SOLUTIONS for Practical session 3:
Causal Survival Analysis using Stata

Network of Applied Statisticians in Health

Part A: Estimate hazard ratios for ITT and PP analysis

Use the script "Final dofile (Hazard Ratios for ITT and PP analysis).do"

1. See Stata script

2. See Stata script and Table in question 8

3. See Stata script and Table in question 8

4. See Stata script and Table in question 8

5. Results between question 1 and question 2 should be very similar, as the estimand is the same (conditional Hazard ratios). If we split the follow-up time into relatively short periods, then pooled logistic regression estimates are almost the same with the Cox regression estimates (the shorter the periods, the closer the results will be). Usually, you have to split the follow-up time into >5 periods, but even with 5 periods, our estimates are very close in this example

   The difference between question 2 and question 3 is that the estimand is different (conditional hazard ratio in question 2 vs margnal hazard ratio in question 3). However, we do expect that the findings will not differ materially

6. See Stata script and Table in question 7

7. What is the difference in the interpretation of the results of questions 3 and 5?

   The estimands are different

   In Question 3, we estimate the HR for the (observational analogue of the) ITT analysis, i.e. what is the effect of initiating vs not initiating treatment on our outcome

   In Question 5, we estimate the HR for the (observational analogue of the) PP analysis, i.e. what is the effect of adhering vs not adhering in treatment on our outcome

8. Summarise your findings in the following table

Table: Hazard ratios of the ITT and PP analysis

	HR (95% CI)
Effect of initiating treatment A on Y using Cox regression and adjusting for baseline confounders in the outcome model for ITT analysis (Q2)	1.04 (0.94-1.15)
Effect of initiating treatment A on Y using pooled logistic regression and adjusting for baseline confounders in the outcome model for ITT analysis (Q3)	1.04 (0.94-1.16)
Effect of initiating treatment A on Y using pooled logistic regression and adjusting for baseline confounders through unstabilised IPW for the intention-to-treat (ITT) analysis (Q4.a)	1.04 (0.94-1.15)
Effect of initiating treatment A on Y using pooled logistic regression and adjusting for baseline confounders through stabilised IPW for the intention-to-treat (ITT) analysis (Q4.b)	1.07 (0.96-1.19)
Effect of adhering on treatment A on Y using pooled logistic regression for the per-protocol (PP) analysis using unstabilised inverse probability of adherence weights for adherence (Q6)	0.93 (0.77-1.12)

Part B: Estimate standardized risk curves

Use the script "Final dofile (Standardised risk curves).do"

In this script, we will estimate standardized risk curves for the PP analysis. The procedure is similar for the ITT analysis (changes only the initial model in step 1)

See the Stata script how we implement the 8 steps for the calculation of standardized risk curves

1. Fit a discrete-time hazards model (eg, a pooled logistic model with relatively short periods) that estimates, at each time and for each person, the probability of developing the outcome. Use the variable "time of follow-up,", along with time^2 and time^3 or a flexible functional form (splines) and estimate time-varying HRs by adding product terms between exposure and "time of follow-up."

2. Recreate all time points and two exposure groups for each subject (we want to estimate their risk curves both under exposure and under no exposure, regardless of the subject's exposure status)

3. Estimate the probability of being at risk at each time point

4. Estimate the probability of being healthy at each time point (1-probability of being at risk)

6. Multiply the model's predicted values of being healthy through time *t* to estimate the probability of remaining healthy (or probability of survival if the outcome is death) at *t* for subjects with their same combination of covariate values.

6. Predict the probability of being at risk at time *t* for each subject both under exposure and under no exposure, regardless of the subject's exposure status (1- probability of remaining healthy)

7. Separately average the adjusted risk curves under exposure and under no exposure, over all subjects. This last step effectively standardizes the curves to the empirical distribution of the covariates in the study, and results in 2 marginal risk curves: one under exposure, another under no exposure.

8. Estimate 95% CI using bootstrap

Results:

Risk in the treatment A=0 group: 35.7% (29.7% , 38.5%)

Risk in the treatment A=1 group: 33.7% (30.0% , 34.8%)

Risk difference (Risk A=1 vs Risk A=0): -2.0% (-6.4% , 3.0%)