Harmonisation project R code

|  |  |  |
| --- | --- | --- |
| Variables | Label | note |
| patid | Patient ID |  |
| DM | Exposure: Diabetes | (1-yes, 0-no) |
| agein | Age at baseline [(index date – date of birth)/ 365.25] | Numeric e.g., 60.56 |
| yearin | Index date | Numeric, e.g., 2011.08 |
| outm | End of follow-up date | Numeric, e.g., 2019.04 |
| crm | Cardio-renal mortality | (1-yes, 0-no) |
| acm | All-cause mortality | (1-yes, 0-no) |

Note:

1. I have a database including people with and without diabetes, if you have separate databases, you don’t have to use subset (db, DM==1);
2. yearin & outm should be stored as numeric, for example, 30/01/2011 should be saved as 2011.08 [(the day number of the year/the total day number of the year) + the year].
3. You can change the age range and period range based on your database.

Table 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Diabetes | | |
| **Country** | Canada | Spain | UK |
| **Total participants** |  |  | 114,909 |
| **Median follow-up, (IQR, years)** |  |  | 9.02 (5.53, 12.83) |
| **Median age at diagnosis (IQR, years)** |  |  | 61 (52, 71) |
| **Median year of entry, (IQR, year)** |  |  | 2008 (2004, 2012) |
| **Diagnosis age group** |  |  |  |
| <40 |  |  |  |
| 40-50 |  |  |  |
| 50-60 |  |  |  |
| 60-70 |  |  |  |
| >70 |  |  |  |
| **number of patients by year of diagnosis** |  |  |  |
| 1998 |  |  |  |
| 1999 |  |  |  |
| 2000 |  |  |  |
| … |  |  |  |
| … |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| Female (%) |  |  |  |
| Ethnicity \* (Caucasian/South Asian, Black, other) |  |  |  |
| Deprivation (quintile)\* |  |  |  |
| \*BMI- most recent around index date |  |  |  |
| \*Smoking-most recent around index date |  |  |  |
| Cardio-renal mortality (%) |  |  |  |
| All-cause mortality (%) |  |  |  |

\* If available