Characterization of non-monotonic relationships between tumor mutational burden and clinical outcomes

Jordan Anaya[1], **Julia Kung**[2], **and Alexander S. Baras**[1,3,4]

[1]**Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA**
[2]**Applied Health Sciences Informatics, Johns Hopkins University School of Medicine, Baltimore, MD, USA**
[3]**The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA**
[4]**Bloomberg∼Kimmel Institute for Cancer Immunotherapy, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA**

Corresponding author:
Alexander S. Baras[1,3,4]

Email address: baras@jhmi.edu

ABSTRACT

Potential clinical biomarkers are often assessed with Cox regressions or their ability to differentiate two groups of patients based on a single cutoff. However, both of these approaches assume a monotonic relationship between the potential biomarker and survival. Tumor mutational burden (TMB) is currently being studied as a predictive biomarker for immunotherapy, and a single cutoff is often used to divide patients. In this study we introduce a two-cutoff approach that allows splitting of patients when a non-monotonic relationship is present, and explore the use of neural networks to model more complex relationships of TMB to outcome data. Using real-world data we find that while in most cases the true relationship between TMB and survival appears monotonic, that is not always the case and researchers should be made aware of this possibility.

INTRODUCTION

When searching for features predictive of survival across different cancer types researchers often use Cox regression[1–3]. The Cox model provides a high level of interpretability in the form of the regression coefficients, but these coefficients simply describe the linear relationship of the feature to the predicted log partial hazard[4]. Often there isn't a clear reason to assume a linear (and more importantly monotonic) relationship; furthermore, neural nets have previously been proposed as a more flexible approach in this context[5], with some recent applications appearing in the literature[6–9].

Tumor mutational burden (TMB) is a commonly investigated biomarker in the context of immunotherapy[10–16], and its prognostic value has also been investigated in the context of heterogeneous treatments[17,18]. When investigating TMB as a biomarker researchers often bin patients into a "TMB low" group and a "TMB high" group, which implicitly assumes a monotonic relationship of TMB with survival, independent of the number of bins used. In this case the relationship is assumed to be a step function with patients below a certain threshold having a certain risk and patients above the threshold having another. The monotonic assumption is that change in risk only increases (or only decreases) with the value of the predictor variable in question.

Similar to how we previously leveraged the flexibility of neural networks to more optimally model the calibration of next-generation sequencing (NGS) gene panel-derived estimates of exomic TMB[19], we wondered whether such flexible modeling approaches could be applied to characterizing the relationship of TMB with clinical outcomes data. While a single cutoff approach may work well for monotonic relationships, it would be expected to be suboptimal in the case of a truly non-montonic relationship. In such a scenario the lowest and highest TMB values would have similar risk, and the moderate values would be associated with a different risk. In this study we explored different approaches to attempt to

better characterize these more complex relationships and investigated whether such relationships exist in the context of TMB and clinical outcomes data, both in the prognostic sense and predictive sense.

RESULTS

Given the long right-tailed distribution of TMB (Supplemental Figure 1A), when modeling the relationship of this data to the log partial hazard some form of a transformation will generally be required, since these extreme values multiplied by the estimated model coefficient will generate unrealistic partial hazards. We can demonstrate this by fitting a Cox model to untransformed TMB values for uterine corpus endometrial carcinoma (UCEC) and the corresponding survival data in the The Cancer Genome Atlas (TCGA) dataset[20]. Generating predicted survival curves for several different TMB values (Supplemental Figure 1B), we see that what the field would consider large differences in TMB result in minimal differences in the curves, and at the highest TMB values patients are predicted to survive longer than humanly possible.

Approaches with Simulated Data

One common approach for transforming TMB is to define a "TMB low" group and a "TMB high" group, which can be described as transformation via a step function with a given cutoff. This cutoff can be defined by taking the median TMB value, or some other percentile of the data; alternatively, a cutoff can be found that optimizes some metric difference between the two groups. We demonstrate what the optimal cutoff approach looks like for simulated survival data with a monotonic relationship (Figure 1A, B). In this simulated data the TMB values are real (the values of the UCEC TCGA samples), but the survival data is generated according to a simple linear mapping of logged TMB values to log risks (representing a monotonic relationship). As expected, in every simulation a single cutoff approach correctly identified a "TMB high" group with a significant hazard ratio.

While a single cutoff approach works well for this data, we wondered what would happen with a simple but non-monotomic relationship. Further, we wanted to understand what approach could be used to identify this type of relationship. Instead of searching for a single cutoff we decided to search for two cutoffs, and merge the lowest and highest groups into a single group. Two cutoffs have been proposed before[21], but in those contexts three groups were generated where the "TMB mid" group was expected to have a risk in between that of "TMB low" and "TMB high", which retains the monotonic nature of the relationship. In contrast, in our two-cutoff approach we assign the same risk to the "TMB low" and "TMB high" groups with the "TMB mid" displaying a different survival risk, which represents a non-monotonic relationship that some describe as a "Goldilocks effect".

		Linear Data		Non-monotonic Data	
Simulated Dataset	Model	LL score	C-index	LL score	C-index
1	Ground Truth	-.895	.601	-.812	.696
	Cox-PH	-.895	.601	-.857	.669
	FCN	-.896	.601	-.817	.689
5	Ground Truth	-.771	.617	-.879	.664
	Cox-PH	-.771	.617	-.915	.611
	FCN	-.772	.617	-.879	.658

Table 1. Simulated data metrics. Log-likelihood and C-indexes for the test folds of two representative monotonic and non-monotonic simulated datasets for either a Cox model or a neural network.

In the case of a monotonic relationship, we would expect grouping the highest and lowest values together would result in poor correlation, and when we applied the new two-cutoff approach to the monotonic data the statistical significance was often lower than the single-cutoff approach (Figure 1C), with the "TMB mid" group often having the higher hazard ratio. However, in the case of a non-monotonic relationship, grouping the lowest and highest values is exactly what should be done and when these two different approaches were applied to data simulating a non-monotonic relationship of TMB with survival we found the two-cutoff approach correctly associated a middle group with a larger hazard ratio, and typically did so with greater statistical significance than the single-cutoff approach (Figure 1D, E, F).

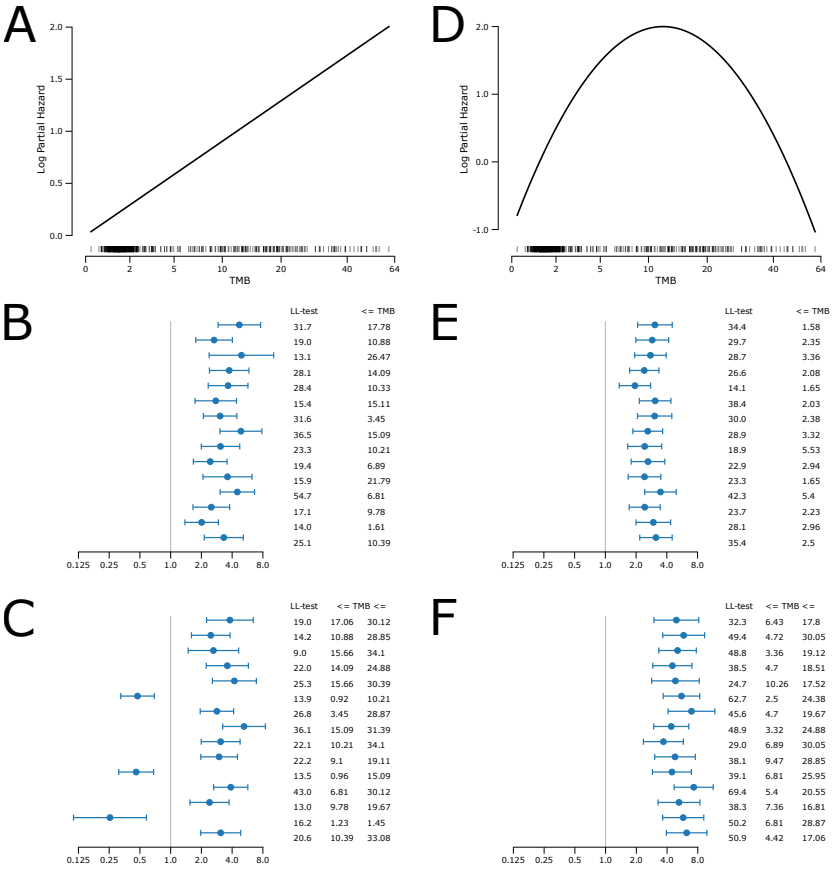

Figure 1. Cutoff analysis with simulated data. 15 simulated survival datasets were generated for either a monotonic relationship with TMB (A) or a non-monotonic relationship (D). B and E show the hazard ratios and associated log-likelihood ratio tests and associated cutoffs of searching for an optimal cutoff with either the monotonic relationship or non-monotonic relationship, respectively, while C and F show the search for the optimal 2 cutoffs of our proposed approach with a monotonic or non-monotonic relationship, respectively. In A, B the TMB distribution is shown as a rug plot.

These results give a potential framework for identifying non-monotonic relationships of an input variable to survival—if the single-cutoff approach has a better test statistic then the relationship is possibly monotonic, and if the two-cutoff approach has a better test statistic the relationship is possibly non-monotonic. However, while we generally saw this expected pattern in the simulated data it was not always the case, and this heuristic doesn't reveal the true underlying relationship, such as a linear relationship versus a step function or others. Instead of comparing different transformation strategies (one cutoff versus two), we can simply allow a fully connected neural network (FCN) to learn the relationship between predictor variable and outcome measures directly from the available data.

We looked at two of the simulated survival datasets for both the linear and non-monotonic relationships, and compared the results of a fully connected network (FCN) to a Cox regression implemented with lifelines. To compare the model fits we looked at the C-index and log-likelihood (a larger value is better in both cases) for data in the test folds of 10 stratified K-folds, and also compared these metrics to what would be obtained with the ground truth relationship (Table 1). With regards to the linear data both the Cox model and the FCN had metrics nearly indistinguishable from the ground truth, but for the non-monotonic data the FCN had noticeably better metrics than the Cox model. Notably, in the first dataset for the non-monotonic data the two-cutoff approach had previously obtained a lower test statistic than the single-cutoff approach and would have incorrectly inferred the relationship while the relationship is clear with an FCN.

We also visualized the model fits of the FCNs, Cox regressions, and the true underlying relationships (Figure 2). Despite having a large number of parameters our FCN produced a very similar fit to a Cox regression for the linear data, with both

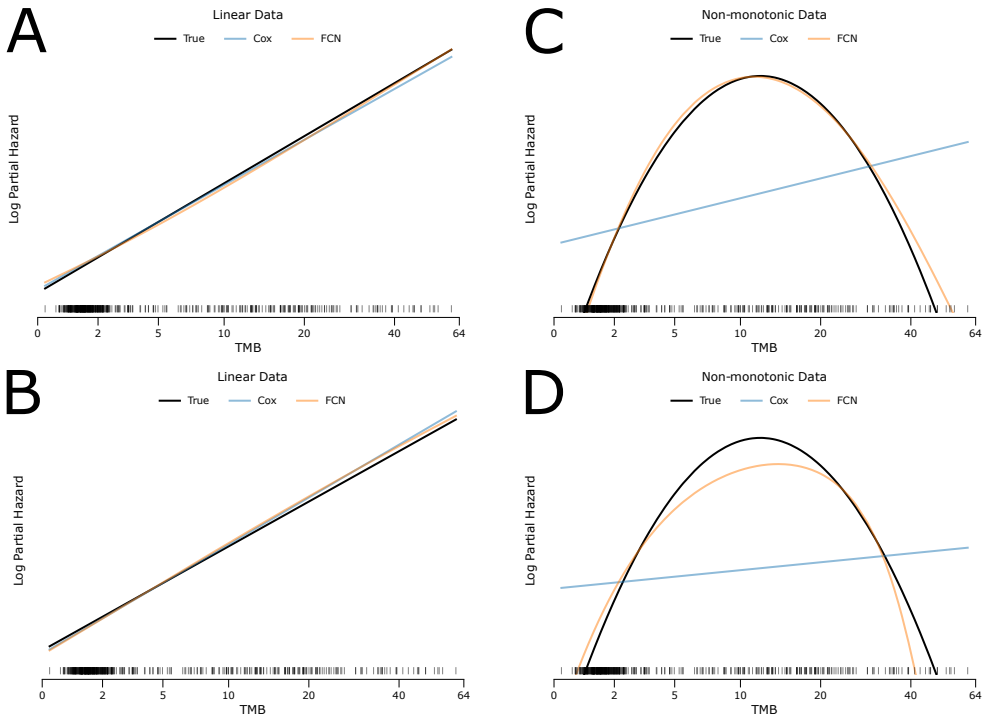

Figure 2. Simulated data model fits. Model fits from a Cox model and a neural net were averaged over 10 K-folds for two representative simulated survival datasets for a linear and non-monotonic relationship. A, B monotonic data. C, D non-monotonic data. TMB distributions shown as rug plots.

closely following the true relationship (Figure 2A, B). Visually, the extra parameters of the FCN allowed it to correctly follow the shape of the non-monotonic data (Figure 2C, D), while the Cox model was a poor fit to the data, which is consistent with the model metrics.

Applying Neural Networks to the TCGA

The simulated data gave us confidence that if the relationship strongly deviates from a monotonic relationship, then an FCN will be able to better model this and this would be reflected in the C-index and model log-likelihood metrics. Our next question was whether this tool would allow us to find evidence for such "Goldilocks effects" in real-world datasets. Using the mutation call and survival data for solid tumors in the TCGA dataset, we compared the model metrics of a Cox model and our FCN (Table 2).

In most cancer types there was limited difference between a Cox model and an FCN, suggesting the relationship likely is monotonic (or in some cases simply no relationship). However, in SKCM the FCN displayed a noticeably better log likelihood and C-index. SKCM is a cancer type where immunotherapy is recommended and where TMB hase been suggested as predictive biomarker. Looking at some of the model fits we see tumor types where the FCN predicted a perfectly linear relationship and followed the Cox predictions, cases where the FCN has slight deviations from the Cox predictions, and then SKCM where the FCN predicts a strong concave up relationship (Figure 3).

Applying Neural Networks to Immunotherapy Datasets

The TCGA dataset contains patients receiving the standard of care at the time, and includes patients at different stages of disease. In this heterogeneous context, exploring the relationship of TMB with survival is more closely associated with some form of prognostication rather than predicting response to a given therapeutic modality. In contrast, when investigating a cohort in which a specific treatment has been administered we generally are considering a biomarker that is predictive of response. It is in this context that TMB has often been explored as a biomarker, and many approaches effectively attempt to stratify the cohort into "TMB low" and "TMB high" and compare clinical outcome measures

	Cox-PH		FCN	
Cancer	LL score	C-index	LL score	C-index
BLCA	-1.337	.573	-1.338	.570
CESC	-.593	.512	-.600	.490
ESCA	-.842	.556	-.842	.556
GBM	-1.967	.491	-1.967	.478
HNSC	-1.381	.532	-1.379	.540
KIRC	-.643	.616	-.643	.616
KIRP	-.372	.573	-.372	.573
LGG	-.630	.712	-.625	.712
LIHC	-.953	.572	-.955	.572
LUAD/LUSC	-1.488	.514	-1.488	.514
OV	-1.415	.569	-1.416	.569
PAAD	-1.177	.551	-1.193	.553
COAD/READ	-.670	.580	-.670	.580
SKCM	-1.346	.578	-1.322	.617
STAD	-1.203	.559	-1.204	.559
UCEC	-.515	.551	-.521	.562

Table 2. TCGA data metrics. Log-likelihood scores and C-indexes for the test folds of different cancers in the TCGA for a Cox model and a neural net.

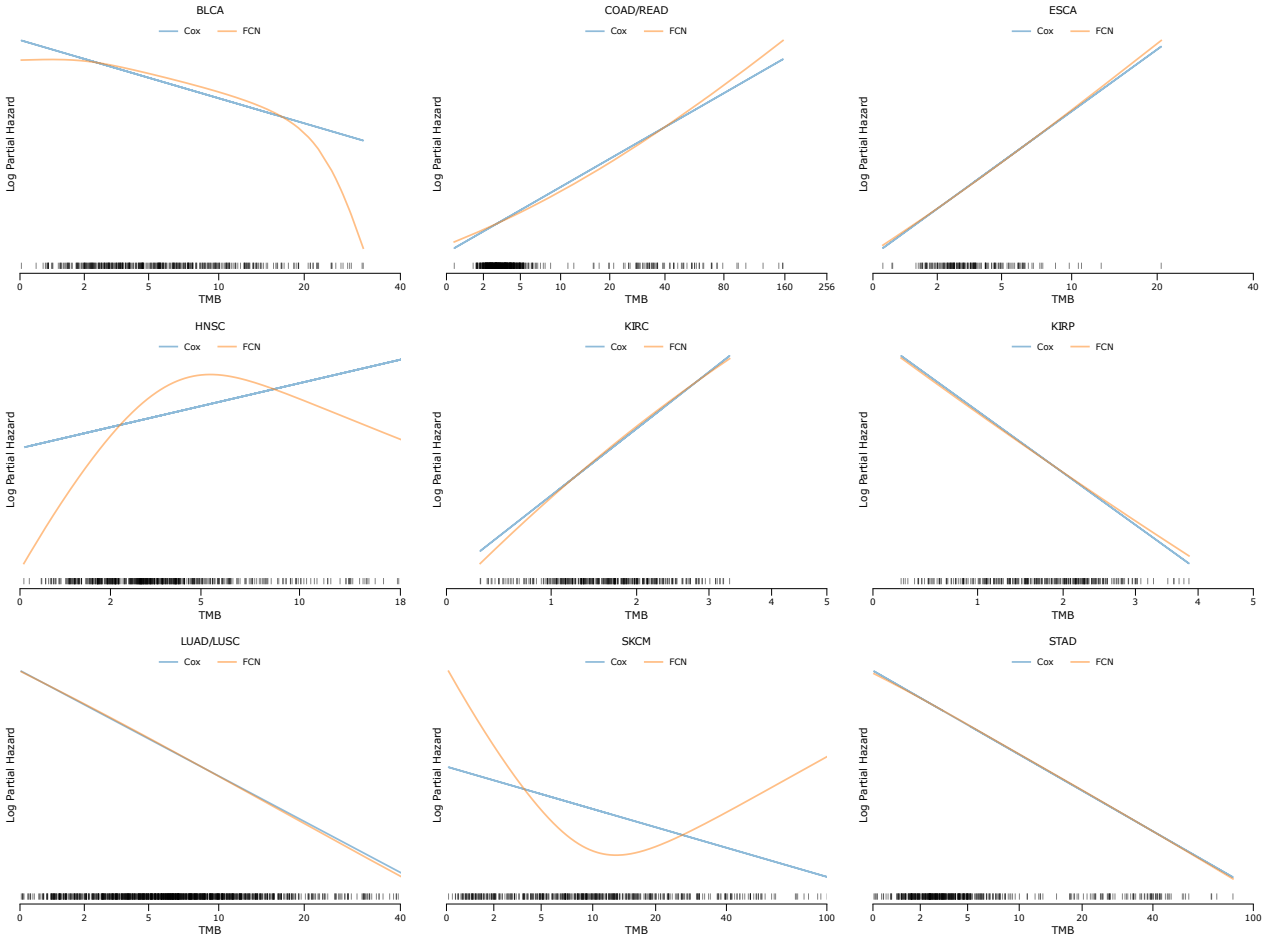

Figure 3. TCGA model fits. Nine selected cancer types are shown. Cox and neural net model fits were averaged over 10 K-folds. TMB distributions shown as rug plots.

(response, disease-free survival, etc). Given our findings of a non-monotonic relationship of survival for SKCM in the TCGA dataset we wondered if we would identify non-monotonic relationships in other datasets, either in a prognostic or predictive context.

The AACR Project GENIE Biopharma Collaborative (BPC) recently released clinical data for lung and colon cancer patients. Using the corresponding panel mutational data for these patients we applied our model to a lung cohort treated without immunotherapy (BPC NSCLC NonIO), a lung cohort treated with immunotherapy (BPC NSCLC IO), and a colon cohort treated without immunotherapy (BPC colorectal NonIO). Similar to the TCGA data our neural network did not detect non-monotonic relationships in these cancer types (Table 3, Supplemental Figure 2).

	Cox-PH		FCN	
Data	LL score	C-index	LL score	C-index
BPC NSCLC NonIO	-1.183	.568	-1.183	.552
BPC NSCLC IO	-1.486	.532	-1.486	.531
BPC Colorectal NonIO	-1.645	.511	-1.643	.477
MSK SKCM IO (2019 cohort)	-1.026	.565	-1.028	.566
MSK NSCLC IO (2019 cohort)	-1.883	.527	-1.871	.576
MSK NSCLC IO (2021 cohort)	-2.043	.514	-2.037	.551
MSK NSCLC NonIO (2021 cohort)	-1.297	.610	-1.294	.610
MSK Colorectal NonIO (2021 cohort)	-1.001	.484	-0.995	.565
MSK Pancreatic NonIO (2021 cohort)	-2.044	.577	-2.044	.577
MSK Endometrial NonIO (2021 cohort)	-0.375	.648	-0.379	.649

Table 3. BPC and MSK data metrics. Log-likelihood scores and C-indexes for the test folds of different cancers in the BPC and MSK data for a Cox model and a neural net.

Memorial Sloan Kettering (MSK) has released several large datasets of patients assayed with their gene panel and the corresponding clinical information. Without some form of common identifier across these datasets it's difficult to know how much patient overlap exists between the data releases, but we applied our FCN to a 2019 dataset which only included patients treated with immunotherapy[15], and a 2021 dataset that had both IO naive and IO treated patients[18]. While we had observed TMB to have a non-monotonic relationship with survival in the TCGA dataset for SKCM, in the IO treated MSK melanoma cohort we did not observe a difference between a Cox regression and our FCN (Table 3). This is consistent with a monotonic relationship of TMB and clinical outcome in the context of melanoma treated with IO, and supports a TMB-high vs TMB-low stratification in this context. The FCN model showed better performance over Cox modeling in the non-IO treated colorectal cancer cohort (MSK Colorectal NonIO), in which a concave down model was apparent (Supplemental Figure 3). Some improvement in performance was seen in the MSK non-small cell lung cancer cohorts treated with IO, and the observed relationship was also concave down (Supplemental Figure 3).

Model Application

The benefit of directly modeling the relationship with survival is that any deviations from a linear fit will be accounted for in the model predictions. This allows a researcher wanting to identify a cutoff for splitting patients to simply use a cutoff based on model risks. Then, working backwards from the model risks the corresponding input variable cutoff(s) can be found. If the relationship of the input variable to survival is non-monotonic then a single model risk will result in two input variable values, while a monotonic relationship will result in a single input variable cutoff.

We can demonstrate what this might look like by splitting each cancer in the TCGA dataset by either the median TMB value or median FCN model risk value. When splitting by median TMB the higher TMB group may or may not have a higher hazard, while when splitting by median model risk the higher risk group should have a higher hazard. Ignoring this difference in sign, in most cases there is almost no difference in the test statistic since the FCN predicted a monotonic relationship for most cancers (Figure 4A, B). However, for SKCM we see that a single cutoff of TMB is inappropriate. In fact, with a median cutoff the relationship of TMB to survival is barely significant while it is highly significant with a FCN model output-based risk cutoff (Figure 4C, D).

Discussion

Identifying consistent cutoffs for biomarkers has always been challenging[22] since only relationships that resemble a step function result in stable optimal cutoff values[23], a result recapitulated in our Figure 1 and Supplemental Figure 4. It is possible to identify a step function with a neural net (Supplemental Table 1, Supplemental Figure 4), but here we used

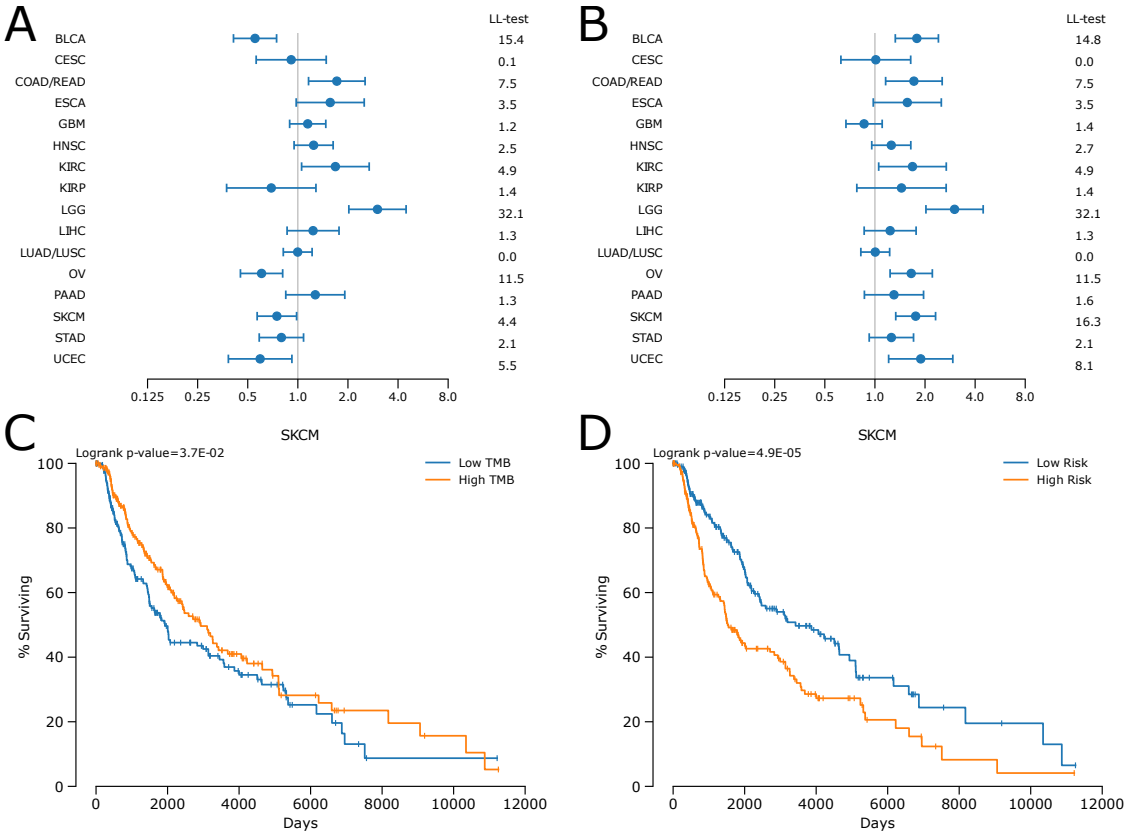

Figure 4. Model effect on a binary label. Patients were split by either median TMB (A, C), or median risk from a neural net (B, D). A, C show the hazard ratios and log-likelihood tests of each cancer type for these splits while B, D show the Kaplan curves for SKCM with these splits along with the logrank p-value.

neural nets to investigate yet another consideration when identifying potential biomarkers and cutoffs: the assumption of a monotonic relationship may not hold. If this scenario occurs it has several important implications. First, when no relationship is found with a Cox regression a very strong relationship could still be present. Second, the predictions for values at the extremes, i.e. patients who would be predicted to have the best or worst survival according to a Cox model, will contain the largest errors (see the fits for SKCM in Figure 3). Third, if a non-monotonic relationship is present then a single cutoff is inappropriate and two-cutoffs of the input variable will be needed to split patients into the appropriate two groups.

The possibility of a non-monotonic relationship is not just theoretical. While approaches to correlate TMB to survival have generally assumed a monotonic relationship, using a more flexible approach we have identified multiple datasets where the relationship between TMB and survival appears more complex. This study has a few important limitations, the first of which is that we did not undertake a more complex multivariate modeling of clinical outcome using features such as age, stage, grade, etc. This was done on purpose as the intent of this work is to highlight how the relationship of TMB to clinical outcome can be modeled to account for potential non-monotonic relationships and not to comprehensively approach clinical outcome modeling across the relevant predictors. The second important limitation is related in that we did not design these experiments such that we would be "validating" a specific TMB $\leftrightarrow$ clinical outcome correlation across tumor type and clinical setting. Again, this was intended given the modeling focus of the work. We believe what we have developed and presented herein has important implications for studies that investigate TMB as a predictive and/or prognostic biomarker. Further, we show that fairly simple neural networks like we have presented here can help avoid the pitfalls of ordinary Cox-PH based regression with respect to the potential of non-monotonic relationships.

Methods

Code for reproducing the results in this manuscript is available at https://github.com/OmnesRes/tmb_survival and has been archived at Zenodo: https://zenodo.org/doi/10.5281/zenodo.10520419.

Modeling

Cox-PH models were implemented with lifelines version 0.27.7[24] and the neural nets with TensorFlow 2.12[25]. TMB values were log transformed before inclusion into the models. The neural nets were designed to have two dense layers of 128 with softplus activation and a dropout of .05, followed by a dense to 1 with no activation. A batch normalization layer was utilized to keep the output values centered at 0. This output was interpreted as the log risk and used to calculate the negative partial likelihood, mimicking the loss of a Cox model.

When training with the real-world datasets the TMB values in the top 1% were discarded to avoid training with extreme values. Stratified K-fold training was performed with 10 train/test splits and stratified by whether the TMB value was in the top 20th percentile to ensure a somewhat consistent range of TMB values across training splits. For each fold the ranks of the test fold data were recorded in order to calculate the concordance indexes of the models. To ensure a comparable loss calculation between the neural net and Cox models lifelines was used to calculate all losses, with the predicted risks from the neural nets being passed into a Cox model from lifelines.

Data Processing

Simulated survival data was generated by utilizing an exponential distribution[26], and a uniform distribution was used for censoring with approximately 30% of the data censored. For the simulated data only TMB values below 64 were used since it was difficult to prevent extreme simulated risks with our quadratic equation for the non-monotonic relationship, which highlights a potential issue of using polynomials for modeling data with a long tail.

The TCGA somatic mutation calls[27] were processed as previously described to calculate exomic TMB[19]. Corresponding clinical data were obtained from Liu. et al[20]. We used all available panels in the BPC data except for VICC-01-SOLIDTUMOR and UHN-48-V1 due to their small sizes. We used GENIE release 14.1[28] to obtain somatic mutations and panel coordinates, and defined TMB as nonsynonymous mutations per Mb of panel coding sequence. We only included patients whose pathology procedure occurred within 180 days of diagnosis.

AUTHORS' DISCLOSURES

The authors have nothing to disclose.

ACKNOWLEDGMENTS

The results here are in whole or part based upon data generated by the TCGA Research Network. The authors would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of study authors.

FUNDING

This research was supported in part by the Leon Troper Professorship in Computational Pathology at Johns Hopkins.

REFERENCES

1. Gentles, A. J. *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature medicine* **21,** 938–945 (2015).
2. Anaya, J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Computer Science* **2,** e67 (2016).
3. Smith, J. C. & Sheltzer, J. M. Genome-wide identification and analysis of prognostic features in human cancers. *Cell reports* **38,** 110569 (2022).

4. Cox, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34,** 187–202 (1972).

5. Faraggi, D. & Simon, R. A neural network model for survival data. *Statistics in medicine* **14,** 73–82 (1995).

6. Ching, T., Zhu, X. & Garmire, L. X. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology* **14,** e1006076 (2018).

7. Katzman, J. L. *et al.* DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology* **18,** 1–12 (2018).

8. Wang, D., Jing, Z., He, K. & Garmire, L. X. Cox-nnet v2. 0: improved neural-network-based survival prediction extended to large-scale EMR data. *Bioinformatics* **37,** 2772–2774 (2021).

9. Zhan, Z. *et al.* Two-stage Cox-nnet: biologically interpretable neural-network model for prognosis prediction and its application in liver cancer survival using histopathology and transcriptomic data. *NAR genomics and bioinformatics* **3,** lqab015 (2021).

10. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine* **371,** 2189–2199 (2014).

11. Allen, E. M. V. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350,** 207–211 (2015).

12. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* **348,** 124–128 (2015).

13. Hellmann, M. D. *et al.* Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *New England Journal of Medicine* **378,** 2093–2104 (2018).

14. Hellmann, M. D. *et al.* Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell* **33,** 853–861.e4 (2018).

15. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics* **51,** 202–206 (2019).

16. Chowell, D. *et al.* Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nature Biotechnology* (2021).

17. Wu, H.-X. *et al.* Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers. *Annals of translational medicine* **7** (2019).

18. Valero, C. *et al.* The association between tumor mutational burden and prognosis is dependent on treatment context. *Nature genetics* **53,** 11–15 (2021).

19. Anaya, J., Sidhom, J.-W., Cummings, C. A. & and, A. S. B. Probabilistic Mixture Models Improve Calibration of Panel-derived Tumor Mutational Burden in the Context of both Tumor-normal and Tumor-only Sequencing. *Cancer Research Communications* **3,** 501–509. https://doi.org/10.1158/2767-9764.crc-22-0339 (Mar. 2023).

20. Liu, J. *et al.* An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173,** 400–416 (2018).

21. Goodman, A. M. *et al.* Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Molecular Cancer Therapeutics* **16,** 2598–2608 (2017).

22. Altman, D. G., Lausen, B., Sauerbrei, W. & Schumacher, M. Dangers of Using "Optimal" Cutpoints in the Evaluation of Prognostic Factors. *JNCI Journal of the National Cancer Institute* **86,** 829–835. https://doi.org/10.1093/jnci/86.11.829 (June 1994).

23. Giannoni, A. *et al.* Do Optimal Prognostic Thresholds in Continuous Physiological Variables Really Exist? Analysis of Origin of Apparent Thresholds, with Systematic Review for Peak Oxygen Consumption, Ejection Fraction and BNP. *PLoS ONE* **9** (ed Lionetti, V.) e81699. https://doi.org/10.1371/journal.pone.0081699 (Jan. 2014).

24. Davidson-Pilon, C. lifelines: survival analysis in Python. *Journal of Open Source Software* **4,** 1317. https://doi.org/10.21105/joss.01317 (2019).

25. Abadi, M. *et al. TensorFlow, Large-scale machine learning on heterogeneous systems* Nov. 2015.

26. Bender, R., Augustin, T. & Blettner, M. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* **24,** 1713–1723. ISSN: 1097-0258. http://dx.doi.org/10.1002/sim.2059 (May 2005).

27. Ellrott, K. *et al.* Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Systems* **6,** 271–281 (2018).

28. AACR Project GENIE Consortium and others. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discovery* **7,** 818–831 (2017).