Cohort define steps

1. From the entire database (**ED**), extract T2DM cohort with any codes in sheet “exposed”; then remove patients with any codes in sheet “exclude” before the end of the study (31/12/2018); use the first appearance diagnosis code of T2DM as the index date for exposed patient cohort (**T2C**).
2. From this cohort (**T2C**), remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; this is the exposed cohort (**EC**).
3. From the entire database (**ED**), remove patients with any codes in sheet “non-exposed pool” before the end of the study (31/12/2018), to get the candidate non-exposed patients (**CNE**).
4. Exact matching the exposed cohort (**EC**) to the candidate non-exposed patients (**CNE**) with a ratio 1:10 (**EC:CNE**) by year of birth (+/-1year), sex, and practice, without replacement (each candidate non-exposed patient can be only matched once). This is the matched cohort (**MC**), and the index date is the same as the matched exposed patient.
5. From the matched cohort (**MC**), remove patients died before the index date; then remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; then keep a randomly selected 5 matched non-exposed patients (good to set a seed to make the random selection replicable). This is the final non-exposed cohort (**FNE**).
6. The final study cohort is the combination of the exposed cohort (**EC**) and the final non-exposed cohort (**FNE**).