Examining the cancer mean survival time between Asian American females and white females

Emily Murphy [1] Michael Bryant [1] Monte Davityan [1] Vanessa Avilez [1] Dr. Valerie Poynor (Advisor) [1]

[1]California State University, Fullerton

Abstract

Ovarian cancer is a rare but deadly gynecological cancer. The Surveillance, Epidemiology, and End Results Program (SEER) estimate new ovarian cancer cases with comprise 1% of all new cancer cases in 2022, however, the 5-year age-adjusted survival rate after diagnosis for ovarian cancer is 50%. On average, Asian-Americans are known to have better survival compared to non-Hispanic white (NHW) Americans. However, in these studies, Asian Americans have been grouped under a single category despite encompassing numerous cultures of people. To aid in the effort of separating these populations, we analyze 9 sub-Asian populations using non-Hispanic white as the base line. In particular, we compare Kaplan-Meier survival curves and the restricted mean survival times of these data. From this, we determine significant populations to consider in regards to ovarian cancer survival time.

Data Description

The dataset we worked with consisted of 40,616 observations with 31 variables including Race, Age at diagnosis, Tumor histology, Marital status, Stage of the tumor, etc.

Ethnicity	0	1	2	3	4	5	6	7	8	NHW
Total	576	833	1113	396	332	383	330	119	649	35885
Age %										
20-29	2.6	1.2	1.1	1.0	1.8	1.3	2.4	2.5	3.5	1.0
30-39	8.2	5.4	6.0	3.8	4.2	7.3	6.7	9.2	6.5	2.9
40-49	23.4	23.2	18.1	15.7	22.0	24.5	17.3	24.4	9.4	11.2
50-59	26.7	31.9	28.8	28.0	32.8	32.1	38.2	32.8	31.4	24.9
60-69	22.0	19.7	27.0	21.7	18.4	23.0	22.1	22.7	19.9	29.8
70-85+	17.0	18.6	19.1	29.8	20.8	11.7	13.3	8.4	13.9	30.2
Stage %	0	1	2	3	4	5	6	7	8	NHW
Local-ized	16.8	21.7	18.9	16.4	15.4	20.4	15.8	21.8	25.4	15.6
Regional	23.9	26.3	30.3	30.1	24.7	34.2	30.6	24.4	31	23.2
In Situ	55.9	48	48.2	52.3	57.3	43.1	51.8	47.9	41.1	58.9
Missing	3.3	3.9	2.6	1.3	2.7	2.3	1.8	5.9	2.5	2.3

When Proportional Hazard Fails

The Cox Proportional Hazard Rates Model, as a parametric model, relies heavily on fundamental assumptions which rarely pass inspection in real world datasets. According to Kuitunen et al., only 20% out of 318 articles surveyed utilizing Cox Proportional Hazards actually mentioned, tested and passed the assumptions [3]. The two main assumptions are:

1. **Proportional Hazard Rates:** For each population of interest (strata), the hazard rates are proportional; and as such, the ratio between hazard rates is constant.
2. **Linear Covariates:** The covariates are linear in relation to the log of the hazard rates. [1]

These assumptions can be violated under the following circumstances:

1. **Proportional Hazard Rates:** If Kaplan- Meier (KM) survival curves cross, that is one indication the hazard rates are not proportional.
2. **Linear Covariates:** Many situations can arise in the real world when covariates are non-linear as well as non-additive which will violate the linearity assumption. This can be checked with residual plots.

Restricted Mean Survival Time (RMST)

Restricted Mean Survival Time (RMST) is the area under the survival curve, restricted up to some time interval. Mathematically, it is defined as

$$\mu_{\tilde{x}}^{T} = \int_0^T S(t|\tilde{x})dt,$$

where $S(t) = P(T > t)$ is the survival curve and $\tilde{x}$ is a linear combination of covariates.

Advantages:

1. **Easy interpretation:** RMST is easily interpretable even in comparison to hazard rates. An RMST score is interpreted as average time patients will survive if they are followed up at time T.
2. **Assumptions:** RMST can be used as an alternative when proportional hazards model assumptions are not met.
3. **Easy Use:** Popular statistical languages such as R and SAS have readily implemented packages that can perform RMST.

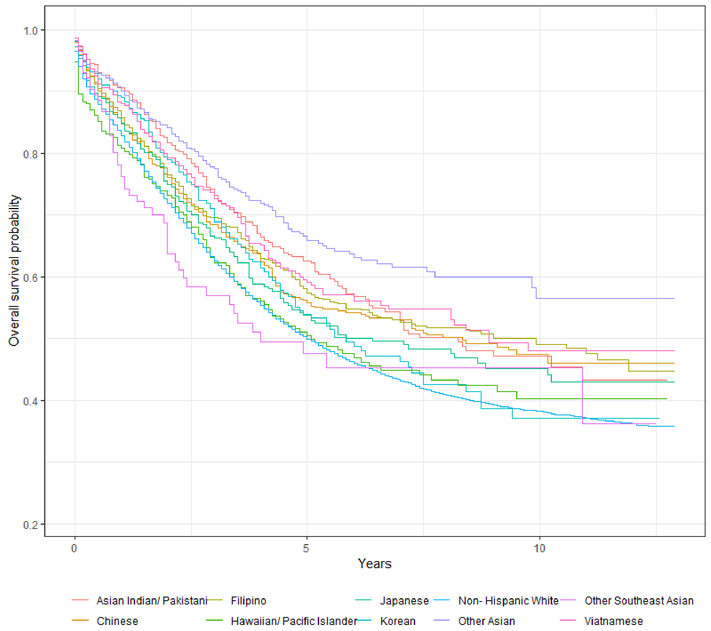
Disadvantages:

1. **Choosing T:** It may not be obvious at which T (time), the survival curve should be restricted to. Relevant exploration and dialogue must be made in order to choose an appropriate T.

Statistical Analysis

- To test for significance among Asian sub populations and the baseline (Non- Hispanic Whites), we partitioned NHW's into 8 subsamples and compared the difference in RMST
- Covariates such as Age, Site of the tumor, Marital status and more were used to control for potential confounding variables
- These covariates enter the model through the mean use identity link function to the linear combination of covariates
- Kaplan- Meyer estimates were utilized for the survival curves

Survival curve(s) / Results table



Survival Curves By Race

Race	Tau(years)	Rmst	CI (Lower)	CI (Upper)	P value
0 = Asian Indian/Pakistani	5	1.158	-0.483	2.799	0.166
	10	3.797	-0.729	8.323	0.1
1 = Chinese	5	-0.558	-2.000	0.884	0.448
	10	-1.306	-4.669	2.057	0.447
2 = Filipino	5	0.484	-0.812	1.780	0.464
	10	-0.031	-2.698	2.636	0.982
3 = Japanese	5	0.277	-1.651	2.206	0.778
	10	-1.424	-5.446	2.598	0.488
4 = Korean	5	1.854	-0.203	3.911	0.077
	10	0.496	-4.307	5.299	0.84
5 = Vietnamese	5	1.531	-0.409	3.471	0.122
	10	3.165	-1.031	7.361	0.139
6 = Hawaiian/Pacific Islander	5	-2.297	-4.472	-0.121	**0.039**
	10	-2.669	-7.156	1.818	0.244
7 = Other Southeast Asian	5	-5.368	-9.203	-1.533	**0.006**
	10	-8.654	-15.823	-1.486	**0.018**
8 = Other Asian	5	0.777	-0.765	2.319	0.323
	10	1.138	-2.817	5.092	0.573

RMST Results

Conclusions

The results showcased in the RMST results table show:

- Other Southeast Asian is significantly different than NHW at the 5 and 10 restricted levels
- Hawaiian/ Pacific Islander is significant at the 5 year restricted level
- Korean shows some evidence of difference at the 5 year level
- Finally, Vietnamese and Asian Indian showed slight evidence of difference to NHW at the 0.15 level

In conclusion, the data implies a significant difference in Restricted Mean Survival Time, when controlled for relevant covariates, between certain Asian sub populations and the baseline of Non- Hispanic Whites (NHW).

Future Work

While the question of interest was solved, we would like to continue exploring variable selection methods using LASSO in order to determine significance of covariates in the proportional hazards model. We would also like to continue exploring the potential of RMST given its ease of interpretation and little reliance on model assumptions.

References

[1] Dana Hashim and Elisabete Weiderpass. Cancer survival and survivorship. pages 250–259, 2019.

[2] National Cancer Institute. Surveillance, epidemiology, and end results program (seer), 2018.

[3] Ponkilainen V.T. Uimonen M.M. et al Kuitunen, I. Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: methodological perspectives and review. *BMC Musculoskelet Disord*, 22(489), 2021.