Investigating Exposures of Ascertained Menopause

Robert Tumasian III (rat2134)

April 19, 2021

Introduction

According to an October 2020 report by the Mayo Clinic [1], the average age at which women experience menopause in the United States is about 51 years old. This study was conducted to identify possible exposures of ascertained menopause using data collected from women who have not had a hysterectomy and did not experience menopause before intake/recruitment. "Ascertained menopause" was defined as when a woman had no menstrual periods for twelve consecutive months and will be referred to here as "menopause" for simplicity. All women were followed until they experienced menopause, died, or were censored due to dropout or study conclusion.

Data Description

Longitudinal data was collected from 380 women, in which 75 women experienced menopause during follow-up. Data was left-truncated and right-censored, and there was no missingness.

Our variables of interest were:

1. Menopause (1=experienced menopause, 0=censored)
2. Intake Age (age at recruitment, in years)
3. Menopause Age (age at menopause or censoring, in years) [*outcome #1*]
4. Menopause Time (menopause age - intake age, in years) [*outcome #2*]
5. Race (White and non-Hispanic, Black and non-Hispanic, or Other) [*exposure #1*]
6. Education (Post-graduate, graduate, some college, or high school or less) [*exposure #2*]

Objectives

The goals of this study were to:

1. Analyze and compare median menopause times under the assumption that they follow an exponential distribution (parametric), and then assuming that their distrbution is unknown (non-parametric), disregarding all other covariates.

2. Compute a non-parametric estimate of the survival function for menopause age and evaluate the median survival time. Compare it to the median obtained under the exponential assumption in objective #1.

3. Determine whether the non-parametric survival distributions for menopause age are different between the three race groups.

4. Consider if race provides additonal information about menopause age beyond that provided by education level.

Methods

For objective #1, under the exponential assumption, we estimated the median menopause time by building a survival model and obtaining the rate parameter ($\hat{\lambda}$) with a corresponding 95% CI. Then, since the menopause times are assumed to follow an exponential distribution, the estimated median menopause time would be $\log(2)/\hat{\lambda}$. To obtain a 95% CI for the median menopause time, the classical CLT was applied. Since $X \sim Exp(\lambda)$, then $\bar{X} \sim N(\lambda, \frac{\lambda^2}{n})$ for large n, where X denotes the menopause times. Since $\mu = median(X) \approx \frac{log(2)}{\hat{\lambda}}$ under the normal distribution, we can derive a 95% CI for the true median menopause time. We can show:

$$
\begin{aligned}
\alpha &= Pr\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma} \leq z_{\alpha/2}\right) \\
&= Pr\left(-z_{\alpha/2} \leq \frac{\hat{\lambda}-\lambda}{\lambda/\sqrt{n}} \leq z_{\alpha/2}\right) \\
&= Pr\left(\frac{-z_{\alpha/2}}{\sqrt{n}} \leq \frac{\hat{\lambda}}{\lambda} - 1 \leq \frac{z_{\alpha/2}}{\sqrt{n}}\right) \\
&= Pr\left(\frac{1-\frac{z_{\alpha/2}}{\sqrt{n}}}{\hat{\lambda}} \leq \frac{1}{\lambda} \leq \frac{1+\frac{z_{\alpha/2}}{\sqrt{n}}}{\hat{\lambda}}\right) \\
&= Pr\left(\frac{log(2)}{\hat{\lambda}} * (1-\frac{z_{\alpha/2}}{\sqrt{n}}) \leq \frac{log(2)}{\lambda} \leq \frac{log(2)}{\hat{\lambda}} * (1+\frac{z_{\alpha/2}}{\sqrt{n}})\right)
\end{aligned}
$$

Therefore, a 95% CI for the true median menopause time is

$$
\left[\frac{log(2)}{\hat{\lambda}} * (1-\frac{1.96}{\sqrt{n}}), \frac{log(2)}{\hat{\lambda}} * (1+\frac{1.96}{\sqrt{n}})\right],
$$

where $z_{0.05/2} = z_{0.025} = 1.96$.

Without specifying the distribution of the menopause times, a reasonable estimate for the median menopause time cannot be obtained since the curve never reaches down to 0.5 and because of the non-parametric nature of the Kaplan-Meier (KM) curve. A plot of the non-parametric KM curve overlapped by the parametric exponential curve was created to visualize and compare the two survival functions. A survival table was also created for the KM model. No exposures were considered in objective #1.

For objective #2, a KM model was used to estimate the survival curve for menopause age. Then, an associated survival table was generated to obatin an estimated median menopause age. This estimate was compared to the estimate obtained under the exponential assumption in objective #1.

Next, for objective #3, three separate KM curves for menopause age were generated for each race group. To determine if the survival distributions for menopause age were statistically different, we conducted a score test using the Cox proportional hazards (PH) model, since a log-rank test cannot be completed with left-truncated data. However, this score test is equivalent to the log-rank test.

For objective #4, to determine if race was a significant predictor of menopause age after controlling for education level, we built a PH model, which can be expressed as

$$
Y_i = h_0(t) * exp(\beta_1 * Race_i + \beta_2 * Education_i),
$$

where $i = 1, ..., 380$ denotes each woman in the study, Y_i denotes the expected hazard for woman i at time t, and $h_0(t)$ is the baseline hazard (i.e., the hazard when all predictors are 0). After constructing the model, the PH assumption was checked by plotting the Schoenfeld residuals across time and running global and individual Schoenfeld tests. The PH assumption is verified if all Schoenfeld tests yield a p-value >0.05.

Finally, we supplemented our primary objectives with the following requested information:

1. A point estimate and associated 95% CI for the hazard ratio of menopause age for a Black woman with an Other ethnicity woman, controlling for education level, with interpretations.
2. An estimate of the baseline survival function for White and non-Hispanic women with post-graduate education.

Results

Objective #1

Assuming the menopause times follow an exponential distribution, the model esimated the rate parameter to be 0.057 (95% CI: [0.045,0.071]). Since the exponential curve does not reach down to a survival probbility of 0.5 (Figure 1), we can infer the median menopause time using this rate estimate. Under the exponential assumption, we obtain an estimated median menopause time of $\log(2)/\hat{\lambda} = \log(2)/0.057 = 12.16$ (95% CI: [10.94,13.38]), where the CI is calculated using the formula in the Methods section with $n = 380$. Therefore, we expect that 50% of the women in the study will experience menopause about 12.16 years after intake/recruitment, and we are 95% confident that the true median menopause time lies in the interval [10.94,13.38].

Next, a non-parametric KM curve was fit to the data, in which the underlying distribution of the menopause times was assumed to be unknown. Similar to the exponential curve above, the KM curve also did not reach a survival probability of 0.5. However, extrapolating an appreciable median menopause time here is not possible due to the non-parametric nature of the KM model. The maximum menopause time is about 4.04 years after intake (last row of Table 1), which is the closest we can get to the median value. Thus, a one-sided 95% CI for the median menopause time via KM would be $(4.04,\infty)$. Compared to the median menopause time calculated under the exponential assumption, the median value we obtain from the non-parametric KM method is very unreliable, since we can only use the maximum menopause time.

Objective #2

Using the method of KM, we obtained an estimated median menopause age of about 55 years old (Table 2), which is close to the national average reported by the Mayo Clinic [1]. Thus, half of the women in the study experienced menopause by age 55. Here, the median can be obtained directly, since the KM curve goes below 0.5 (Figure 2). Note that the menopause age data is left-truncated, so the estimated median here is much higher than the one calculated under the exponential assumption using the non-truncated menopause time data (menopause age - intake age). The median menopause time from the exponential curve represents the expected number of years *after intake* that 50% of the women will experience menopause, whereas the median menopause age represents the expected number of years *from birth* that 50% of the women will experience menopause.

Objective #3

To determine whether the survival distributions for the race groups were different, we plotted three KM curves for menopause age for each race group (Figure 3) and used the score test described in the Methods section. Based on the score test, the survival curves for the three race groups are statistically different ($p = 0.04$) at the $\alpha = 0.05$ significance level.

Objective #4

To further explore the role of race in menopause age and the effect of education level, we built the PH model given in the Methods section (reference level for race = White and non-Hispanic, reference level for education level = high school or less). After controlling for education, race is a significant predictor of menopause age at the $\alpha = 0.05$ significance level ($p = 0.006$) (Table 3). Specifically, the hazard ratio of experiencing menopause at a certain age for a Black and non-Hispanic woman compared to a White and non-Hispanic woman is 2.50 (95% CI: [1.30,4.82]), adjusting for education level (Table 3).

Supplemental Information

After refitting the PH model using a different reference level for race (Other), we no longer achieve significance at the $\alpha = 0.05$ level (all $p > 0.05$) (Table 4). The hazard ratios for experiencing menopause at a particular age for a White and non-Hispanic woman and for a Black and non-Hispanic woman are not statistically different from a women of an Other race. Here, the hazard ratio of experiencing menopause at a certain age for a Black and non-Hispanic woman compared to a woman of an Other race is 2.64 (95% CI: [0.98,7.10]), adjusting for education level (Table 4).

We also estimated the survival curve for menopause age for White and non-Hispanic women with a post-graduate education using the KM method (Figure 4).

Lastly, since the Schoenfeld residuals for race and education are spread randomly above and below zero, and the global and individual Schoenfeld tests yielded p-values >0.05 (Figure 5), the PH assumption was verified. Thus, all of our PH model results are reliable at the $\alpha = 0.05$ significance level.

Limitations

This study had several notable limitations. First, only about 20% of the women in the study (75/380) experienced menopause. A larger sample would allow us to obtain more dependable results. Second, there was disproportionate censoring among Black women, which further restricted our results. Lastly, the data we analyzed was observational, and thus unobserved or unmeasured factors may have affected our findings.

Appendix

Tables

Table 1: Survival table for menopause time via KM.

Time	N.at.risk	N.events	Survival	lower.CI	upper.CI
0.033	379	1	0.997	0.992	1.000
0.186	376	1	0.995	0.987	1.000
0.413	373	2	0.989	0.979	1.000
0.501	370	1	0.987	0.975	0.998
0.515	368	1	0.984	0.971	0.997
0.537	367	1	0.981	0.968	0.995
0.674	364	1	0.979	0.964	0.993
0.701	363	1	0.976	0.961	0.992
0.813	361	1	0.973	0.957	0.990
0.824	360	1	0.971	0.954	0.988
0.868	359	2	0.965	0.947	0.984
0.931	355	1	0.962	0.943	0.982
0.950	353	1	0.960	0.940	0.980
0.999	339	1	0.957	0.936	0.978
1.032	338	1	0.954	0.933	0.976
1.038	336	1	0.951	0.929	0.973
1.169	333	1	0.948	0.926	0.971
1.224	332	1	0.945	0.923	0.969
1.276	328	1	0.943	0.919	0.967
1.328	327	1	0.940	0.916	0.964
1.342	324	1	0.937	0.912	0.962
1.355	323	1	0.934	0.909	0.960
1.421	317	1	0.931	0.905	0.957
1.484	316	1	0.928	0.902	0.955
1.503	315	1	0.925	0.898	0.953
1.506	314	1	0.922	0.895	0.950
1.533	313	1	0.919	0.891	0.948
1.555	312	1	0.916	0.888	0.945
1.624	309	1	0.913	0.884	0.943
1.634	308	1	0.910	0.881	0.941
1.697	303	1	0.907	0.878	0.938
1.782	301	1	0.904	0.874	0.935
1.843	299	1	0.901	0.871	0.933
1.930	298	1	0.898	0.867	0.930
1.936	297	1	0.895	0.864	0.928
1.944	295	1	0.892	0.860	0.925
2.286	286	1	0.889	0.857	0.923
2.294	285	1	0.886	0.853	0.920
2.505	282	1	0.883	0.850	0.917
2.546	280	1	0.880	0.846	0.915
2.650	278	1	0.876	0.842	0.912
2.669	277	1	0.873	0.839	0.909
2.672	276	1	0.870	0.835	0.906
2.713	275	1	0.867	0.832	0.904
2.730	274	1	0.864	0.828	0.901
2.858	271	1	0.861	0.825	0.898

Time	N.at.risk	N.events	Survival	lower.CI	upper.CI
2.867	270	1	0.857	0.821	0.895
2.951	265	1	0.854	0.817	0.893
2.960	264	1	0.851	0.814	0.890
3.064	258	1	0.848	0.810	0.887
3.116	253	1	0.844	0.806	0.884
3.149	251	1	0.841	0.803	0.881
3.165	250	1	0.838	0.799	0.878
3.176	249	1	0.834	0.795	0.875
3.233	246	1	0.831	0.792	0.872
3.296	244	1	0.827	0.788	0.869
3.329	242	1	0.824	0.784	0.866
3.365	238	1	0.821	0.780	0.863
3.395	236	1	0.817	0.776	0.860
3.398	235	1	0.814	0.772	0.857
3.406	234	1	0.810	0.769	0.854
3.425	233	1	0.807	0.765	0.851
3.455	232	1	0.803	0.761	0.848
3.485	229	1	0.800	0.757	0.845
3.532	225	1	0.796	0.753	0.841
3.546	223	1	0.793	0.749	0.838
3.636	221	1	0.789	0.745	0.835
3.685	220	1	0.785	0.742	0.832
3.803	215	1	0.782	0.738	0.828
3.814	214	1	0.778	0.734	0.825
3.885	208	1	0.774	0.730	0.822
4.041	191	1	0.770	0.725	0.818
4.044	190	1	0.766	0.721	0.815

Table 2: Survival table for menopause age via KM.

Time	N.at.risk	N.events	Survival	lower.CI	upper.CI
47.296	182	1	0.995	0.984	1.000
47.348	182	1	0.989	0.974	1.000
47.524	190	1	0.984	0.966	1.000
47.792	202	1	0.979	0.959	1.000
47.869	202	1	0.974	0.952	0.997
48.025	208	1	0.969	0.946	0.994
48.181	213	1	0.965	0.940	0.991
48.632	220	1	0.960	0.934	0.988
48.652	219	1	0.956	0.928	0.985
48.780	222	1	0.952	0.923	0.981
49.002	224	1	0.948	0.918	0.978
49.016	221	1	0.943	0.913	0.975
49.207	216	1	0.939	0.907	0.972
49.232	214	1	0.935	0.902	0.968
49.276	213	1	0.930	0.897	0.965
49.284	212	1	0.926	0.891	0.961
49.407	205	1	0.921	0.886	0.958
49.522	202	1	0.917	0.880	0.954
49.700	194	1	0.912	0.875	0.951
50.012	187	1	0.907	0.869	0.947
50.100	187	1	0.902	0.863	0.943
50.103	186	1	0.897	0.858	0.939
50.245	180	1	0.892	0.852	0.935
50.253	179	1	0.887	0.846	0.931
50.316	171	1	0.882	0.840	0.927
50.363	167	1	0.877	0.834	0.923
50.390	165	1	0.872	0.827	0.918
50.535	159	1	0.866	0.821	0.914
50.700	152	1	0.860	0.814	0.909
50.897	144	1	0.854	0.807	0.904
51.055	133	1	0.848	0.800	0.899
51.258	126	1	0.841	0.792	0.894
51.351	125	1	0.835	0.784	0.889
51.507	119	1	0.828	0.776	0.883
51.548	117	1	0.820	0.767	0.877
51.562	117	1	0.813	0.759	0.872
51.589	114	1	0.806	0.751	0.866
51.893	103	1	0.799	0.742	0.860
52.068	100	1	0.791	0.732	0.853
52.200	94	1	0.782	0.723	0.846
52.509	90	1	0.773	0.712	0.840
52.545	87	1	0.765	0.702	0.832
52.687	87	1	0.756	0.692	0.825
52.769	87	1	0.747	0.682	0.818
52.778	86	1	0.738	0.672	0.811
52.912	82	1	0.729	0.662	0.804
53.029	77	1	0.720	0.651	0.796
53.040	75	1	0.710	0.640	0.788
53.092	72	1	0.700	0.629	0.780
53.169	69	1	0.690	0.618	0.771
53.185	67	1	0.680	0.606	0.763

Time	N.at.risk	N.events	Survival	lower.CI	upper.CI
53.194	66	1	0.670	0.595	0.754
53.287	63	1	0.659	0.583	0.745
53.426	59	1	0.648	0.571	0.736
53.443	58	1	0.637	0.558	0.726
53.457	57	1	0.626	0.546	0.717
53.634	53	1	0.614	0.533	0.707
53.695	50	1	0.601	0.519	0.696
53.873	43	1	0.587	0.504	0.685
53.916	40	1	0.573	0.487	0.673
53.922	39	1	0.558	0.471	0.661
53.958	37	1	0.543	0.455	0.649
54.420	27	1	0.523	0.431	0.634
54.850	26	1	0.503	0.409	0.619
54.976	24	1	0.482	0.385	0.602
55.023	23	1	0.461	0.363	0.586
55.693	18	1	0.435	0.334	0.567
55.740	17	1	0.410	0.307	0.548
55.756	16	1	0.384	0.280	0.527
55.896	16	1	0.360	0.256	0.506
56.268	15	1	0.336	0.233	0.485
56.402	14	1	0.312	0.210	0.463
56.490	12	1	0.286	0.186	0.440
56.638	10	1	0.257	0.160	0.415
56.742	9	1	0.229	0.135	0.389

Table 3: Cox PH model results for race (reference = White and non-Hispanic), adjusting for education.

Variable	**Hazard Ratio**	**95% CI**		**p-Value**
Education = Some College	1.945	0.831	4.554	0.125
Education = College Graduate	1.006	0.426	2.374	0.989
Education = Post-Graduate	1.939	0.872	4.311	0.104
Race = Black	2.50	1.300	4.819	**0.006**
Race = Other	0.948	0.405	2.218	0.902

Table 4: Cox PH model results for race (reference = Other), adjusting for education.

Variable	**Hazard Ratio**	**95% CI**		**p-Value**
Education = Some College	1.945	0.831	4.554	0.125
Education = College Graduate	1.006	0.426	2.374	0.989
Education = Post-Graduate	1.939	0.872	4.311	0.104
Race = Black	2.640	0.982	7.095	0.054
Race = White	1.055	0.451	2.469	0.902

Figures

Figure 1: Estimated survival functions for menopause time under the exponential assumption (red curve) and under no distributional assumption using KM (black curve).

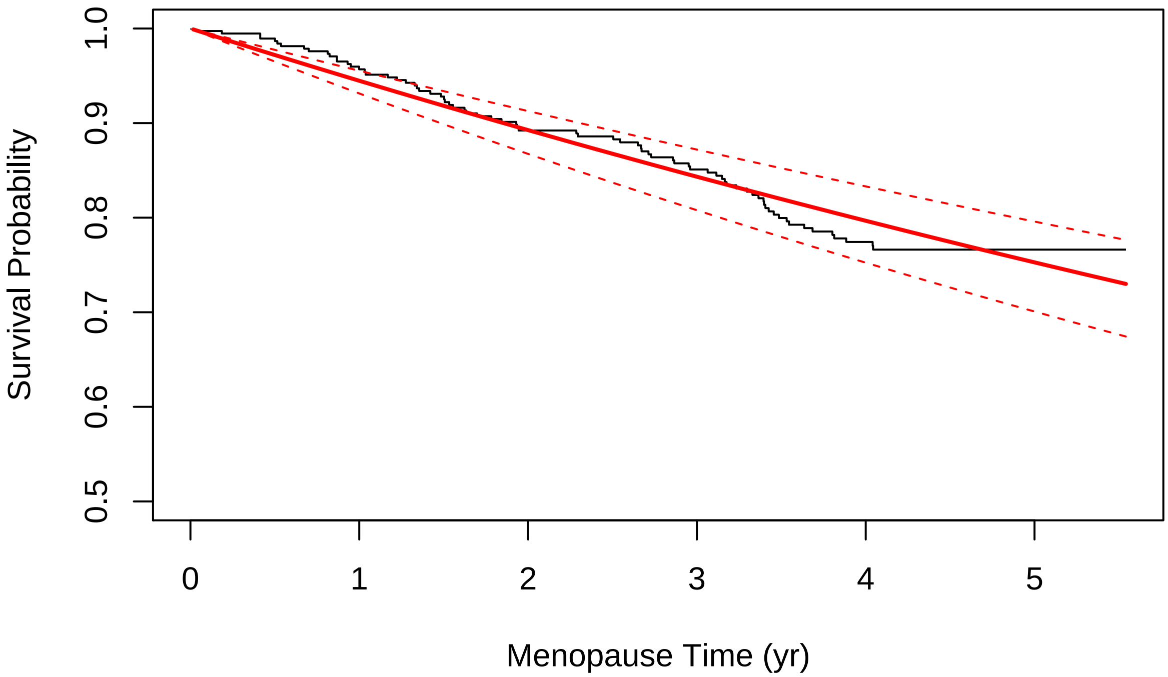

Figure 2: Estimated survival function for menopause age via KM.

Figure 3: KM survival curves for menopause age by race.

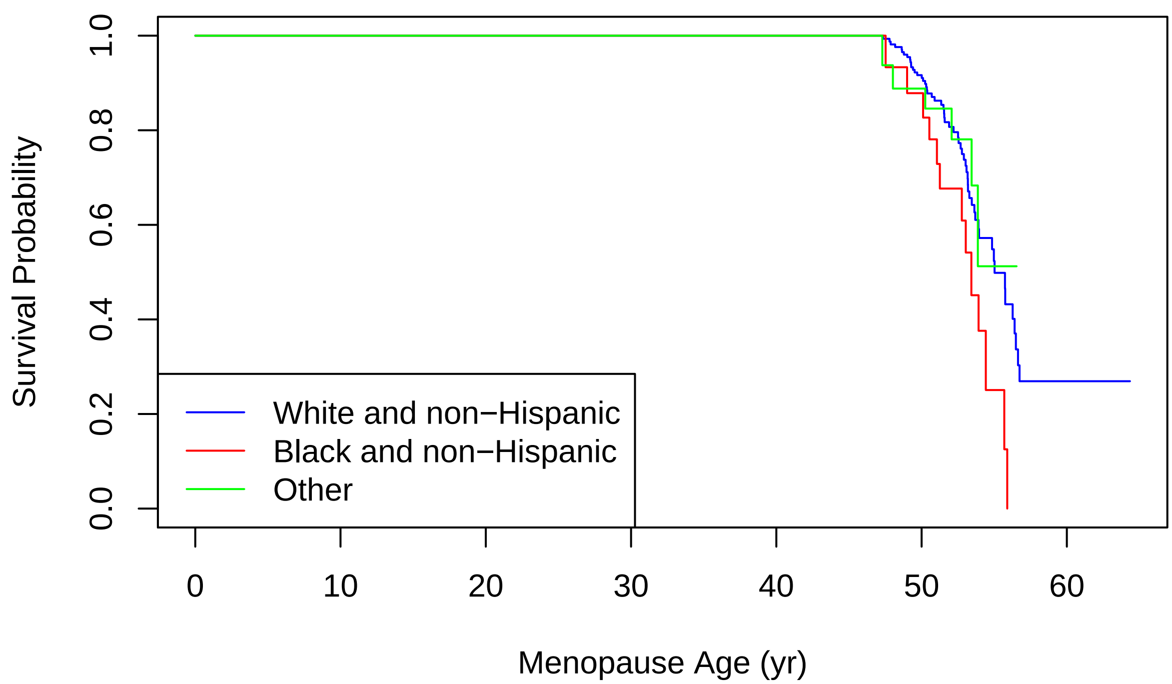

Figure 4: KM curve for menopause age for White and non-Hispanic women with a post-graduate education.

Figure 5: Schoenfeld residual plots and test results for race and education.