare diagnosis. There are very few reports on the clinical history of this disease reports so far have been small series and case reports(Takahashi et al., 2011; Zhu and Li, 2011; Garg and Arora, 2012). This study is the largest study using the Surveillance Epidemiology and End Results (SEER) cancer registry data to analyze the prognostic and socioeconomic factors affecting the outcome of glassy cell carcinoma.

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 benchmark data source for studying cancer outcomes in US and in other countries (Ognjanovic et al., 2009; Sultan et al., 2009; Cheung et al., 2010; McDowell et al., 2010; Pappo et al., 2010; Bhatia, 2011; Perez et al., 2011). 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 for cancers in U.S. and could be served as a model for global public health registries (Cheung, 2014a; 2014b; 2014c; 2014d; 2015a; 2015b; 2016 (in press)).

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 explore the potential barriers to good treatment outcome that may be discernable from a national database.

Materials and Methods

The SEER registry has a massive amount of data available for analysis, however, manipulating this data pipeline could be challenging. SEER Clinical Outcome Prediction Expert (SCOPE) used mine SEER datasets to construct accurate and efficient prediction models (Cheung, 2014a; 2014b; 2014c). 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 = '8015/3: Glassy cell carcinoma'. This study explored a long list of socio-economic, staging and treatment factors

*Bryn Mawr Ave K43, Bryn Mawr, PA, USA *For correspondence: cheung.r100@gmail.com*

that were available in the SEER database. 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). 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; Cheung et al., 2001b).

Results

SEER-18 started recording the data on glassy carcinoma in 2001. There were 79 patients (Table 1) diagnosed from 2001 to 2009 were included in this study. The mean follow up time (S.D.) was 37 (32.8) months. Female patients outnumbered male ones 4:1. The mean (S.D.) age was 54.35 (19.84) years. The overall crude risk

of death from glassy cell carcinoma was about 27% but it (Cheung, 2015 (In press))varied greatly with the stage of the disease and other factors (Table 1 and Table 2). About half of the cases were cervix uteri (Table 3) but it also occurred in lung, corpus uteri, pancreas, breast and prostate. More than fifty percent of patients did not receive radiation treatment (Figure 1 and Table 1). The predictive power of each model was measured by the area under the receiver operating characteristic curve (Table 1).

Table 1 shows that the SEER stage was the most predictive factor of outcome (ROC area of 0.69). The risks of cause specific death were, respectively, 9.4% for localized, 16.7% for regional, 35% for the un-staged/others category, and 60% for distant disease. After optimization (Table 1), separation between the regional and unstaged/others category was removed with a higher ROC area of 0.72. Radiotherapy has not been used in 90% of patients with regional disease (Figure 1). Poorly and undifferentiated glassy carcinoma had a 52% risk

Table 1. Risk Models Developed Including Socio-demographic, Tumor and Treatment Factors for Glassy Cell Carcinoma

Initial univariate risk models		Number	%	Model	average ROC Area	S.D.	ROC Area				
Study population			79								
Age of diagnosis	Mean	35									
	S.D.	20									
	< 20 years	1	1.27		0.51	0.01	0.52	0.5	0.5	0.53	0.51
	>=20 years old	78	98.73								
Follow up (months)	Mean	37									
	S.D.	33									
Sex	Female	64	81.01		0.55	0.03	0.58	0.57	0.55	0.52	0.52
	Male	15	18.99								
SEER historic stage A	Unstaged, III	11	13.92	I, II, III, IV	0.69	0.12	0.54	0.82	0.71	0.61	0.79
	Distant, IV	15	18.99	optimized							
	Localized, I	32	40.51	I, (II,III),IV	0.72	0.07	0.73	0.8	0.68	0.63	0.77
	Regional, II	20	25.32								
	Localized/regional (Prostate cases), III	1	1.27								
Site of disease	Cervix uteri	44	55.70		0.55	0.06	0.5	0.55	0.66	0.52	0.52
	Others	35	44.30								
Grade	Unknown	34	43.04		0.57	0.03	0.57	0.57	0.58	0.6	0.53
	Poorly differentiated; Grade III	35	44.30								
	Undifferentiated; anaplastic; Grade IV	7	8.86								
	Well differentiated; Grade I	2	2.53								
	Moderately differentiated; Grade II	1	1.27								
Rural-Urban Continuum	Counties in metropolitan areas ge 1 million pop	42	53.16		0.58	0.02	0.54	0.59	0.57	0.59	0.59
Code 2003	Counties in metropolitan areas of 250,000 to 1 million pop	16	20.25								
	Urban pop of 2,500 to 19,999, adjacent to a metro area	6	7.59								
	Urban pop of 2,500 to 19,999, not adjacent to a metro area	2	2.53								
	Counties in metropolitan areas of lt 250 thousand pop	7	8.86								
	Urban pop of ge 20,000 not adjacent to a metropolitan area	1	1.27								
	Unknown/missing/no match (Alaska - Entire State)	2	2.53								
	Urban pop of ge 20,000 adjacent to a metropolitan area	2	2.53								
	Comp rural lt 2,500 urban pop, adjacent to a metro area	1	1.27								
County Family Income	>= $50000	25	31.65		0.54	0.02	0.52	0.56	0.56	0.53	0.53
	< $5000054		68.35								
County % college graduate	>=25%	21	26.58		0.56	0.04	0.53	0.59	0.6	0.55	0.51
	< 25%	58	73.42								
Race	White/others	68	86.08		0.56	0.06	0.52	0.53	0.59	0.65	0.5
	Black	11	13.92								
Radiation treatment given	None	40	50.63		0.56	0.02	0.58	0.55	0.59	0.55	0.55
	Combination of beam with implants or isotopes	14	17.72								
	Refused	2	2.53								
	Beam radiation	17	21.52								
	Recommended, unknown if administered	1	1.27								
	Radioactive implants	2	2.53								
	Unknown	2	2.53								
	Radiation, NOS method or source not specified	1	1.27								
Reason no cancer-directed surgery	Not recommended	37	46.84		0.68	0.05	0.61	0.73	0.72	0.67	0.67
	Surgery performed	36	45.57								
	Recommended but not performed, unknown reason	2	2.53								
	Recommended but not performed, patient refused	2	2.53								
	Unknown; death certificate or autopsy only case	2	2.53								
SEER cause specific Survival	Dead	21	26.58								
	Alive or dead of other cause	51	64.56								
	N/A not first tumor	7	8.86								

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		
	>=20 years old		
Sex	Female	64	0.27
	Male	15	0.27
Site rec with Kaposi and mesothelioma			
Primary site	Cervix uteri	44	0.23
	Others	35	0.31
Grade	Well differentiated; Grade I	2	0.00
	Moderately differentiated; Grade II	1	1.00
	Poorly differentiated; Grade III	35	0.23
	Undifferentiated; anaplastic; Grade IV	7	0.00
	Unknown	34	0.35
Rural-Urban Continuum	Counties in metropolitan areas ge 1 million pop/ Code 2003 Counties in metropolitan areas of 250,000 to 1 million pop/ Urban pop of ge 20,000 adjacent to a metropolitan area	65	0.29
	versus		
	Others	14	0.14
County Family Income	>= $50000	25	0.32
	< $50000	54	0.24
County % college graduate	>= 25 college graduate	21	0.29
	< 25 % college graduate	58	0.26
Race	White/others	68	0.26
	Black	11	0.27
Radiation treatment given	Beam radiation, Beam and Brachytherapy	31	0.29
	Others	48	0.25
Reason no cancer-directed surgery	Surgery performed	36	0.11
	Others	43	0.40
SEER Staging	Localized	32	0.09
	Regional	12	0.17
	Un-staged/others	20	0.35
	Distant	15	0.60

Table 3. Site Distribution of Glassy Cell Carcinoma

Site	Number	Percentage
Lung and Bronchus	9	11.39
Ovary	1	1.27
Cervix Uteri	44	55.70
Breast	4	5.06
Pancreas	4	5.06
Corpus Uteri	7	8.86
Prostate	3	3.80
Thyroid	1	1.27
Urinary Bladder	2	2.53
Sigmoid Colon	1	1.27
Esophagus	1	1.27
Miscellaneous	1	1.27
Kidney and Renal Pelvis	1	1.27
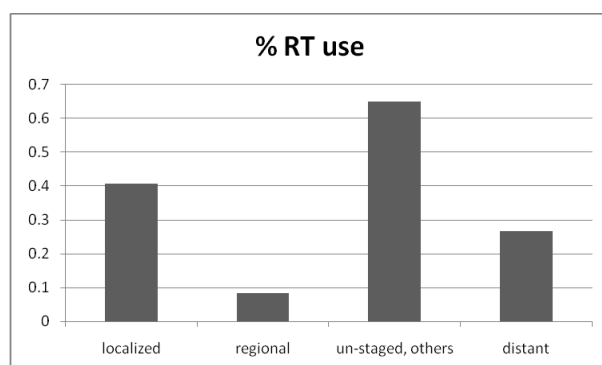
Figure 1. Probability of Being Treated with Radiotherapy by SEER Stage

of cause specific death. 42.5% of patients did not have their tumor graded. Being un-graded carried a 35% risk of cause specific death (Table 1 and Table 2). Among the socio-economic factors, lower county family income level and rural residence had better prognosis (Table 2). This is worth further investigation.

Discussion

This study is interested in constructing models that will aid patient and treatment selection for glassy cell 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). 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 was the most predictive of patient outcome (Table 1). SEER stage has ROC of 0.69 that was the highest among the other factors tested. Thus complete staging is important for patient selection and council.

After binary fusion by SCOPE (Cheung, 2014a; Cheung, 2014b), the 4 tiered grade was reduced to a 3 tiered grade based on ROC area calculations (Table 1). Un-staged/others was associated with intermediate risk of cause specific death (Table 2). Although in this study,

the un-staged was fused with the regional category, there was no a priori reason to put un-staged category between regional and metastatic categories. The solution to the uncertainty of placement of these cases is to complete the staging. 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.

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. 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 build efficient and accurate prediction models.

For radiotherapy, the ROC area of 0.56 (Table 1) was modestly more than 0.5. 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.

Glassy cell carcinoma is an aggressive disease when it is metastatic, there was a 16.7% risk of glassy cell carcinoma death (Table 2) despite treatments. There was only 8% use of RT (Figure 1) for these patients. Thus radiation oncologists should be more attentive in recommending RT for these patients.

In conclusion, this study has identified the staging models are the most prognostic of treatment outcomes of biliary tract cancer 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

Bhatia S (2011). Disparities in cancer outcomes: lessons learned 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). Surveying and optimizing the predictors for ependymoma specific survival using SEER data. *Asian Pac J Cancer Prev*, **15**, 867-70.

Cheung MR (2014c). 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 (2014d). 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.

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, 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.

Garg MM, Arora VK (2012). Clear cell adenosquamous carcinoma of the cervix: a case report with discussion of the differential diagnosis. *Int J Gynecol Pathol*, **31**, 294-6.

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

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**, 243-51.

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.

Takahashi Y, Sasaki H, Mogami H, et al (2011). Adjuvant combined paclitaxel and carboplatin chemotherapy for glassy cell carcinoma of the uterine cervix: report of three cases with clinicopathological analysis. *J Obstet Gynaecol Res*, **37**, 1860-3.

Zhu HT, Li SX (2011). Glassy cell carcinoma of cervix: a clinicopathologic analysis of 5 cases. *Zhonghua Bing Li Xue Za Zhi*, **40**, 523-7 (in Chinese).