Supplementary file 2- A modified dataset to run R and Stata code.

**The dataset**

The dataset is a set of 200 simulated lung cancer patients. These patients are followed up for a year following their cancer diagnosis: 106 of them received surgery within six months of their diagnosis and 48 died in the year.

**Variables**

\*id: patient identifier

\*fup\_obs: observed follow-up time (time to death or 1 year if censored alive)

\*death: observed event of interest (all-cause death) 1: dead, 0:alive

\*timetosurgery: time to surgery (NA if no surgery)

\*surgery: observed treatment 1 if the patient received surgery within 6 month, 0 otherwise

\*age: age at diagnosis

\*sex: patient's sex

\*perf: performance status at diagnosis

\*stage: stage at diagnosis

\*deprivation: deprivation score

\*charlson: Charlson's comorbidity index

\*emergency: route to diagnosis

**Stata Results**

Using 100 bootstrap samples, these are the results:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Original cohort** |  | **1-year survival (%)** | **95% CI\*** | |  | **RMST (days)** | **95% CI\*** | |
| **Treated** | Yes | 84.9 | 76.0 | 90.6 |  | 279 | 232 | 289 |
|  | No | 66.0 | 54.2 | 75.5 |  | 300 | 262 | 318 |
| ***Differences1*** | | *18.9* | *6.3* | *32.4* |  | *-21* | *-73* | *15* |
|  |  |  |  |  |  |  |  |  |
| **Emulated cohort** |  | **1-year survival (%)** | **95% CI\*** | |  | **RMST (days)** | **95% CI\*** | |
| **KM** |  |  |  |  |  |  |  |  |
| Surgery arm | | 81.7 | 73.4 | 87.8 |  | 326 | 309 | 339 |
| No surgery arm | | 67.4 | 56.0 | 76.2 |  | 314 | 289 | 333 |
| ***Differences2*** | | *14.3* | *0.0* | *24.2* |  | *12* | *-9* | *21* |
| **Weighted KM** |  |  |  |  |  |  |  |  |
| Surgery arm | | 81.6 | 73.7 | 88.4 |  | 326 | 309 | 339 |
| No surgery arm | | 69.7 | 58.6 | 80.9 |  | 321 | 299 | 339 |
| ***Differences3*** | | *11.9* | *3.4* | *24.1* |  | *5* | *-2* | *29* |

\* The 95% CI were calculated using 100 bootstrap replicates

1 These differences are prone to both confounding and immortal-time biases

2 These differences are prone to informative censoring

3 These differences account for all types of biases, under the assumptions detailed in method