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 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 ($\hat{\lambda}$) and a corresponding 95% CI. Then, since the menopause times are assumed to follow an exponential distribution, the 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 the menopause times $X \sim Exp(\lambda)$, then $\bar{X} \sim N(\lambda, \frac{\lambda^2}{n})$ for large n. 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 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 Cox proportional hazards (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 relative risk of menopause age for a Black patient with an Other ethnicity patient, controlling for education level, with interpretations.
2. An estimate of the baseline survival function for White, non-Hispanic patients 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, 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 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].

This result agrees with lit…

Supplemental Information

Limitations

This study had several notable limitations. First,