1. Consider survival data from a study conducted on $n = 13$ subjects. Suppose there are 6 unique failure times and k censored observations, occuring at k unique censoring times. Let $\tau_1 < \tau_2 < \cdots < \tau_6$ denote the ordered failure times, and let $C_1 < C_2 < \cdots < C_k$ denote the k ordered censoring times.

   Let $\hat{S}$ denote the Kaplan-Meier estimator of the failure time distribution. Suppose that

$$\begin{aligned}
\hat{S}(\tau_1) &= (10/11) \\
\hat{S}(\tau_2) &= (10/11) \times (9/10) \\
\hat{S}(\tau_3) &= (10/11) \times (9/10) \times (7/8) \\
\hat{S}(\tau_4) &= (10/11) \times (9/10) \times (7/8) \times (4/6) \\
\hat{S}(\tau_5) &= (10/11) \times (9/10) \times (7/8) \times (4/6) \times (1/2) \\
\hat{S}(\tau_6) &= 0
\end{aligned}$$

   (a) (6) What is a possible value of k that is consistent with the above estimator of $S(t)$? Why?

   (b) (6) For the value of k that you've identified in (a), list the number of failures at each of the 6 failure times.

(c) (6) Give an overall ordering of the combined set of failure and censoring times (based on your identified k).

(d) (7) For the ordering of failure and censoring times you gave in (c), calculate the Kaplan-Meier estimate of the distribution of the time to censoring.

2. A follow-up study of post-menopausal women diagnosed with breast cancer was performed to examine whether the estrogen receptor (ER) status of the tumor was related to prognosis, adjusting for stage of disease at diagnosis and age at diagnosis.

   Let $h(t, X)$ be the hazard of death at month t after diagnosis for an individual with covariates X, for $X = (X_E, X_A, X_S)$ where

$$X_E = \begin{cases} 1 & \text{ER positive tumor} \\ 0 & \text{ER negative tumor} \end{cases}$$

$$X_A = \text{ age in years at diagnosis}$$

   and

$$X_S = \text{ stage } = \begin{cases} 1 & \textit{in situ} \\ 2 & \text{local} \\ 3 & \text{regional} \\ 4 & \text{distant} \end{cases}$$

   Suppose the following proportional hazards model was found to fit the data:

$$h(t, X) = h_{0X_S}(t) \exp(-.971 X_E + .003 X_A),$$

   where $h_{0X_S}(t)$ is the baseline hazard function, specific for stage (i.e. value of X_S).

   (a) (5) Based on this model, what is your best estimate of the hazard ratio (i.e. relative risk) of death for a woman with an ER positive tumor relative to a woman of the same age and with the same stage ER negative tumor, the same number of months beyond diagnosis?

   (b) (5) Based on this model, what can you say about the hazard ratio of death for a woman 52 years old at diagnosis with a localized (i.e., local stage) ER positive tumor, 24 months beyond diagnosis, relative to a woman 57 years old at diagnosis with a localized ER negative tumor, 24 months beyond diagnosis?

(c) (5) Based on this model, what can you say about the hazard ratio of death for a woman 50 years old at diagnosis with an *in situ* ER positive tumor, 36 months beyond diagnosis, relative to a woman 50 years old at diagnosis with an *in situ* ER negative tumor, 24 months beyond diagnosis?

(d) (5) Based on this model, what can you say about the hazard ratio of death for a woman 50 years old at diagnosis with an ER positive tumor with regional spread, 24 months beyond diagnosis, relative to a woman 50 years old at diagnosis with an ER negative *in situ* tumor, 24 months beyond diagnosis?

(e) (5) All other things being equal, according to this model which women have more favorable survival, women with ER positive tumors, or women with ER negative tumors?

4. A study was published on February 25, 2016 in the *New England Journal of Medicine* on weekly versus every-3-week Paclitaxel for treatment of ovarian cancer. The Kaplan Meier estimate for progression-free survival for all subjects is shown in Figure A and for subjects who did not take Bevacizumab is shown in Figure C.

   (a) (3) What can you conclude about weekly versus every-3-week treatment from the results listed in Figure A?

   (b) (3) What can you conclude about weekly versus every-3-week treatment from the results listed in Figure C?

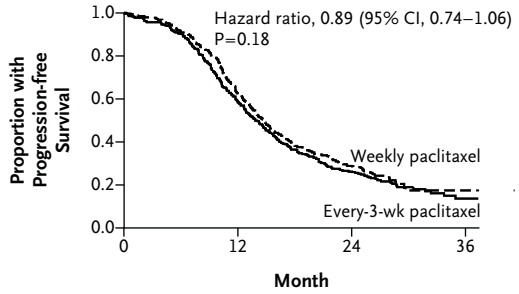
   (c) (2) Is the proportional hazards assumption justified for the Figure A analysis? For the Figure C analysis? Why or why not?

(d) (7) Express the analysis of Figure C in a proportional hazards model (i.e., please write down the model and identify coefficient(s) in the model using the results given in Figure C).

(e) (10) Suppose that the results reported in Figures A and C came from fitting a single model. Please write down this model and identify the coefficient(s) of the model using the results given in Figures A and C.

A **Progression-free Survival**

	No. of Events	Total No. of Patients	Median *mo*
Weekly paclitaxel	256	346	14.7
Every-3-wk paclitaxel	272	346	14.0

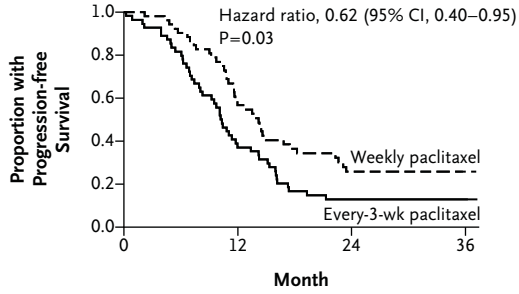


No. at Risk

Weekly paclitaxel	346	206	84	1
Every-3-wk paclitaxel	346	200	82	5

C **Progression-free Survival without Bevacizumab**

	No. of Events	Total No. of Patients	Median *mo*
Weekly paclitaxel	37	55	14.2
Every-3-wk paclitaxel	47	57	10.3



No. at Risk

Weekly paclitaxel	55	28	12	1
Every-3-wk paclitaxel	57	20	6	1