The University of Hong Kong
School of Public Health

Master of Public Health
Advanced Statistical Methods II
June 26, 2023

Assignment 2 – Survival analysis
(*Submit to Moodle by 23:55, July 6, 2023*)

Explore time to cessation of symptoms and viral shedding in subjects with confirmed influenza virus infection

A study was conducted in 2007 to investigate household transmission of influenza virus (Cowling et al., 2008, PLoS ONE). An extract of the data regarding duration of illness in index cases may be downloaded on moodle (named as "pilot2007.csv").

The description of the variables is as follows:

- *hhID* is a subject identifier.
- *male* indicates male sex
- *onsettime* is the delay between symptom onset and recruitment to the study (not used here).
- *vac1* = 1 indicates receipt of vaccine prior to the flu season
- *flu.type* indicates the virus type
- *antibiotics, antiviral, antihistamine, antipyretic, steroid* indicate treatments prescribed
- Individuals with chronic diseases were recorded with *chor_disease*=1,
- *time*, *event* is the outcome – time to cessation of symptoms (*event*=1) or censoring (*event*=0)
- *age* is in years
- *timeL, timeR* are the interval-censored times to cessation of viral shedding

All answers may be given to 2 decimal places, or if quoted as percentages may be rounded to the nearest percentage point.

1. Fit a proportional hazards model to time to cessation of symptoms adjusting for age (0-6y, 7-15y and ≥16y), flu type, antibiotics, antiviral and antihistamine prescription. Present your results in a table. [3 mark]

2. Assess the proportional hazards assumption for the model in (1). [2 mark]

3. Fit 3 parametric accelerated failure time models to time to cessation of symptoms with the same variables as in (1), using lognormal, Weibull and exponential distributions. Present your results in a table. [3 mark]

4. Selection the best model in (3). [1 mark]

5. Assess the model fitness of the selected model in (4). [1 mark]

6. Using the same predictor variables, fit 3 parametric accelerated failure time models to time to cessation of viral shedding, using lognormal, Weibull and exponential distributions. Present your results in a table. [3 marks]

   [Hint: The left censored time at time=0 may cause problem. In that case you may change it to a small number, e.g. 0.0001 without affecting the results practically]

7. Selection the best model in (6). [1 mark]

8. Assess the model fitness of the selected model in (7). [1 mark]

9. Draw an overall conclusion based on the above analyses. [2 marks]