(19) **Task 1**

(2) **A** **Generalized Type I right censoring**, since individuals enter the study at different time points and the end of the study is predetermined.

(Depending on how precise the time of death is registered, interval censoring could also be present – if we only know that the patient died between two visits and not exactly when this happened. Not that likely, though.)

Left truncation, since only individuals with a diagnosis are included and there might be individuals with oseophageal cancer that aren't diagnosed which we would have wanted to include in the study if we had known about them.

(1) **B** To control for any effect caused by these other variables (confounders), i.e. to be able to focus on the marginal effect of volume separated from the effect of the other variables.

(2) **C** Yes, "postoperative mortality" indicates that there is a risk of dying due to the surgery, which would be a competing risk. The fact that the event is defined as "death from any cause" only tells us how the researchers have chosen to handle this, not how they *should* have.

The estimated survival will be biased (underestimated).

(4) **D** The hazard ratio for volume is below 1 for all volume levels, compared to the reference 1-10 surgeries/year. This means that for hospitals with higher volume levels, the risk of dying is on average lower than for hospitals with the lowest volume.

The two highest volumes are significant (CI doesn't include 1).

Generally, the larger the volume, the lower the risk of dying, and the test of trend is significant,

(2) **E** Relative risk = $\exp(\beta_{>75} - \beta_{60\text{-}75}) = \exp(\beta_{>75})/\exp(\beta_{60\text{-}75}) = 1.78/1.40 = 1.3357$

This means that the risk of death for individuals over 75 years of age is approximately 34% larger than the risk of death for individuals 60-75 years of age (all other covariates held constant).

(4) **F** *i*. **No**. The hazard ratio decreases with approximately 8-10% for each higher volume, except for the last one. The hazard ratio then *increases* for the highest volume, which indicates that the effect is not linear.

ii. Each increase in volume, except for the highest volume, decreases the risk of death by approximately 8-10%. If this would have been the case for also the highest volume, the hazard ratio of hospital volume as a continuous covariate would be approximately 0.91 (on average a 9% risk reduction if the volume increases by 10 surgeries/year). But since the highest volume contradicts this pattern, the risk doesn't reduce by as much on average, and the hazard ratio should be closer to 1. The hazard ratio must thus be **0.95**.

(4) **G** Centralizing surgery is a large undertaking, and we as statisticians have to do our very best to ensure that the analyses used for this kind of decision-making are ethically sound. There are many aspects from the guidelines that can be referred to here, some examples of important points:

- Pursue objectivity
- Avoid preempted outcomes
- Protect the interest of the subjects
- Strive to produce results that reflect the observed phenomena

(Write a bit about these points to get full score)

(23) **Task 2**

A (8) Age ≤30 years:
25th percentile = 3 months. This means that 25% of patients 30 years of age or younger have had a remission within (at or before) 3 months from the study start.

Median = 8 months. This means that 50% of these patients have had a remission within 8 months from the study start.

75th percentile= 15 months. This means that 75% of these patients have had a remission within 15 months from the study start.

Age >30 years:
25th percentile = 18 months. This means that 25% of patients over 30 years of age have had a remission within (at or before) 18 months from the study start.

Median = at least 34 months. The median cannot be estimated since less than 50% of these patients have had a remission during the study. The median must however be at least 34 months, which is the longest observed time in this group.

75th percentile = at least 34 months. The 75th percentile cannot be estimated either, since less than 75% of these patients have had a remission during the study. The 75th percentile must also be at least 34 months, which is the longest observed time in this group.

These measures all have larger values for the patients over 30 years than for the younger patients, which means that the risk of remission is lower ("survival" is larger) for the older patients.

B (3) 1-year probability of not experiencing a remission, i.e. P(*time* > 12):
Age ≤ 30 years: **30.97%**
Age > 30 years: **79.28 %**

The 1-year probability is a lot larger for patients over 30 years of age, more than twice as large as for the younger patients.

C (12) H_0: $h_{\leq 30}(t) = h_{>30}(t)$ for all $t \leq \tau$
H_1: $h_{\leq 30}(t) \neq h_{>30}(t)$ for some $t \leq \tau$

Where τ = largest time at which both groups have at least one subject at risk.

Test: Log-rank or Gehan's/Wilcoxon test. Motivation needed.

Assumptions:

- Random sample – not specified in the task
- Independent samples – OK (reasonable to assume that the times to depression in the two different groups are independent)

- Non-informative censoring (i.e. that the censoring times and event times are independent) – not specified in the task (should be discussed with the clinician/psychiatrist), but the censoring plot doesn't contradict this since the censoring pattern is similar in both groups.
- Right censored data – OK (reasonable to assume from the given information)
- Survival probabilities are the same for subjects recruited early and late in the study - not specified in the task (should be discussed with the clinician/psychiatrist), but reasonable to assume
- Large samples (both tests are based on large-sample approximations to the distribution of the chi-square statistics) – OK, 54 events in the smallest group.

Choice of significance level:
Wrongly rejecting the null hypothesis here would mean that we claim that there is a difference in time to remission between the two age groups, when in fact there is no difference. No decisions will be made based on the results, difficult to see that there could be any severe consequences of committing a Type I error. The standard 5% is fine to use.

Result:
P-value < 0.0001 (for both options).
The P-value is smaller than (any) α, thus H_0 is rejected.

Conclusion:
The test suggests that there is a significant difference in time to remission between the two age groups in the investigated population of depression patients.

NOTE! The generalization to the population is only valid if the sample is random.

(36) **Task 3**

A (2) Parametric regression can be used when we know the probability distribution for time to event. Nothing is said about previous knowledge here, which means that Cox regression is more suitable. (One option could be to estimate the distribution from the data, but if the wrong distribution is chosen, and the model doesn't fit the data well, the results can be misleading.)

B (4) The two variables *gender* and *civil status* have to be recoded to 0/1 variables, or denoted as "class" variables in proc phreg.

The output shows that there is one single individual in the third gender category, which means that there is too little information for this category to be used. One of the 'female' variable options proposed on page 12 could be used instead.

The Martingale plot suggests that age can be used as a continuous covariate.

C (18) Any chosen model would have to fulfill the **assumptions** for the Cox model. Many of them are already discussed in Task 2 (OK to refer to that):
• Random sample (for inference to be correct). No information in task.
• Non-informative censoring. Reasonable to assume.
• Right-censored or left truncated data. Right censoring can be seen from task (observation period 30 months).

New for this task:
• Large sample (common rule of thumb: ≥ 10 events per covariate). We have a total of 754 observations, and 486 events. This is sufficient for any model you've chosen.
• Proportional hazards (to be checked when building the model)

Check of **proportionality (PH) assumption**:

1) include time-dependent covariate in model
To test the assumption of proportional hazards you can include the time-dependent covariate ln(t)*covariate in the model (if significant, the PH assumption is rejected)

H_0: The hazards for different values of covariate i are proportional
(all i=1 to p covariates are to be examined)
H_a: The hazards are not proportional

Significance level $\alpha = 5\%$ fine to use (no serious consequences if we claim that the hazards are not proportional when they in fact are)

The test above is rejected for the *age* and *civil status* covariates, but fine (non-significant) for *trt* and *female*.
(Significant also for *gender*, but we've already stated to use *female* instead.)

2) Score plots / Arjas plots
You should always make use of a graphical method in addition to the test above. Score and Arjas plots have been provided, which both show that the PH assumption looks okay for *trt* and *female*, but not for *age* and *civil status*.

This concludes that the PH assumption doesn't hold for *age* and *civil status*.

Comparison of AIC values
Model 5 has the lowest AIC value, which suggests that this model is to be used.

Test of equality of strata for Model 5 (or Model 3):
For Model 5 (or Model 3) to be valid, we need to check that it is reasonable to assume that the regression coefficients are the same in each stratum.

H_0: All β's are the same for all s strata
H_a: At least one of the β's is/are different

This can be tested, using the Likelihood ratio test.

Significance level α=5% fine to use (no serious consequences if we claim that the covariates are not the same when they in fact are, then we estimate separate models instead)

Test statistic:

$$-2\left[LL(\mathbf{b}) - \sum_{j=1}^{s} LL_j(\mathbf{b}_j)\right] \sim \chi^2_{(s-1)p}$$

where s= no. of strata and p=no. of covariates

$LL(\mathbf{b})$ = -2889.152/2 = -1444.576
$LL(\mathbf{b})_{\text{Married}}$ = -822.234/2
$LL(\mathbf{b})_{\text{Single}}$ = -1459.118/2
$LL(\mathbf{b})_{\text{Divorced}}$ = -593.481/2

Sum $LL_j(\mathbf{b}_j)$ = -2874.833/2 = -1437.4165

Test statistic = -2(-2889.152/2 + 2874.833/2) = -2889.152 + 2874.833 = 14.319
df = (3-1)*4 = 8 df

According to Table c.2 the corresponding p-value is larger than 0.05, which means that the null hypothesis is not rejected (reject if χ^2_{test} larger than χ^2_{crit} = 15.50731).

Conclusion:
It is okay to assume that the regression coefficients are the same in each of the two strata and the stratified model can be used.

<u>Cox-Snell plots</u>
All models provide Cox-snell plots that suggest that the model fits the data.

<u>Choice of model:</u>
<u>All of the above suggests that **Model 5** is a good choice.</u>

D (10) All covariates but *female* are significant at the 5% level. The marginal effects of the covariates are presented below, i.e. the effect of each covariate holding the other covariates constant.

Individuals who received the new treatment have a 22.8% lower risk of depression on average, compared to individuals who received the standard care (hazard ratio 0.772, 95% confidence interval 0.596 to 0.987).

Women have a 5.5% lower risk of depression on average, compared to men (hazard ratio 0.945, 95% confidence interval 0.789 to 1.133). This is quite close to 1 and also not significant.

Age is time-dependent, which means that the hazard ratio is not constant over time.

Hazard ratio for a one unit increase in Age:

At 3 months: hazard ratio = exp(-0.05773-0.01236*ln(3)) = 0.931

At 12 months: hazard ratio = exp(-0.05773-0.01236*ln(12)) = 0.915

At 25 months: hazard ratio = exp(-0.05773-0.01236*ln(25)) = 0.907

For every year older the patient is, the risk of depression reduces by approximately 7% at 3 months, by approximately 8.5% at 12 months, and by approximately 9.3% at 25 months. In other words, the longer time in remission the larger the risk reduction by older age.

Civil status cannot be interpreted in terms of hazard ratios, since the model is stratified on civil status, but the estimated survival plot shows that the risk of depression is substantially lower (the "survival" is higher) for patients who are married or in an equivalent partnership than for single patients or divorced/separated/widowed patients.

F (2) Generalized $R^2 = 1 - e^{-(LRT/n)}$

where $LRT = -2\log L(0)-[-2\log L(p)] = 2999.934 - 2889.152 = 110.782$

$R^2 = 1 - \exp(-110.782/754) = 1 - \exp(-0.1469) = 1 - 0.863 = 0.137$

According to the generalized R^2 the model shows at least some association between the covariates and time to depression remisison, but the association is not very strong.