International trends in cause-specific mortality among people with and without diabetes

September 2, 2024
https://github.com/jimb0w/CM

Correspondence to:
Jedidiah Morton
Jedidiah.Morton@monash.edu
Research Fellow
Baker Heart and Diabetes Institute, Melbourne, Australia
Monash University, Melbourne, Australia

Contents

1 Data cleaning

This is the protocol for an analysis of trends in cause of death (COD) in people with and without diabetes across several countries over the period spanning 2000 to 2021.

We have been provided with many different variables and some countries have restrictions on what data they can provide, so we need to harmonize and clean the data into an analysable format.

The variables we will derive are:

- Calendar year
- Sex
- Mid-point age for the age-group
- Person-years of follow-up in people with diabetes
- Person-years of follow-up in people without diabetes
- Number of deaths for each COD in people with diabetes
- Number of deaths for each COD in people without diabetes

The COD are shown in Table 1.1.

1.1 Australia

For Australia, we have the following variables (by age, sex, and calendar year): total population size, person-years in people with diabetes, deaths in people with diabetes, and deaths in the total population. We can calculate person-years in the total population by assuming that the person-years of follow-up in a given calendar year are equal to the population size in the current year plus the population size in the next year, divided by two [this has been performed before I got the dataset–JM]. From there, person-years in people without diabetes is just person-years in the total population minus person-years in people with diabetes. Similarly, for deaths in people without diabetes, we can subtract the deaths in people with diabetes from the total deaths.

Australian data restrictions prohibit the use of any cell count <6 for the diabetes population; thus, there are many blank values (see below). I will fill them in randomly, where the number can be any number from 0 to 5 with equal probability, unless the number of deaths in the total population for the age/sex group is <5, in which case the upper bound will be the number of deaths in the total population. Further, because of this, data has been provided in both 10-year age groups and overall (i.e., the actual counts). My intuition is that the small cell counts won't drive any overall results anyway, which I check below (Figure **??**), and that the uncertainty associated with such low numbers will be reflected in very wide confidence intervals for the younger ages.

Table 1.1: Causes of death in the present analysis

Causes of death	Abbreviation	ICD-10	ICD-9
Cardiovascular diseases	CVD	I00-I99	390-434, 436-459
Ischaemic heart diseases	CHD	I20-I25	410-414, 429.2
Cerebrovascular diseases	CBD	I60-I69	430-434, 436-438
Heart failure	HFD	I50	428
Cancer	CAN	C00-C97	140-208
Diabetes	DMD	E10-E14	250
Infectious diseases	INF	A00-B99	001–033, 034.1–134, 136–139, 771.3
Influenza and pneumonia	FLU	J09-J18	480-487
Chronic lower respiratory diseases	RES	J40-J47	490-494, 496
Liver diseases	LIV1	K70-K76	570-572, 573.0, 573.3-573.9
Liver diseases (exclude alcoholic liver disease)	LIV2	K71-K76	570, 571.4-571.9, 572, 573.0, 573.3-573.9
Renal diseases	CKD	N00-N08, N17-N19, N25-N27	580-589
Dementia	AZD	F00, F01, F03, G30	290.0-290.2, 290.4, 331.0

References