Available online at www.elixirpublishers.com (Elixir International Journal)

Human Physiology

Elixir Human Physio. 52 (2012) 11363-11366

ISSN: 2229-712X

Role of tumor markers AFP, HCG, and LDH in survivability studies of testicular cancer patients

Srinivasa Lingireddy[1] and Ajay Lingireddy[2]

[1]KYPIPE LLC, Lexington, KY, USA. Formerly, Department of Civil Engineering, University of Kentucky, USA.

[2]IHHS, Cincinnati, OH, USA.

ARTICLE INFO

Article history:

Received: 26 September 2012;

Received in revised form:

30 October 2012;

Accepted: 6 November 2012;

Keywords

Survivability,

Testicular,

Tumor,

Cancer.

ABSTRACT

Survivability studies are integral part of cancer research and are particularly important in testicular cancer (TC) studies as TC appears to affect young adults most and is the most curable of all cancers leaving a long life ahead for TC survivors. Numerical models were developed correlating several different parameters documented in SEER (Surveillance, Epidemiology and End Results) database to survivability of TC patients and comparing the model predicted results with actual survival times. The study demonstrates the importance of tumor markers AFP, hCG, and LDH in maintaining strong correlations.

Introduction

Testicular cancer is one of the most curable forms of cancers; nevertheless nearly 9000 deaths are attributed to this cancer every year worldwide (Juška et al 2011). Significant advances in techniques for treating germ cell tumors over the past 4 decades has improved the overall survival rates for testicular cancer patients from about 70% in early 1970s to the current rates of over 90%. In the United States of America, the 5 year survival rate for testicular cancer patients is at 97% with about 8000 patients diagnosed with cancer and about 350 deaths every year (Howlader et al 2012). Incident rates of testicular cancer are very high in Western and Northern Europe and are the lowest in Asia and Africa. However, the mortality rates are the highest in Central America (0.7%) and Western Asia (0.6%) (Rosen et al 2011 and NCRI 2012). Testicular cancer is the most common form of cancer in men between 15 and 40years age which puts a lot of emphasis on survivability studies. The incidence of testicular cancer has been on the rise worldwide and the rate is going up more rapidly in the industrialized world (Huyghe et al 2003).

Serum markers are used extensively in early detection of many forms of cancers, and testicular cancer is no exception. Three different serum tumor markers appear to play a dominant role in testicular cancer studies: 1. Alpha Fetoprotein (AFP), 2. Human Chorionic Gonadotropin (hCG), and 3. Lactate Dehydrogenase (LDH). While elevated levels of these serum markers signal the need for further investigations and/or initiation of treatment, they also seem to help in diagnosis, staging and risk assessment, evaluation of response to therapy and early detection of relapse (Stenman et al 2009).

Much of the published work on testicular cancer (outside the realm of research on causes, prevention and treatment) revolved around statistical analysis of large cancer patient databases deciphering trends in incidence, mortality, survival rates in relation to age, gender, ethnicity and type of testicular cancer and type of treatment. Schairer et al (2007) studied mortality rates of second cancer patients among testicular cancer survivors using Cox proportional hazard models in an effort to compare their mortality with those of comparable first cancers. Master et al (2010) examined changing patterns of testicular cancer among Hispanic and non-Hispanic white Americans using univariate chi-square tests and multivariate Cox proportional analyses. Travis et al (2005) employed Poisson regression analysis to model relative risks and excess absolute risks of second cancer for testicular cancer survivors in Europe and USA. Age-specific incident data among testicular cancer patients along with the Armitage and Doll equation was used by Brody (2011) in determining the number of mutations required to transform normal human cells into tumor cells. Beard et al (2010) used the Kaplan-Meier method to estimate long-term survival rates in testicular cancer patients with stage I seminoma. Oshima et al (2001) employed Cox proportional hazard model to estimate 5-year survival rates of testicular cancer patients in Osaka, Japan, based on patient's age, year of diagnosis, clinical stage, histology type, and size of hospital where the patients were treated.

The present study was aimed at developing a robust survival prediction model for testicular cancer patients based on a several different routinely documented parameters in cancer registries and then to evaluate the relative importance of each of the three tumor markers (AFP, hCG, and LDH) in accurately predicting the survival rate.

Materials and Methods

National Cancer Institute of USA collects data on cancer cases from various locations and sources throughout the country and documents the data in a database called SEER or Surveillance, Epidemiology and End Results (SEER 2012). This data collection effort started in 1973 and accumulated a large database of over 10000 testicular cancer cases documenting over 100 different parameters including age, race, marital status, age at diagnosis, type and extent of testicular cancer, tumor marker levels, treatment methods, survival time, cause of death. The

Tele:

E-mail addresses: srini@fluidhammer.com

present study employed SEER database to develop survival models for testicular cancer patients. Because of the significant advancements in cancer detection and treatment methods since the initiation of SEER, the data collection and documenting format has changed over a period of time which necessitated writing special tools to extract data in required format for the present study. This effort has ensured that data was translated accurately into the required format.

The study has employed neural network modeling tools (NeuroSort 3.0), previously developed by one the authors of this article (Lingireddy et al 2004), for the proposed survival time prediction models. The choice of neural networks was obvious for this study as neural networks offer several robust ways to work with complex multivariate problems (Lingireddy and Brion 2005, Brion et al 2000, 2002, 2005, Neelakantan et al 2001, 2002). Besides, NeuroSort 3.0 was successfully applied for a similar cancer modeling study (Baron et al 2004). Out of the 124 parameters listed in the SEER database, about 50 parameters were identified as relevant for this study based on manual screening of parameters not related to testicular cancer, incomplete data etc. For example, the use of certain parameters that were added to the database several years later might drastically reduce the total number of usable datasets, and were therefore discarded. Preliminary neural network models indicated relative insignificance of several other parameters and were dropped from the initial list. Table 1 lists the remaining parameters that showed higher relative strength indices (Chandramouli et al 2008). The earliest record (Year of diagnosis) in the resulting database was 1998 and the latest was 2003 and the total number of records was about 2200.

Table 1. List of parameters retained in database after preliminary modeling

SEER Parameter	SEER Code	Description
Marital status at diagnosis	MAR_STAT	Marital status patient at diagnosis
Race/ethnicity	RACE	Race/ethnicity
Age at diagnosis	AGE_DX	Patient's age at diagnosis
Year of diagnosis	DATE_yr	Year when the tumor was first diagnosed by a recognized medical practitioner
Histologic Type ICD-O-3	HISTO3V	Basis for staging and determination of treatment options
EOD 10 - size (1988+)	EOD10_SZ	Largest dimension of primary tumor
EOD 10 - extension	EOD10_EX	Extension of tumor away from primary site
Tumor marker 1 (AFP)	TUMOR_1V	Prognostic indicator
Tumor marker 2 (hCG)	TUMOR_2V	Prognostic indicator
Tumor marker 3 (LDH)	TUMOR_3V	Prognostic indicator
RX Summ--surg prim site	SURGPRIM	Surgical procedure that removes and/or destroys tissue of the primary site performed as part of the initial work-up or first course of therapy
Radiation	RADIATN	Method of radiation therapy performed as part of the first course of treatment
Radiation sequence with surgery	RAD_SURG	Order in which surgery and radiation therapies were administered for those patients who had both surgery and radiation
SEER other cause of death classification	o_dth_cl	Whether a patient died from something other than their cancer
Survival time recode	SURV_TM	Number of years (and months) a patient has survived after diagnosis

Neurosort3.0 provides 4 different training algorithm options: (i) Cumulative Error Back Propagation, (ii) Iterative Error Back Propagation, (iii) Radial Basis Function, and (iv) Genetic Algorithm. Preliminary modeling has indicated iterative error back propagation algorithm to be the most effective for testicular cancer dataset. All subsequent model runs were based on iterative error back propagation algorithm and with the same set of training data (learning and momentum rates, number of iterations, error tolerance, number of hidden layer nodes, choice of normalization for input and output variables, and transformation functions for hidden and output layers), for unbiased comparison of results from different model runs.

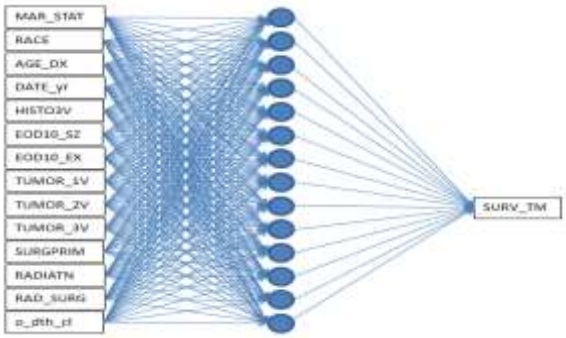

Figure 1. Schematic representation of neural network model

Figure 1 shows schematic representation of neural network model for the baseline case that includes all three tumor markers (AFP, hCG, and LDH) in input parameters list and survival time as output parameter. All records in the dataset were thoroughly shuffled before the training process and the first $3/4^{th}$ of the dataset was used for training the neural network model and the remaining $1/4^{th}$ dataset was used for testing and validation of the trained neural network model. The order of records (after shuffling) was maintained for all subsequent runs, for unbiased comparison of results. The baseline case employed all parameters listed in Table 1 for training and testing the neural network model for predicting survival times. Three other scenarios were run by removing one of the three tumor markers (AFP, hCG, and LDH) at a time to study the influence of tumor markers on the prediction efficiency of neural network models.

Results and Discussion

Case A: Survival time prediction based on all parameters listed in Table 1 including all three tumor markers – AFP, hCG, and LDH). Figure 2 compares model predicted survival times with SEER reported survival times for all 2217 records in the final dataset. Figure 3 shows a similar plot for the testing dataset comprising 555 records ($1/4^{th}$ of total number of records). Tables 2 and 3 list number of records with deviation in model predicted and SEER reported survival times at 10% intervals in training set and testing set, respectively.

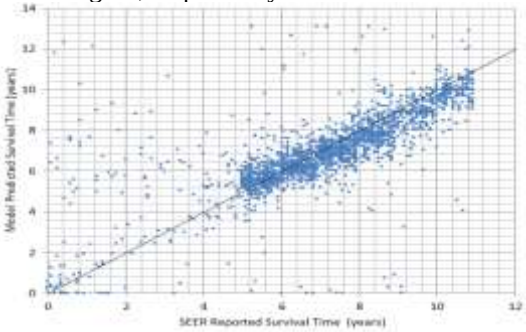

Figure 2. Comparison between SEER reported and model predicted survival times (complete dataset) for Case A

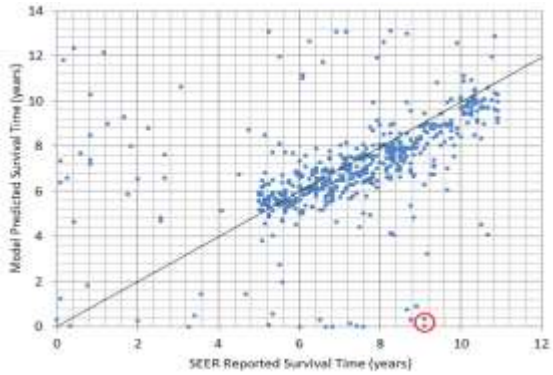

Figure 3. Comparison between SEER reported and model predicted survival times (testing dataset) for Case A

Table 2. Number of records in training set with percent deviation in model predicted and SEER reported survival times in 10% intervals (total number records = 1663)

	Number of Records from Training Set			
% Deviation	Case A All Tumor Markers	Case B AFP Removed	Case C hCG Removed	Case D LDH Removed
0-10	1446	1226	1292	1014
10-20	178	352	233	563
20-30	17	48	75	50
30-40	14	15	24	16
40-50	4	10	24	9
50-60	4	11	6	7
60-70	0	0	5	3
70-80	0	1	3	1
80-90	0	0	0	0
90-100	0	0	1	0

Table 3. Number of records in testing set with percent deviation in model predicted and SEER reported survival times in 10% intervals (total number records 554)

	Number of Records from Testing Set			
% Deviation	Case A All Tumor Markers	Case B AFP Removed	Case C hCG Removed	Case D LDH Removed
0-10	408	380	428	294
10-20	74	118	72	186
20-30	18	20	20	34
30-40	8	10	8	6
40-50	12	4	4	12
50-60	14	14	0	12
60-70	12	4	14	6
70-80	6	2	4	0
80-90	0	2	4	2
90-100	2	0	0	2

Of the 1663 records in the training set, the predicted values of survival times for over 85% of the records were within ±10% of the SEER reported survival times. Similarly, nearly 75% of the records in testing set had model predicted survival times within 10% of the SEER reported values. However, 2 of the 554 records in testing had diagonally opposite prediction in survival time compared to SEER reported values. For these two records, the SEER reported survival times were 9.08years each, but the model predicted values were 0.03 and 0.33 years, respectively. These two records are highlighted by a circle in Figure 3. Table 4 shows the input parameters for these two records (Record X and Record Y) along with the corresponding mean and standard deviation values in the complete dataset of 2217 records. A close observation of these two records shows that the values for AGE_DX parameter were significantly different from rest of the dataset. Number of records in the dataset with age parameter less than 10years was 9 (out of 2217). Only 33 of 2217 records (1.5%) were with AGE_DX parameter value greater than 60 years. The fact that AGE_DX for Record X was 61years (close to the top 1.5% of records with respect to AGE_DX parameter) coupled with a very high EOD10_SZ (representing the size of tumor in mm) value must have triggered the model to predict a very low survival time for this record. In addition, Record X had a value of 5 for TUMOR_3V parameter, which indicates an LDH value of 1.5x to 10x normal value, which also could have contributed to significantly low predicted survival time. However, for Record Y, age alone appears to be significantly off, and that would be the only possible explanation for extremely low predicted value of survival time other than more than normal AFP (1000-10000 ng/ml) and LDH (1.5x–10x normal value) tumor marker levels.

Table 4. Data for Records X and Y and along with database statistics

Input Parameter	Record X	Record Y	Database Average	Database Standard Deviation
MAR_STAT	0	1	0.4	0.5
RACE	1	1	0.9	0.3
AGE_DX	61	1	34.2	10.3
DATE_yr	1999	1999	2001.0	1.6
HISTO3V	1	0	0.6	0.5
EOD10_SZ	140	25	42.5	28.2
EOD10_EX	15	10	21.5	22.1
TUMOR_1V	2	5	2.7	1.5
TUMOR_2V	2	2	2.7	1.5
TUMOR_3V	5	5	1.9	1.9
SURGPRIM	40	40	43.0	12.3
RADIATN	0	0	0.5	0.7
RAD_SURG	0	0	1.4	1.5
o_dth_cl	0	0	0.0	0.2

While significantly low survival time predictions are certainly of great concern and represent model's inefficiency in predicting right survival times, significantly high survival time predictions compared to SEER reported values may not be interpreted the same way. The earliest value for Date_yr (year of diagnosis of testicular cancer) in the entire data set was 1999, the latest was 2003 and the largest survival time was 10.9 years (SEER database updated in Nov 2010). Though the maximum SEER reported survival time for records with DATE_yr of 2003 can only be 6.9 years, it is possible that the patient may survive much longer than that period. This appears to be the reason for more than 13years of predicted survival times for certain records shown in Figures 2 and 3. Therefore, it is essential to go over each record for which the predicted survival time was much greater than SEER reported value before treating that as improperly predicted record. This exercise was not attempted in this study but should be addressed in future for proper quantification of neural network model efficiency.

Cases B, C, and D: Same as the baseline case (Case A) but with one of the three tumor markers removed from input parameters for each case. Tables 2 and 3 show number of records with deviation in model predicted and SEER reported survival times at 10% intervals in training set and testing set, respectively, for each of these three cases. Removal of LDH (Lactate Dehydrogenase) from the input parameter set (Case D) appears to have significantly affected model's efficiency. Only 294 of 554 records in testing set had model predicted survival times within 10% of the SEER reported values. 4 records had diagonally opposite predicted values compared to SEER reported values. Similar drops in efficiency were noted in training set as well. Removal of hCG from the input parameter list (Case C) appears to have mixed effect on prediction efficiency. More records had better predicted values for survival times for testing set compared to results from Case A (all tumor

markers included in input parameters) while fewer number of records with better predicted values for training set compared to results from Case A. Results from Case B were worse than those reported in Case A but were slightly better than those reported in Case D.

Conclusions

Neural network models were able to predict survival times with reasonable accuracy based on 14 of 100+ different parameters documented in SEER database. Model predicted survival times were within 10% of SEER reported survival times for nearly 75% of the records in testing dataset comprising $1/4^{th}$ of well shuffled 2217 record database. The prediction efficiency could be higher considering that the actual survival times for certain patients could be more than those reported in SEER database, unless those patients died by the time SEER database was last updated. The importance of tumor markers in the input parameter set for efficient prediction of survival times was clearly evident from the results reported in this study. Removing LDH from input parameter set dropped the efficiency by nearly 20% in predicting survival times within 10% of SEER reported values for testing set, and by about 25% for training set. Overall prediction efficiency was consistently low when any of the three tumor markers (AFP, hCG, and LDH) were dropped from the input parameter set demonstrating the importance of tumor markers for more accurate prediction of survival times for testicular cancer patients.

References

- Baron, A.T., King, M.S., Dickerson, K., Durbin, E., Williams, J., Redmond, J., Hopenhayn, C.,Kelly, K., Lynn, B., Brion, G.M., Lingireddy, S., Chandramouli, V., Shelton, B., and Cibull, M. "Cancer Biomarkers and their Relationship to Bioinformatics", Proc. University of Kentucky Markey Cancer Center Symposium on The Relationship of Tobacco and HPV with Lung, Head & Neck, and Cervical Cancer, Lexington, KY, USA, May 2004.
- Beard, C.J., Chen, M., Arvold, N.D., Nguyen, P.L., Ng, A.K., Hoffman, K.E., Long-term Survival and Competing Causes of Death in Men with Stage I Seminoma, Int. J. of Radiation Oncology, Biology, Physics, Vol 78(3) Nov 2010
- Brion G.M., Neelakantan, T., and Lingireddy, S. "Using Neural Networks to Predict Peak Cryptosporidium Concentrations," Journal American Water Works Association, 93(1), 99-105, 2000
- Brion, G.M., Neelakantan, T.R., and Lingireddy, S. A Neural Network Based Classification Scheme for Sorting Sources and Ages of Fecal Contamination in Water, Water Research, Vol. 36, No. 15, 3765-3774, 2002
- Brion, G., Chandramouli V., T.R.Neelakantan , Lingireddy, S., Rosina Girones, David Lees, Annika Allard, Apostolos Vantarakis, Probing Virus Presence in Shellfish with Artificial Neural Networks, Applied and Environmental Microbiology, Vol 71(9) 5244-5253, 2005.
- Brody, J.P. Age-Specific Incidence Data Indicate Four Mutations Are Required for Human Testicular Cancers, J. PLoS One, Vol 6 (10) 1-5, Oct 2011.
- Chandramouli, V. Lingireddy, S. and Brion, G.M. A Robust Training Terminating Criterion for Neural Network Modeling of Small Datasets, ASCE Jl. of Computing in Civil Engineering. Vol 21(1) 39-46, 2008
- Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations), National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2009_pops09/, based on November 2011 SEER data submission, posted to the SEER web site, April 2012.
- Huyghe, E., Matsuda, T., and Thonneau, P. Increasing incidence of testicular cancer worldwide: a review., Journal Urology V 170(1) 5-11, July 2003.
- Juška, A., Ulys, A., Kairevičė, L., Smailytė, G., Stankevičius, E., Jančiauskienė, R., Jievaltas, M. Survival of Patients With Testicular Cancer in Lithuania During 1999–2002. Medicina (Kaunas) Vol 47(1):52-6, 2011.
- Lingireddy, S., Brion G.M., Neelakantan, T., and Chandramauli, V. NeuroSort3.0, Department of Civil Engineering, University of Kentucky, Lexington, Kentucky, USA, 40506, 2004.
- Lingireddy, S., and Brion, G.M., Editors, Artificial Neural Network Applications in Water Supply Engineering. Task Committee Report. Water Supply Engineering Committee, ASCE, 2005
- Master, V.A., Johnson, T.V., Hsiao, W., Jani, A.B., Changing patterns of testicular cancer among Hispanic and non-Hispanic white Americans in the twenty and twenty-first centuries: A study of the SEER database. 2010 Genitourinary Cancers Symposium, American Society of Clinical Oncology, 2010.
- NCRI, Cancer Trends No. 15, Cancers of Testis, National Cancer Registry Ireland, August 2012.
- Neelakantan, T., Brion, G.M., and Lingireddy, S. Neural Network Modeling of Cryptosporidium and Giardia Concentrations in the Delaware River, Water Science and Technology, Vol 43 (12), 125-132, 2001
- Neelakantan, T.R., Lingireddy, S., and Brion, G.M., Relative Performance of Different ANN Training Algorithms in Predicting Protozoa Concentration in Surface Waters, ASCE Journal of Environmental Engineering, Vol 128(6), 533-542, 2002
- Oshima, A., Kitagawa, T., Ajiki, W., Tsukuma, H., Takenaka, S., Iura, A., Survival of Testicular Cancer Patients in Osaka, Japan. Japan J. Clinical Oncology, Vol 31 (9), 438-43, 2001.
- Rosen, A., Jayram, G., Drazer, M., and Eggener, S.E., Global Trends in Testicular Cancer Incidence and Mortality, European Urology Vol 60, 374-79, 2011.
- Schairer, C., Hisada, M., Chen, B.E., Brown, L.M., Howard, R., Fosså, S.D., Gail, M., Travis, L.B., Comparative Mortality for 621 Second Cancers in 29 356 Testicular Cancer Survivors and 12 420 Matched First Cancers, J. National Cancer Institute Vol 99(16), 1248-56, August 2007.
- SEER. *About the SEER Program*. Web. 17 Sept. 2012. <http://seer.cancer.gov/about/>.
- Stenman, U., Lamerz, R., and Looijenga, L.H., Tumor Markers in Testicular Cancers, Chapter 2, *Laboratory Medicine Practice Guidelines, Use of Tumor Markers in Testicular, Prostate, Colorectal, Breast, and Ovarian Cancers*, Sturgeon and Diamandis, Eds. National Academy of Clinical Biochemistry, AACC, 2009.
- Travis, L.B., ., Fosså, S.D., Schonfeld, A.J., McMaster, M.L., Lynch, C.F., Storm, H., Hall, P., Holowaty, E., Andersen, A., Pukkala., E., Andersson, M., Kaijser, M., Gospodarowicz, M., Joensuu, T., Cohen, R.J., Boice, J.D. Jr., Dores, G.M., and Gilbert, E.S. Second Cancers Among 40 576 Testicular Cancer Patients: Focus on Long-term Survivors, J. National Cancer Institute, Vol 97 (18), 1354-65, Sept 2005.