|  |  |  |  |
| --- | --- | --- | --- |
|  | Death | Recovered | Under treatment |
| Onset Date – Event Date |  |  |  |
|  |  |  |  |
|  |  |  |  |

Data pre – processing

One Imputation - Latest

1. After exporting imputed file from python
2. Death Null or blank imputed with 0 , dates fixed
3. Recovered null or blank imputed with 0 , dates fixed
4. Added as of date to calculate number of days
5. Put formula – Recovery\_ Time & Survival\_Time if 0 – as of date – symptom onset else recovered date – symptom onset
6. Put Recovery\_Status & Survival\_Status value based on event happen (2) or censor(1)
7. Drop records that say 2/30/1899 as recovery dates – 12 rows
8. For recovery as 1 – found the date of recovery by reading the summary – Last 4 data rows do not have the date in summary – dropped the two rows
9. Symptom Onset – Blank –with hospital visited date – copied the date to Symptom onset date
10. Equated symptom onset date with NA/0 to Exposure end date for 228 rows
11. 2405 rows blank symptom onset date deleted
12. 975 rows of data left
13. For Death 0 but recovered 1 manually change the survival time to 5000
14. For Death 1 but recovered 0 manually change the survival time to 8000

Recovery Data

1. Check the blanks in recovery. For these blanks check blanks in death . This data set means that the patient status is unknown – he is recovering because he is not dead - Removed 13th records where the data in death was not blank
2. Drop records that say 2/30/1899 as recovery dates – 12 rows
3. For recovery as 1 – found the date of recovery by reading the summary – Last two data rows do not have the date in summary – dropped the two rows
4. Create new column – Recovery\_time condition – if 0 – reported date – symptom onset else recovered date – symptom onset
5. Reporting Date – NA – One row removed
6. Symptom Onset – Blank – For 78 rows with hospital visited date – copied the date to Symptom onset date.
7. Deleted 70 rows with Death as non 0
8. Equated symptom onset date with NA/0 to Exposure end date for 228 rows
9. For case 237 – manually changed the symptom onset date to reporting date
10. Remove symptom onset date with nulls and blanks
11. Left with 927 rows of data
12. 110 rows with gender NA removed – Imputed Male after reading the summary column for one row
13. Put death as infinite number ( here 400 )
14. Removing 2565 with gender = 20
15. Female = 2 & Male = 1 ; Deleted NotKnown (1591 rows of data)
16. Copy hospital visit date if symptom onset date is missing
17. Symptom Onset – Blank – For 68 rows with hospital visited date – copied the date to Symptom onset date
18. Equated symptom onset date with NA/0 to Exposure end date for 228 rows
19. Removed 926 rows without symptom onset date
20. For case 236 – manually changed the symptom onset date to reporting date
21. Put recovery as infinite number ( here 400 )

Two cases :

1. Event – Death - Death and under treatment - Have a new data file – death\_covid
2. Event – Recovered – Recovered and under treatment – Have a new data file – Recovered\_covid

Assumption :

What we are trying to analyze :

1. Who has a better survival rate when it comes to death – male/female
2. Who has a better recovery rate – male female ( here event is recovery)