Imperial College
London

MATH97075 MATH97185

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May-June 2021

This paper is also taken for the relevant examination for the
Associateship of the Royal College of Science

Survival Models and Actuarial Applications

Date: Friday, 14 May 2021

Time: 09:00 to 11:30

Time Allowed: 2.5 hours

Upload Time Allowed: 30 minutes

This paper has 5 Questions.

Candidates should start their solutions to each question on a new sheet of paper.

Each sheet of paper should have your CID, Question Number and Page Number on the top.

Only use 1 side of the paper.

Allow margins for marking.

Any required additional material(s) will be provided.

Credit will be given for all questions attempted.

Each question carries equal weight.

SUBMIT YOUR ANSWERS AS SEPARATE PDFs TO THE RELEVANT DROPBOXES ON BLACKBOARD INCLUDING A COMPLETED COVERSHEET WITH YOUR CID NUMBER, QUESTION NUMBERS ANSWERED AND PAGE NUMBERS PER QUESTION.

1. (a) Find the density function $f(t;\theta)$, survivor function $S(t;\theta)$, and hazard function $\mu(t;\theta)$ for a random lifetime T with distribution function

$$F(t;\theta) := \frac{t\theta}{1+t\theta}, \quad t > 0$$

where $\theta > 0$ is an unknown parameter. Show your work. (6 marks)

(b) Two studies are performed to compare the survival distributions of components generated by machines A and B using proportional hazards regression. The lifetimes of the components are known to follow the distribution in (a), with respective parameters $\theta_A \neq \theta_B$. Study 1 finds strong evidence of a difference between populations A and B based on a sample consisting primarily of shorter lifetimes (longer lifetimes were right-censored). Study 2 finds no evidence of a difference based on a sample consisting primarily of longer lifetimes (this could happen if shorter lifetimes were left-censored). Explain why these results are not contradictory.
(Hint: For $\theta_A \neq \theta_B$, the ratio $\mu(t;\theta_A)/\mu(t;\theta_B)$ is not constant.) (5 marks)

(c) Find the median lifetime m as a function of θ. Recall that the median for continuous random variables is defined by $P(T \leq m) = 1/2$. (2 marks)

(d) Suppose an initial estimate $\tilde{m}$ of the median is available. Find a natural estimate $\tilde{\theta}$ of θ based on $\tilde{m}$. (2 marks)

(e) Explain how Newton's method can be used to obtain an approximate maximum likelihood estimator $\hat{\theta}_{\text{NEWTON}}$ based on an initial estimator $\tilde{\theta}$. Derive all quantities required to implement Newton's method. State an approximation to the sampling distribution of $\hat{\theta}_{\text{NEWTON}}$.
(5 marks)

(Total: 20 marks)

2. Suppose we observe n independent and identically distributed life times $T_1, \ldots, T_n$ with hazard rate $\mu(t)$ and integrated hazard rate $M(t)$. Some of the observations are right-censored.

For the ith individual, let $Y_i(t)$ denote whether the individual is at risk at time t, and let $N_i(t) = 1$ if the individual died at or before time t and $N_i(t) = 0$ otherwise.

Define $Y(t) := \sum_{i=1}^{n} Y_i(t)$ and $N(t) := \sum_{i=1}^{n} N_i(t)$.

(a) For individual i, state the intensity $\lambda_i(t)$ of $N_i(t)$. Use this to find the cumulative intensity $\Lambda(t)$ of the counting process $N(t)$. (3 marks)

(b) Show that $D(t) := N(t) - \Lambda(t)$ is a martingale. (3 marks)

(c) Derive the Nelson-Aalen estimator of $M(t)$,

$$\hat{M}(t) = \sum_{T_j \leq t} \frac{dN(T_j)}{Y(T_j)},$$

based on the processes $N(t)$ and $Y(t)$. (6 marks)

(d) Let the k unique ordered deaths times be denoted by $t_{(1)} < \ldots < t_{(k)}$. Express the usual Kaplan-Meier estimate $\hat{S}(t)$ of the survivor function $S(t)$ as a function of $N(\cdot)$ and $Y(\cdot)$. (4 marks)

(e) Show that $\hat{S}(t)$ places probability mass

$$\pi(t) = \hat{S}(t-)\frac{dN(t)}{Y(t)}$$

at each time $t > 0$ such that $Y(t) \geq 1$.

(4 marks)

(Total: 20 marks)

3. A study is investigating the association of smoking and systolic blood pressure with survival. The investigators propose the following proportional hazards model for $\mu_i(t)$, the hazard function of individual i, with

$$\mu_i(t) = \mu_0(t) \exp\left\{\texttt{sbp}_i \cdot \beta_1 + \texttt{smoke}_i \cdot \beta_2 + \texttt{smoke}_i \cdot \texttt{sbp}_i \cdot \beta_3\right\}$$

for baseline hazard function $\mu_0(t)$ and $(\beta_1, \beta_2, \beta_3)' \in \mathbb{R}^3$. The variable $\texttt{sbp}_i$ is a measurement of systolic blood pressure at the beginning of the study in units of mm Hg. The variable $\texttt{smoke}_i = 1$ if individual i is a smoker at the start of the study, and $\texttt{smoke}_i = 0$ otherwise. A large sample is collected, and the following point estimates and associated standard errors are obtained:

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Point estimate	-0.1	-2.5	0.05
Standard error	0.04	0.9	0.02

(a) Estimate $\log[\mu_i(t)/\mu_j(t)]$, where individual i is a non-smoker with systolic blood pressure of 140 mm Hg and individual j is a non-smoker with systolic blood pressure of 120 mm Hg. Based on your estimate, which individual is at greater instantaneous risk of death?
(3 marks)

(b) Estimate $\log[\mu_i(t)/\mu_j(t)]$, where individual i is a smoker, individual j is not, and both individuals have systolic blood pressure of 120 mm Hg. Based on your estimate, which individual is at greater instantaneous risk of death? (3 marks)

(c) For both of the log hazard ratios in (a) and (b), state whether it is possible to construct a 95% confidence interval from the available information. If it is possible, write and simplify an expression for the confidence interval. Otherwise, explain what additional information is required. *You may use the approximation* $P(|Z| > 2) \approx 0.05$ *for* $Z \sim N(0, 1)$. (6 marks)

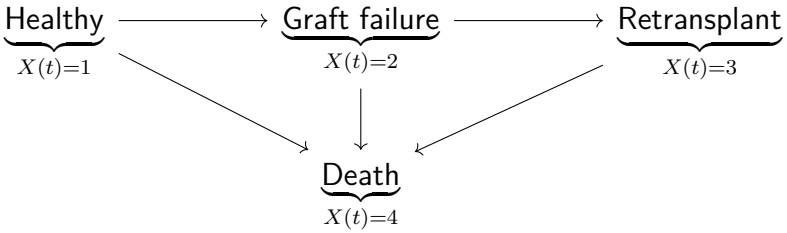
(d) Use a Wald statistic to conduct a level 0.05 hypothesis test of $H_0 : \beta_3 = 0$ versus $H_1 : \beta_3 \neq 0$. Clearly state the value of the test statistic and what your conclusion is. Do not try to calculate an exact p-value. *You may use the approximation* $P(|Z| > 2) \approx 0.05$ *for* $Z \sim N(0, 1)$.
(4 marks)

(e) For the setting of the study above, describe the baseline group of individuals that has hazard rate $\mu_0(t)$. Is this baseline group scientifically plausible? If not, suggest an alternative baseline group that could be defined in this study. (4 marks)

(Total: 20 marks)

4. Let $X(t) \in \{1, 2, 3, 4\}$ be a stochastic process denoting the current status of a patient in the first t months after receiving a bone graft. If the first graft fails, at most one additional transplant is possible.

The diagram below shows all possible transitions for a four-state Markov model for this scenario:



Let μ_{ij} denote the (constant) transition intensity from state i to state j.

(a) Use the Kolmogorov forward equations to derive the third row of the transition probability matrix for the above process,

$$p^{3j}(t), \quad j = 1, 2, 3, 4.$$

Clearly state the general form of the forward equations before using them. You may use the result that if $dy/dt = ay + b$, then $y = Ce^{at} - b/a$ for some constant C.

(Hint: you should not need to solve a differential equation for every j.)

(10 marks)

(b) The following data were obtained from bone graft recipients at a particular hospital:

V^1	V^2	V^3	D^{12}	D^{14}	D^{23}	D^{24}	D^{34}
84	4	3	21	3	1	1	1

Here, V^i denotes the total time spent by patients in state i in months and D^{ij} denotes the total number of transitions by patients from state i to state j.

(i) Estimate the *odds* that a patient experiences graft failure, conditioning on the patient leaving the healthy state. Note: Odds(A) = P(A)/P(not A). (3 marks)

(ii) Use a goodness of fit statistic to test the null hypothesis that $\mu_{12} = 1/3$ at the 0.05 level. Simplify the statistic as much as possible and clearly state your conclusion. Note that $P(\chi^2_1 > 3.84) \approx 0.05$. (3 marks)

(c) Healthy patients might have a higher risk of graft failure or death in the first three months after receiving their initial graft. What is one way we can modify the model to allow for this possibility? Describe a hypothesis test that could be used to test a null hypothesis of homogeneity using your model.

(4 marks)

(Total: 20 marks)

5. This question is based on the following paper:

> *Royston, P., Parmar, M.K. "Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome." BMC Med Res Methodol* ***13****, 152 (2013). https://doi.org/10.1186/1471-2288-13-152*

The notation used in this question follows the conventions used in that paper.

(a) Let T denote a random survival time with survivor function $S(t)$. The *restricted mean survival time (RMST)* of T is defined as

$$\mu = \int_0^{t^*} S(t)dt$$

for *suitable time horizon* t^*. Consider the following data from a small clinical trial testing novel treatment *TX* against *placebo (PL)*:

TX group

Time	At Risk	Deaths
10	9	1
100	8	3
140	5	1
160	4	2
180	2	1

Placebo (PL) group

Time	At Risk	Deaths
20	9	2
40	7	1
50	6	3
110	3	1
130	2	1
160	1	1

Use the Kaplan-Meier survival estimates in each group to estimate $\mu_{TX} - \mu_{PL}$, the difference in *RMST* with $t^* = 140$ based on the data above. Based on your estimate, which group has more favorable survival outcomes?

(8 marks)

(b) In Table 5 of the paper, comparisons are drawn between various measures of treatment effects in a trial with survival data. For the *log HR* and *RMST*, explain *in your own words* why both measures "reflect the entire survival history" but the *log HR* does "not change with extended follow-up." (6 marks)

(c) For the design of a clinical trial using the RMST, the authors propose the use of piecewise exponential distributions. For analysis of the clinical trial, the authors suggest the use of alternative flexible parametric models. Why is this approach reasonable and why might we prefer parametric models over a fully nonparametric approach? (6 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2021

This paper is also taken for the relevant examination for the Associateship.

MATH96048/MATH97075/MATH97183

Survival Models and Actuarial Applications (Solutions)

Setter's signature	Checker's signature	Editor's signature
.....	.....	.....

1. (a) The density function is

sim. seen ⇓

$$f(t;\theta) = \frac{d}{dt}F(t;\theta) = \frac{\theta}{(1+t\theta)^2}, \quad t > 0.$$

2, A

The survivor function is

2, A

$$S(t;\theta) = 1 - F(t;\theta) = \frac{1}{1+t\theta}, \quad t > 0.$$

The hazard function is

2, A

$$\mu(t;\theta) = \frac{f(t;\theta)}{S(t;\theta)} = \frac{\theta}{1+t\theta}, \quad t > 0.$$

unseen ⇓

(b) Let $g(t) := \mu(t;\theta_A)/\mu(t;\theta_B)$. From (a), we find that

$$g(t) = \frac{\theta_A + t\theta_A\theta_B}{\theta_B + t\theta_A\theta_B}.$$

We find that

2, D

$$g(0) = \frac{\theta_A}{\theta_B}$$

and that

$$\lim_{t\to\infty} g(t) = 1.$$

For both large and small values of t, the hazard ratio approaches a constant. Near $t = 0$, this ratio is θ_A/θ_B, so proportional hazards regression can be used to test the hypothesis that this ratio equals one. However, as t increases, this hazard ratio approaches one regardless of the values of θ_A and θ_B. This means that at larger times, we will not be able to distinguish between the survival distributions of the two populations on the basis of their hazard ratio, even when $\theta_A \neq \theta_B$.

3, D

unseen ⇓

(c) The median is the value m such that $F(m;\theta) = 1/2$. Hence,

2, A

$$F(m;\theta) = \frac{1}{2} \quad \Leftrightarrow \quad \frac{m\theta}{1+m\theta} = \frac{1}{2} \quad \Leftrightarrow \quad m = \frac{1}{\theta}\,.$$

unseen ⇓

(d) Using the relationship from (c), a reasonable estimate of θ is

2, B

$$\tilde{\theta} = \frac{1}{\tilde{m}}\,.$$

meth seen ⇓

(e) Let $t_1, \ldots, t_n$ denote a random sample of n lifetimes, $U \subset \{1, \ldots, n\}$ which observations are uncensored and d denote the number of deaths observed.

Newton's method for approximating the maximum likelihood estimator requires the first and second derivatives of the log-likelihood function:

2, B

$$\ell(\theta) = d\log\theta - \sum_{i\in U}\log(1+t_i\theta) - \sum_{i=1}^{n}\log(1+t_i\theta) \tag{1}$$

$$\ell'(\theta) = \frac{d}{\theta} - \sum_{i\in U}\frac{t_i}{1+t_i\theta} - \sum_{i=1}^{n}\frac{t_i}{1+t_i\theta} \tag{2}$$

$$\ell''(\theta) = -\frac{d}{\theta^2} + \sum_{i\in U}\frac{t_i^2}{(1+t_i\theta)^2} + \sum_{i=1}^{n}\frac{t_i^2}{(1+t_i\theta)^2} \tag{3}$$

Let $\theta_0 = \tilde{\theta}$. For $k \geq 1$, the k-th step in Newton's method is

1, B

$$\theta_k = \theta_{k-1} - \frac{\ell'(\theta_{k-1})}{\ell''(\theta_{k-1})}.$$

The approximate distribution of $\hat{\theta}_{\text{NEWTON}}$ obtained by Newton's method is

$$\hat{\theta}_{\text{NEWTON}} \sim N(\theta, -\ell''(\hat{\theta}_{\text{NEWTON}})^{-1}).$$

2, C

2. (a) Find the cumulative intensity $\Lambda(t)$ of the counting process $N(t)$ The intensity for individual i is

seen ⇓

3, A

$$\lambda_i(t) = Y_i(t)\mu(t).$$

It follows immediately that $\lambda(t) = Y(t)\mu(t)$. Integrating,

$$\Lambda(t) = \int_0^t Y(s)\mu(s)ds.$$

meth seen ⇓

(b) Let $\mathcal{H}_{t-}$ denote the history of $N(t)$ up to, but not including time t. We have that

$$E[dD(t) \mid \mathcal{H}_{t-}] = E[dN(t) - d\Lambda(t) \mid \mathcal{H}_{t-}] = 0$$

since $d\Lambda(t) = E[dN(t) \mid \mathcal{H}_{t-}]$. Hence $D(t)$ is a martingale.

3, B

(c) We again note that $E\{dN(t) \mid \mathcal{H}_{t-}\} = d\Lambda(t)$. However,

meth seen ⇓

$$d\Lambda(t) = Y(t)dM(t).$$

Using this and substituting $dN(t)$ for its expectation, we have

2, A

$$dN(t) = Y(t)d\hat{M}(t).$$

Rearranging this and integrating with respect to t, we find

2, B

$$\hat{M}(t) = \int_0^t \frac{dN(s)}{Y(S)} = \sum_{T_j \le t} \frac{dN(T_j)}{Y(T_j)}$$

.

2, B

(d) The usual Kaplan-Meier estimator is

sim. seen ⇓

$$\hat{S}(t) = \prod_{j:T_{(j)} \leq t} \left[1 - \frac{d_j}{n_j}\right]$$

where d_j is the number of deaths occurring at time $t_{(j)}$ and n_j is the number of individuals at risk as we approach time $t_{(j)}$ from below.

3, A

Note that $d_j = dN(t_{(j)})$ and $n_j = Y(t_{(j)})$. Hence,

1, B

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left[1 - \frac{dN(t_{(j)})}{Y(t_{(j)})}\right].$$

unseen ⇓

(e) We first consider the case that $t > 0$ is not an event time. By construction, $\hat{S}(t)$ only places positive probability mass on the event times. Since $dN(t) = 0$ for any other t, the identity holds trivially.

2, D

If $t = t_{(j)}$ for some $1 \leq j \leq k$, the mass is

$$\begin{aligned}
\hat{S}(t_{(j)}-) - \hat{S}(t_{(j)}) &= \hat{S}(t_{(j-1)}) - \hat{S}(t_{(j)}) \\
&= \hat{S}(t_{(j-1)}) - \prod_{i=1}^{j} \left[1 - \frac{dN(t_{(i)})}{Y(t_{(i)})}\right] \\
&= \hat{S}(t_{(j-1)}) - \hat{S}(t_{(j-1)}) \left[1 - \frac{dN(t_{(j)})}{Y(t_{(j)})}\right] \\
&= \hat{S}(t_{(j-1)}) \frac{dN(t_{(j)})}{Y(t_{(j)})} \\
&= \hat{S}(t_{(j)}-) \frac{dN(t_{(j)})}{Y(t_{(j)})},
\end{aligned}$$

which verifies the identity.

2, D

3. (a) In both (a) and (b) we use

meth seen ⇓

$$\log\{\mu_i(t)/\mu_j(t)\} = \log\{\mu_i(t)/\mu_0(t)\} - \log\{\mu_j(t)/\mu_0(t)\}.$$

Other equivalent approaches are accepted. For individual i, we estimate that

$$\log\{\mu_i(t)/\mu_0(t)\} = (140)(\hat{\beta}_1) + (0)(\hat{\beta}_2) + (0)(120)(\hat{\beta}_3) = -14$$

and for individual j,

$$\log\{\mu_j(t)/\mu_0(t)\} = (120)(\hat{\beta}_1) + (0)(\hat{\beta}_2) + (0)(120)(\hat{\beta}_3) = -12.$$

Hence, we estimate

2, A

$$\log\{\mu_i(t)/\mu_j(t)\} = -14 - (-12) = -2$$

which means that $\mu_i(t) < \mu_j(t)$. Hence, individual j has a higher instantaneous risk of death.

1, A

(b) For individual i, we estimate that

meth seen ⇓

$$\log\{\mu_i(t)/\mu_0(t)\} = (120)(-0.1) + (1)(-2.5) + (1)(120)(0.05) = -8.5$$

and for individual j,

$$\log\{\mu_j(t)/\mu_0(t)\} = (120)(-0.1) + (0)(-2.5) + (0)(120)(0.05) = -12.$$

Hence, we estimate

2, A

$$\log\{\mu_i(t)/\mu_j(t)\} = -8.5 - (-12) = 3.5$$

which means that $\mu_i(t) > \mu_j(t)$. Hence, individual i has a higher instantaneous risk of death.

1, A

(c) In (a), the log hazard ratio estimated is equivalent to $20\hat{\beta}_1$. We can construct an approximate 95% confidence interval with limits given by

unseen ⇓

3, C

$$20(-0.1) \mp 2(20)(0..04) \Rightarrow -2.0 \mp 1.6 \Rightarrow (-3.6, -0.4).$$

In (b), the log hazard ratio estimated is equivalent to $\hat{\beta}_2 + 120\hat{\beta}_3$. It is not possible to construct a 95% confidence interval in this case, since the standard error of this estimator depends on the covariance between $\hat{\beta}_2$ and $\hat{\beta}_3$. An estimate of this could be obtained from, e.g, an estimate of the inverse information matrix.

2, C

1, D

(d) The Wald test statistic is

meth seen ⇓

2, B

$$Z = \frac{\hat{\beta}_3 - 0}{s.e.(\hat{\beta}_3)} = \frac{0.05 - 0}{0.02} = 2.5$$

Since $Z = 2.5 > 2$, we reject H_0 at the 0.05 level.

2, B

(e) The baseline group of individuals corresponds to non-smokers with systolic blood pressure of 0 mm Hg. If we are interested in a population of living humans, a blood pressure of 0 is not scientifically plausible. To make $\mu_0(t)$ more interpretable, we could consider transforming blood pressure by subtracting a reference value from each measurement (e.g. 90 mm Hg, the sample mean, etc...).

unseen ⇓

2, B

2, D

4. (a) The Kolmogorov forward equations are

meth seen ⇓

4, A

$$\frac{d}{dt}p^{ij}(t) = \sum_{k\neq j} p^{ik}(t)\mu_{kj} - p^{ij}(t)\mu_{jk}.$$

In the present case, we know that transitions from state 3 to states other than 4 are not possible. Hence, $p^{31}(t) = p^{32}(t) = 0$.

Since the rows of the transition probability matrix must add to one, we have

3, A

$$p^{33}(t) = 1 - p^{34}(t).$$

Hence, it suffices to consider the forward equation

1, B

$$\frac{d}{dt}p^{34}(t) = \sum_{k\neq 4} p^{3k}(t)\mu_{k4} - p^{34}(t)\mu_{4k} = \{1 - p^{34}(t)\}\mu_{34}.$$

This is of the form of the provided result with $a = -\mu_{34}$, $b = \mu_{34}$, and $y = p^{34}(t)$. Hence, for some constant C,

$$p^{34}(t) = Ce^{-\mu_{34}t} - (-\mu_{34}/\mu_{34}) = 1 + Ce^{-\mu_{34}t}.$$

Using the initial condition that $p^{34}(0) = 0$, we find $C = -1$. Therefore,

2, C

$$p^{34}(t) = 1 - e^{-\mu_{34}t}.$$

In total, the third row of the transition probability matrix is

$$[0 \quad 0 \quad e^{-\mu_{34}t} \quad 1 - e^{-\mu_{34}t}].$$

unseen ⇓

(b) (i) The odds in question are the ratio of the jump probabilities

$$\frac{\mu_{12}/(\mu_{12}+\mu_{14})}{\mu_{14}/(\mu_{12}+\mu_{14})} = \frac{\mu_{12}}{\mu_{14}}.$$

Note that the maximum likelihood estimator for any μ_{ij} is $\hat{\mu_{ij}} = D^{ij}/V^i$. Hence, a (maximum likelihood) estimator for the odds in question is

$$\frac{\hat{\mu}_{12}}{\hat{\mu}_{14}} = \frac{D^{12}/V^1}{D^{14}/V^1} = \frac{D^{12}}{D^{14}}.$$

Based on the data, we estimate these odds to be $21/3 = 7$.

3, A

(ii) We can directly apply the goodness of fit test from Chapter 7 of the lecture notes, noting that there is only one "age group" in this setting.

1, C

The test statistic is

$$X^2 = \frac{\{D^{12} - V^1(1/3)\}^2}{V^1(1/3)} = \frac{(21-28)^2}{28} = 7/4 = 1.75.$$

Hence, we do not reject the null hypothesis at the 0.05 level.

2, C

(c) Perhaps the simplest modification is to assume that the Markov model holds piecewise within individual months (students may specify a different time unit).

$\boxed{2, \text{D}}$

This allows the transition intensities to vary depending on the length of time since the initial graft. To test the hypothesis that the intensities do not vary in time, a likelihood ratio test could be implemented. This would involve comparing the maximised likelihoods under the homogeneous and piecewise homogeneous models.

$\boxed{2, \text{D}}$

Note: any proposal that results in nested models will be similar. Proposing a valid test for non-nested models can be done, but would likely involve materials outside the scope of the module.

5. (a) We note that there are no censoring events in either group up to time $t^* = 140$, so we can simply take $1-$(Deaths up to Time t)$/9$ for the Kaplan-Meier (KM) estimates (since 9 individuals were at risk to begin with in each group):

TX group

Time	At Risk	Deaths	KM
10	9	1	8/9
100	8	3	5/9
140	5	1	4/9
160	4	2	2/9
180	2	1	1/9

Placebo (PL) group

Time	At Risk	Deaths	KM
20	9	2	7/9
40	7	1	6/9
50	6	3	3/9
110	3	1	2/9
130	2	1	1/9
160	1	1	0

The estimate RMST in the TX group is found by computing the area under the KM estimate up to time $t^* = 140$. This amounts to adding the area of rectangular regions corresponding to the time intervals [0,10), [10,100), and [100,140): 3

$$1 \times 10 + \frac{8}{9} \times 90 + \frac{5}{9} \times 40 = \frac{1010}{9} \approx 112.2$$

The estimate RMST in the PL group is found by computing the area under the KM estimate up to time $t^* = 140$. This amounts to adding the area of rectangular regions corresponding to the time intervals [0,20), [20,40), [40,50), [50,110), [110,130), and [130,140): 3

$$1 \times 20 + \frac{7}{9} \times 20 + \frac{6}{9} \times 10 + \frac{3}{9} \times 60 + \frac{2}{9} \times 20 + \frac{1}{9} \times 10 = \frac{610}{9} \approx 67.8$$

Hence $\hat{\mu}_{TX} - \hat{\mu}_{PL} = (1010 - 610)/9 = 400/9 \approx 44.4$. 1

Based on this the TX group has better survival outcomes, surviving longer on average over the time period $[0, 140)$. 1

(b) Representative discussions for each part of the question follow:

On the RMST: Regardless of how RMST is estimated, it necessarily makes use of the full data available up to time t^* in order to construct estimates of the survival curves. Additional data on survival beyond time t^* typically improves estimation of the survival curve when using parametric methods. Due to the definition of RMST changing the time horizon of interest, t^*, may lead to different treatment effects under non-proportional hazards. 3

On the log HR: As we have noted, the use of the log HR makes the assumption that the HR is constant across the full time period. Use of the Cox model leads to estimates that make use only of the relative ordering of the death times (hence, depending on the history of survival). Under the Cox model, the log HR is assumed to remain constant, so the same treatment effect is expected regardless of how long individuals are followed on the study. 3

(c) Representative discussions for each part of the question follow:

The **piecewise exponential model** is desirable at the design stage of the analysis because it allows specification of time varying hazard ratios across prespecified time intervals. However, the piecewise nature of this model means that estimates based on it will not be smooth functions (which are expected to better reflect underlying survival patterns). 2

Flexible, smooth, parametric models avoid the discontinuities in the hazard function inherent to the piecewise exponential model. This is desirable for estimation based on data from a trial, but it is no longer obvious how to specify curves a priori at the design stage of the trial given the extreme flexibility of, e.g., spline models. [2]

Nonparametric models face similar difficulties at the design stage of a trial in that there is no straightforward approach to specifying survival patterns for each group. For estimation and inference, nonparametric models are viable as an alternative to flexible parametric methods if individuals are observed in both groups to at least time t^*. Unlike a parametric approach, the Kaplan-Meier estimates cannot be used to extrapolate beyond the observed data. [2]

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

Please record below, some brief but non-trivial comments for students about how well (or otherwise) the questions were answered. For example, you may wish to comment on common errors and misconceptions, or areas where students have done well. These comments should note any errors in and corrections to the paper. These comments will be made available to students via the MathsCentral Blackboard site and should not contain any information which identifies individual candidates. Any comments which should be kept confidential should be included as confidential comments for the Exam Board and Externals. If you would like to add formulas, please include a sperate pdf file with your email.

ExamModuleCode	QuestionNumber	Comments for Students
MATH96048 MATH97075 MATH97185	1	This question was well answered in most instances. Parts (a,c,d) were nearly always done correctly. In part (b), some students were not able to clearly demonstrate why the non-constant hazard ratio reconciles the differences (responses such as "the results are consistent because of nonproportional hazards" were inadequate for full marks). In part (e), partial credit was awarded for students that discussed general properties of the Newton algorithm. Full marks were awarded only to students that derived the update for the parametric distribution in this question.
MATH96048 MATH97075 MATH97185	2	This question was addressed reasonably well by most candidates. Parts (a) and (b) were generally done well, with few issues. The derivations in part (c) were mostly done well, however a common issue was a failure to state either that dN(t) is substituted for its expectation, or that the resulting estimator was a moment-based estimator; a statement to this effect was required for full marks. Part d) went well, though several candidates lost marks through not providing sufficient working and explanation. Part e) challenged most candidates: in order to obtain full marks, candidates were required to clearly consider both the cases where t is and is not a death time; few candidates did this.

MATH96048 MATH97075 MATH97185	3	This question was tackled well by most students. Parts (a) and (b) were done well with few issues. In part (c), most students sought the correct CI for the log hazard ratio in (a); a reasonable number of students correctly recognised that the CI for the log hazard ratio in (b) is unavailable as we don't have the required covariance information. Very few students obtained full marks in part (c), however; to do so, candidates could have briefly suggested how to obtain a suitable covariance estimate. Part (d) was generally done well, although a handful of students failed to address this question using a Wald test statistic. Part (e) was also done well by most, however only a handful of candidates obtained full marks by correctly describing the required changes to the proportional hazards model, as well as the interpretation of the altered baseline group.
MATH96048 MATH97075 MATH97185	4	While a large number of students tackled this problem well, there were many students that had some difficulty. In part (a), many students had difficulty solving the differential equations, sometimes using incorrect initial conditions for the first two transition probabilities. Part (b) was well answered, though some students conflated odds and probabilities or did not use the goodness of fit statistic we defined in terms of the number of deaths. Partial marks in part (c) were awarded to a number of students that did not adequately describe the modification to the model or the hypothesis test they would use.
MATH96048 MATH97075 MATH97185	5	More students had difficulty with the mastery material question. Most students did well on part (a) of the question, which asked them to derive an estimate of the difference in restricted mean survival time using the Kaplan-Meier survival curves. Students primarily lost points only if they did not integrate correctly or did not follow the instructions (attempting to fit the piecewise model, which was not asked for). In parts (b,c), students often addressed only part of the prompt, neglected to include specific details or did not provide a balanced critique of the methods in question.