**qwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnm**

|  |
| --- |
| Examination Project  Survival Analysis  15338673  Paul-Willem Janse van Rensburg |

Table of Contents

[Part A 1](#_Toc23856978)

[Question 1 1](#_Toc23856979)

[Question 2 2](#_Toc23856980)

[Question 3 6](#_Toc23856981)

[Question 4 7](#_Toc23856982)

[Question 5 8](#_Toc23856983)

[Part B 9](#_Toc23856984)

[Question 1 9](#_Toc23856985)

[Question 2 9](#_Toc23856986)

[Part C 10](#_Toc23856987)

[Question 1 10](#_Toc23856988)

[Question 2 10](#_Toc23856989)

[Question 3 10](#_Toc23856990)

[Question 4 10](#_Toc23856991)

[Part D 11](#_Toc23856992)

# Part A

A medical collaborator is interested in how time to re-infection depends on the type of initial infection.

## Question 1

Consider the three types of initial infections and compute a survival curve of the time to re-infection. Plot them on the same graph. What does the graph suggest?

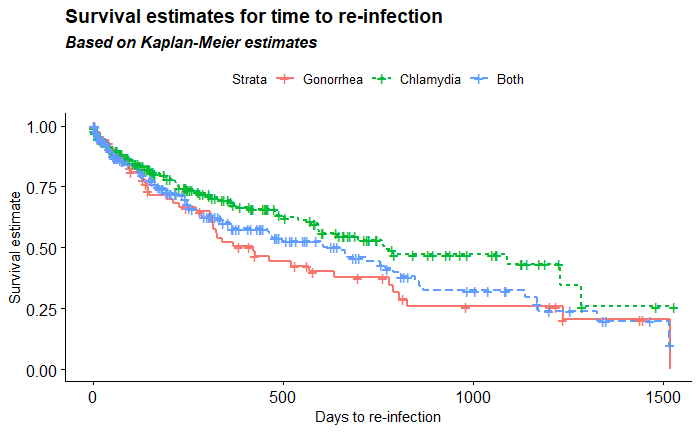


Figure 1 - Kaplan-Meier Estimates

We implement the Kaplan-Meier survival estimate as the majority of the groups have their final

From Figure 1 we can see that Chlamydia has an overall higher survival rate in terms of re-infection than does Gonorrhea. A person who has been infected with both appears to lie in between the two singular infections.

This essentially tells us that your chance of re-infection is lower if you’ve initially been infected with Chlamydia than if you had been infected with both Chlamydia and Gonorrhea or just Gonorrhea. The suspicion is that Gonorrhea pulls down the survival rate for initial infection of both, seeing as it standing alone has the lowest survival rate.

## Question 2

Obtain an appropriate estimator and confidence interval for the 3 quartiles of the survival curves for the three types of initial infections. Interpret the results.

Table 1 - Survival Estimates for Gonorrhea

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| obs | t | nrisk | nevent | ncensored | |  | | σ | | upper | | lower | | strata | |
| 4 | 5 | 73 | 1 | 0 | 0.986301 | | 0.013793 | | 1 | | 0.959994 | | 1 | |
| 5 | 9 | 72 | 1 | 0 | 0.972603 | | 0.019644 | | 1 | | 0.935868 | | 1 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 24 | 126 | 51 | 1 | 0 | 0.779978 | | 0.06458 | | 0.885223 | | 0.687245 | | 1 | |
| 26 | 136 | 49 | 1 | 0 | 0.76406 | | 0.067792 | | 0.872633 | | 0.668995 | | 1 | |
| 28 | 143 | 47 | 1 | 0 | 0.747803 | | 0.071121 | | 0.859658 | | 0.650502 | | 1 | |
| 29 | 144 | 46 | 1 | 0 | 0.731547 | | 0.07444 | | 0.846458 | | 0.632235 | | 1 | |
| 31 | 146 | 44 | 1 | 0 | 0.71492 | | 0.077909 | | 0.832864 | | 0.613679 | | 1 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 44 | 338 | 30 | 1 | 0 | 0.522395 | | 0.123089 | | 0.664925 | | 0.410417 | | 1 | |
| 45 | 367 | 29 | 1 | 0 | 0.504381 | | 0.127994 | | 0.648198 | | 0.392474 | | 1 | |
| 48 | 420 | 26 | 1 | 0 | 0.484982 | | 0.133869 | | 0.630486 | | 0.373058 | | 1 | |
| 49 | 426 | 25 | 1 | 1 | 0.465583 | | 0.139956 | | 0.61253 | | 0.353889 | | 1 | |
| 50 | 464 | 23 | 1 | 0 | 0.44534 | | 0.146846 | | 0.593865 | | 0.333961 | | 1 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 61 | 805 | 11 | 1 | 0 | 0.289529 | | 0.232895 | | 0.457018 | | 0.183422 | | 1 | |
| 63 | 827 | 9 | 1 | 0 | 0.257359 | | 0.261015 | | 0.429256 | | 0.154299 | | 1 | |
| 67 | 1238 | 5 | 1 | 1 | 0.205888 | | 0.343699 | | 0.403819 | | 0.104972 | | 1 | |
| 70 | 1519 | 1 | 1 | 0 | 0 | | Inf | | NA | | NA | | 1 | |

We highlight the largest smaller than 0.75, 0.5 and 0.25 for Gonorrhea. The upper and lower confidence interval for the survival estimates are displayed in Table 1. We then estimate the time to re-infection quartiles as below:

We calculate a 95% confidence interval using the below equation:

**[i]**

Where is the quartile estimate for re-infection, around the aforementioned quartile, with the standard error as retrieved from the table and with a , with the following result (α = 0.05):

Table 2 - Survival Estimates for Chlamydia

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| obs | t | nrisk | nevent | ncensored | |  | | σ | | upper | | lower | | strata | |
| 71 | 1 | 215 | 1 | 1 | 0.995349 | | 0.004662 | | 1 | | 0.986295 | | 2 | |
| 72 | 2 | 213 | 1 | 1 | 0.990676 | | 0.006624 | | 1 | | 0.977897 | | 2 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 153 | 218 | 103 | 1 | 0 | 0.761108 | | 0.043293 | | 0.828509 | | 0.699191 | | 2 | |
| 154 | 221 | 102 | 1 | 0 | 0.753647 | | 0.044399 | | 0.822168 | | 0.690836 | | 2 | |
| 155 | 223 | 101 | 1 | 0 | 0.746185 | | 0.045501 | | 0.815787 | | 0.682521 | | 2 | |
| 160 | 247 | 96 | 1 | 1 | 0.738412 | | 0.04669 | | 0.809173 | | 0.673839 | | 2 | |
| 164 | 264 | 91 | 1 | 0 | 0.730298 | | 0.04798 | | 0.802307 | | 0.664752 | | 2 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 217 | 698 | 34 | 1 | 0 | 0.530748 | | 0.091675 | | 0.635218 | | 0.44346 | | 2 | |
| 224 | 761 | 27 | 1 | 0 | 0.511091 | | 0.099141 | | 0.620707 | | 0.420832 | | 2 | |
| 225 | 766 | 26 | 1 | 0 | 0.491433 | | 0.106618 | | 0.605645 | | 0.39876 | | 2 | |
| 227 | 790 | 24 | 1 | 0 | 0.470957 | | 0.1148 | | 0.589792 | | 0.376066 | | 2 | |
| 239 | 1090 | 12 | 1 | 0 | 0.431711 | | 0.144065 | | 0.57256 | | 0.32551 | | 2 | |
| 246 | 1230 | 5 | 1 | 0 | 0.345368 | | 0.265998 | | 0.581702 | | 0.205053 | | 2 | |
| 247 | 1284 | 4 | 1 | 0 | 0.259026 | | 0.39254 | | 0.559081 | | 0.120009 | | 2 | |

As before, we highlight the largest smaller than 0.75, 0.5 and 0.25 for Chlamydia. The upper and lower confidence interval for the survival estimate for Chlamydia is displayed in Table 2. We then estimate the time to re-infection quartiles as below:

We calculate a 95% confidence interval using equation **[i]**, with the following result (α = 0.05):

Table 3 - Survival Estimates for Chlamydia and Gonorrhea

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| obs | t | nrisk | nevent | ncensored | |  | | σ | | upper | | lower | | strata | |
| 252 | 3 | 208 | 1 | 0 | 0.995192 | | 0.004819 | | 1 | | 0.985836 | | 3 | |
| 254 | 5 | 206 | 3 | 2 | 0.980699 | | 0.009745 | | 0.99961 | | 0.962146 | | 3 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 311 | 161 | 124 | 1 | 0 | 0.760723 | | 0.041756 | | 0.8256 | | 0.700944 | | 3 | |
| 313 | 164 | 122 | 1 | 0 | 0.754487 | | 0.04256 | | 0.820123 | | 0.694105 | | 3 | |
| 314 | 167 | 121 | 1 | 0 | 0.748252 | | 0.043361 | | 0.814624 | | 0.687288 | | 3 | |
| 316 | 177 | 119 | 1 | 1 | 0.741964 | | 0.044175 | | 0.809067 | | 0.680426 | | 3 | |
| 320 | 193 | 114 | 2 | 0 | 0.728947 | | 0.045914 | | 0.797586 | | 0.666215 | | 3 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 385 | 599 | 41 | 1 | 0 | 0.514905 | | 0.084437 | | 0.607575 | | 0.436369 | | 3 | |
| 386 | 604 | 40 | 1 | 0 | 0.502032 | | 0.088151 | | 0.596713 | | 0.422374 | | 3 | |
| 390 | 650 | 36 | 1 | 0 | 0.488087 | | 0.092543 | | 0.585154 | | 0.407122 | | 3 | |
| 391 | 663 | 35 | 1 | 0 | 0.474142 | | 0.096977 | | 0.573397 | | 0.392067 | | 3 | |
| 392 | 668 | 34 | 1 | 0 | 0.460196 | | 0.101468 | | 0.561453 | | 0.377201 | | 3 | |
| … | … | … | … | … | … | | … | | … | | … | | … | |
| 414 | 1138 | 12 | 1 | 0 | 0.296488 | | 0.184334 | | 0.425513 | | 0.206587 | | 3 | |
| 415 | 1168 | 11 | 1 | 1 | 0.269535 | | 0.207533 | | 0.404825 | | 0.179458 | | 3 | |
| 416 | 1172 | 9 | 1 | 0 | 0.239587 | | 0.23866 | | 0.382481 | | 0.150077 | | 3 | |
| 419 | 1327 | 6 | 1 | 0 | 0.199655 | | 0.300486 | | 0.359795 | | 0.110792 | | 3 | |
| 423 | 1517 | 2 | 1 | 0 | 0.099828 | | 0.768305 | | 0.450021 | | 0.022145 | | 3 | |

Again, we highlight the largest smaller than 0.75, 0.5 and 0.25 for both Chlamydia and Gonorrhea initial infection. The upper and lower confidence interval for the survival estimate is displayed in Table 3. We then estimate the time to re-infection quartiles as below:

We calculate a 95% confidence interval using equation **[i]**, with the following result (α = 0.05):

)

We summarize the results in the below table:

Table 4 - C.I. Summary

|  |  |  |  |
| --- | --- | --- | --- |
|  | 75th Percentile C.I. | Median C.I. | 25th Percentile C.I. |
| Gonorrhea | (60,315) | (144,1238) | (66,NA) |
| Both | (131,480) | (334,1284) | (NA,NA) |
| Chlamydia | (99,257) | (251,1138) | (242,NA) |

From Table 4 we can see that there is overlap for all confidence intervals, giving us an unclear indication whether or not the three curves do indeed differ. There is therefore not clear enough evidence to suggest whether time to re-infection is different for the three states of initial infection.

## Question 3

Conduct a single test for differences between the three survival curves. Justify your choice of the test. Also, give the complete notation for the test.

Testing for differences between the survival curves will involve equating the survival estimates for each of the three initial infections types under the null hypothesis, whereas under the alternative hypothesis at least one of the estimates should be different.

This means we set up our hypothesis like this:

Which translates to:

We employ a Fleming-Harrington test with p=1 and q=1, as we would like to place more weight on early and late re-infections, given the survival curves in Figure 1.

Table 5 - Fleming-Harrington Test

|  |  |  |  |
| --- | --- | --- | --- |
| Test of Equality over Strata | | | |
| Test | **ꭓ2** | **DF** | **p-value** |
| Fleming(1,1) | 6.5408 | 2 | 0.0380 |

Based on the Fleming-Harrington test, with a p-value of 0.038 we reject H0 and conclude that there is indeed a difference between the re-infection rates of the different initial STD infections.

## Question 4

Conduct a trend test for differences between the survival curves, using an ordering which seems natural.

We base our ordering on the evidence presented in Figure 1. This leads us to set up the hypothesis for the trend test as below:

Table 6 - Log-Rank statistics

|  |  |
| --- | --- |
| Rank Statistics | |
| Type | **Log-Rank** |
| 1 | 8.366 |
| 2 | -15.324 |
| 3 | 6.958 |

Using the log-rank statistics, we calculate with and z consisting of the log-rank statistics (i.e. the 3 is associated with the Gonorrhea hazard rates) and using the covariance matrix, we calculate

We then calculate:

This gives us a p-value of 0.4457149, so we do not reject the null hypothesis and can therefore not conclude that there is a trend to the re-infection times with the different initial infections.

## Question 5

Estimate the relative risks of re-infection for all the different risk groups (use as baseline, “gonorrhea alone” infection)

1. Assume first that these relative risks are constant over time. Find an estimator and confidence interval for these risks under this assumption

We specify the model as follows:

where,

)

Table 7 - Maximum likelihood estimates of the coefficients of the proportional hazards model for re-infection times using the Breslow method

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Analysis of Maximum Likelihood Estimates | | | | | | |
| Parameter | **DF** |  |  | **ꭓ2** | **p-value** | **RR** |
|  | 1 | -0.04813 | 0.06912 | 0.4848 | 0.4863 | 0.953 |
|  | 1 | 0.00233 | 0.0004181 | 31.0856 | <.0001 | 1.002 |

From Table 7 we can see that ( and the variable does not appear to be significant with a p-value of 0.4863. We therefore do not reject at a 0.05 confidence level. As for , we have ( and the variable does appear to be significant with a p-value of less than 0.0001 (and thus we reject at a 0.05 confidence level).

For a 95% confidence interval for the relative risk of re-infection for someone who was initially infected by both gonorrhea and chlamydia compared to someone who was only infected with gonorrhea we define relative risk as:

Using the asymptotic normality of a 95% confidence interval for the relative risk is:

Which gives us: (0.8322643,1.091273)

Similarly for , we derive a 95% confidence interval for the relative risk as:

Which gives us: (1.001512, 1.003154)

1. Determine if there is evidence that these risks are indeed constant over time. Verify this with an appropriate hypothesis test procedure, as well as the appropriate graphical checks.
2. If the risks are not constant over time, propose a different model that fits the data better and estimate only the relative risks under this new model.

# Part B

The investigator has a feeling that the use of condoms may play a role in the re-infection time.

## Question 1

Confounding for the variable “condom use”, repeat the analyses in part A, questions (1), (3) and (5a), by adjusting for and/or stratifying upon this factor “condom use”. Compare results briefly with those results in Part A.

## Question 2

Perform a statistical test, which determines what effect condom-use has on the survival curves for the three STD types (i.e. does condom-use have the same effect on all the three types of STD’s or are there different effects for the different users of condoms?)

# Part C

Use a parametric model to estimate the mean time to re-infection for the three types of initial infection (use therefore the “condom use variable” in this parametric model).

## Question 1

Find the best single parametric model, with full justifications (include all relevant model fit statistics).

## Question 2

Give a full interpretations of the estimated parameters (i.e. in terms of risk or acceleration).

## Question 3

Summarise the means in one table for the different categories of condom-use within the STD-types.

## Question 4

Perform all relevant graphical checks to verify that all assumptions have been met (including influential observation- and outlier detection). Make relevant suggestions after investigating these plots.

# Part D

Summarise your findings in about half a page, without mentioning any statistics (consider all the questions you have answered), such that the medical collaborator can understand what you have analysed. Focus on the differences between part A (pure non-parametric model vs proportional hazard assumption) and part B (condom-use included with full parametric model vs proportional hazards). Explain also the effect of condom use from your point of view.