(27) **Task 1**

A (2) Event of interest: death without relapse
Competing risk: death with relapse

The Kaplan-Meier method assumes that all events are independent, and thus, censors events other than the event of interest. When there are competing risks the KM estimates are biased, and therefore the **Cumulative incidence function** is the preferred choice.

B (8) Based on the Cumulative incidence function the percentiles of event times are presented below (for survival times, switch places of 25th and 75th percentiles).

cGVHD group:
25th percentile=around 80-85 months (estimated from the plot). This means that 25% of the cGVHD patients have died without first getting a relapse before 80-85 months from transplant.

Median: at least 150 months. The median cannot be estimated since less than 50% of the cGVHD patients have died without relapse during the study. The median survival time must however be at least 150 months, which is the longest observed survival time.

75th percentile: at least 150 months. Same as for the median above.

Group without cGVHD:
25th percentile: at least 150 months. Not even the 25% percentile can be estimated since less than 20% of the patients without cGVHD have died without relapse during the study. The 25th percentile must however be at least 150 months, which is the longest observed survival time.

Median: at least 150 months. Same as above.

75th percentile: at least 150 months. Same as above.
The 25th percentile is the only estimable measure, for the cGVHD group, and therefore the only measure that can be compared. The estimated survival time is much longer for the group without cGVHD (>150 months vs 80-85 months). The chronic Graft-versus-host-disease seems to have a negative impact on the risk of dying without first getting a relapse.

C (4) 50- and 100-month probabilities of non-relapse death estimated from plot:

cGVHD group:
50 months: ~18%
100 months: ~34%

Group without cGVHD:
50 months: ~18%
100 months: ~18%

The probability of non-relapse death is about the same for the two groups at 50 months post transplant. After 100 months there is a relatively large difference, where the risk of non-relapse death is almost twice as high for the cGVHD group (34% vs 18%).

D (13)

$$H_0 : CI_{cGVHD}(t) = CI_{no\_cGVHD}(t) \text{ for all } t \le \tau$$
$$H_a : CI_{cGVHD}(t) \ne CI_{no\_cGVHD}(t) \text{ for some } t \le \tau$$

where CI= Cumulative Incidence function, and t = largest time at which all of the groups have at least one subject at risk.

Test: Gray's test for equality of cumulative incidence functions (the only test for comparison of Cumulative incidence functions that we've learned about).

Assumptions:
- Random samples – not specified in the task
- Independent samples – OK (reasonable to assume that the survival times in the two different groups are independent)
- Right censored data – OK (reasonable to assume from the given information)
- Time to event and event type independent and identically distributed within each group – not specified in the task (should be discussed with the clinician), but the censoring plot doesn't contradict this since the censoring pattern is similar in both groups.
- Large samples – OK (39/65 events respectively)

Choice of significance level:
Wrongly rejecting the null hypothesis here would mean that we claim that there is a difference in non-relapse survival time between the two groups, when in fact there is no difference. The consequences of this could be that the cGVHD patients will receive special care (with increasing costs), but this would not be of any harm for the patient. a=5% is fine to use.

Result:
P-value = 0.0097.
The P-value is smaller than (any chosen) α, thus H_0 is rejected.

Conclusion:
The test suggests that there is a significant difference in non-relapse survival between patients with and without cGVHD in the population of ALL patients.

NOTE! This conclusion is only valid if the sample is random.

(43) **Task 2**

A (2) Time is measured in weeks. The ties are probably a result of imprecise measurement (the times could be measured more precise), which means that the **exact** method of handling would be the best choice.

B (2) The **profile-likelihood** method is appropriate for all likelihood-based statistical analyses (the other alternative, the Wald method, may work poorly for maximum likelihood estimation).

C (2) **Option *iii)*** is the only option that correctly handles the variable since it is a time dependent covariate.

D (22) Any chosen model would have to fulfill the **assumptions** for the Cox model:
• Random sample (for inference to be correct). No information in task.
• Non-informative/random censoring. Reasonable to assume.
• Right-censored or left truncated data. Right censoring can be seen from task (observation period 1 year).
• Large sample (common rule of thumb: ≥ 10 events per covariate). We have a total of 432 observations, but the no. of events is not stated! There are a maximum of 7 covariates, depending on which model you've chosen. Model no. 4, option *iii*) has 4 covariates, which means that we only need 40 events – reasonable to assume that we have many more than this.
• 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 a=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* covariate, also for *age30*, but fine (non-significante) for the rest of the covariates.

2) Score plots
You should always make use of a graphical method in addition to the test above. Score plots have been provided, which show that the PH assumption looks okay for for all covariates but age (OK for age23 and age30).

This concludes that age23 is the only age covariate where the PH assumption holds.

Comparison of AIC values
Model 4 has the lowest AIC value, which suggests that this model is to be used.

Martingale plot
Suggests a cut-off for age around 23 years, which supports Model 4.

Test of equality of strata for Model 4:
For model 4 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

$LL(\mathbf{b})$ = -1090.100/2 = -545.05

$LL(\mathbf{b})_{cGVHD}$ = -659.964/2 = -329.982
$LL(\mathbf{b})_{not\_cGVHD}$ = -424.324/2 = -212.162
Sum $LL_j(\mathbf{b}_j)$ = -542.144

Test statistic = -2(-545.05 + 542.144) = 5.812
df = (2-1)*4 = 4 df

According to Table c.2 the corresponding p-value is larger than 0.1, which means that the null hypothesis is not rejected (reject if χ^2_{test} larger than χ^2_{crit} = 9.48773).

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.

Cox-Snell plots
All models provide Cox-snell plots that suggest that the model fits the data.

Choice of model:
All of the above suggests that **Model 4** is a good choice.

Model 2 or Model 5 could also be argued for, depending on how the results of the model are to be used (interpreted/explained to non-statisticians). If it e.g. is important to get an estimated hazard ratio for age/age23 then Model 2 would be a better choice. If dichotomization is to be avoided (which is recommended) then Model 5 is a better choice, even though it takes some effort to interpret the effect of age.

E (12) All covariates but *married* 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 financial aid have a 32.6% lower risk of rearrest on average, compared to individuals who didn't receive financial aid (hazard ratio 0.674, 95% confidence interval 0.463 to 0.976).

Individuals who are married have a 35.1% lower risk of rearrest on average, compared to individuals who are not married (hazard ratio 0.649, 95% confidence interval 0.286 to 1.278). This is however not significant, which can be seen by the relatively wide confidence interval that covers 1.

The risk of rearrest increases with 9.7% on average for every extra prior conviction (hazard ratio 1.097, 95% confidence interval 1.037 to 1.155).

Whether an individual manages to get a job or not affects the risk of rearrest positively. At a given time the risk of rearrest is 37% lower on average for individuals who has got a job at or before that time, compared to individuals who hasn't got a job (hazard ratio 0.630, 95% confidence interval 0.416 to 0.969).

Age cannot be interpreted in terms of hazard ratios, since the model is stratified on age23, but the estimated survival plot shows that the risk of rearrest is higher for individuals younger than 23 years of age, compared to those 23 years or older, when setting the other covariates to their reference or mean values. The survival, i.e. time to rearrest, drops faster for the younger age group.

F (3) The relative risk of rearrest for individuals with 10 more prior convictions than others:

$1.097^{10} \approx 2.52$
where 1.097 is the estimated hazard ratio for the covariate prior convictions.

This means that the risk of rearrest is 2.52 times higher (or increases by 152%) for individuals with 10 more prior convictions than others, holding the other covariates constant.

(4) **Task 3**

It is always a good thing to explain that you as a statistician are following the declaration on professional ethics for statisticians. Among other things it is stated there that you as a statistician should pursue objectivity, and strive to produce results that reflect the observed phenomena in an impartial manner.

You have to explain to your employer that you as a statistician are responsible for this, which means that you should use the model that best describes the data, not the model that best fits the purpose of your employer. Since there is only one model yielding a hazard ratio above 1, and several models that show the opposite, the single model your employer wants to use is not trustworthy.

(6) **Task 4**

A (3) In the Cox model, you can make special assumptions about the baseline hazard which leads to parametric models, e.g. the exponential and Weibull distributions.

But these assumptions can also be avoided, you still get estimates of hazard ratios (relative risks). A parametric form is assumed only for the covariate effect. This is the reason why it's said to be semi-parametric.

B (3) Sometimes you might have knowledge about the distribution of survival and censoring times, e.g. in a manufacturing process. In these situations it is better to use parametric models, since they will produce more precise estimates. But if we assume the wrong parametric model, we will get consistently wrong estimates.