## **Supplementary Text: Model Equations**

- $B/B: Surv(X,d) \sim trt\_young+trt\_middle+trt\_old+middleagegrp+oldagegrp\\ + hh\_ward + large\_cluster + hh\_ownhouse + hh\_stableoccupation + hh\_hiexpenditure\\ + hh\_farfromwater + cluster(cl\_id)$
- $T/T: Surv(tstart, tstop, d) \sim trt\_young + trt\_middle + trt\_old + middle agegrp + old agegrp + hh\_ward + large\_cluster + hh\_ownhouse + hh\_stable occupation + hh\_hiexpenditure + hh\_farfromwater + cluster(cl_id)$
- $B/T: Surv(tstart, tstop, d) \sim trt\_byoung+trt\_bmiddle+trt\_bold+middleagegrp+oldagegrp\\ + hh\_ward + large\_cluster + hh\_ownhouse + hh\_stableoccupation + hh\_hiexpenditure\\ + hh\_farfromwater + cluster(cl\_id)$
- $T/B: Surv(tstart, tstop, d) \sim trt\_young+trt\_middle+trt\_old+bmiddleagegrp+boldagegrp\\ + hh\_ward + large\_cluster + hh\_ownhouse + hh\_stableoccupation + hh\_hiexpenditure\\ + hh\_farfromwater + cluster(cl\_id)$

The formula B/B is used together with the un-augmented dataset, and all the age group-related predictors in the formula are calculated at baseline. The other three formulae are used together with the augmented dataset for time-dependent covariate analysis [1], in which the variables trt\_young, trt\_middle, trt\_old, middleagegrp and oldagegrp are calculated using age group during the time interval defined by tstart and tstop, while the variables trt\_byoung, trt\_bmiddle, trt\_bold, bmiddleagegrp and boldagegrp are calculated using baseline age group.

## Reference

[1] Therneau T, Crowson C. Using time dependent covariates and time dependent coefficients in the Cox model. R Survival Package Vignette. 2014.